CHAPTER 6

County Estimation of Crop Acreage Using Satellite Data

Michael Bellow, Mitchell Graham, and William C. Iwig
National Agricultural Statistics Service

1993

6.1 Introduction and Program History

The National Agricultural Statistics Service (NASS) of the U.S. Department of Agriculture (USDA) has published county estimates of crop acreage, crop production, crop yield and livestock inventories since 1917. These estimates assist the agricultural community in local agricultural decision making. Also the Federal Crop Insurance Corporation (FCIC) and the Agricultural Stabilization and Conservation Service (ASCS) of the USDA use NASS county crop yield estimates to administer their programs involving payments to farmers if crop yields are below certain levels. The primary source of data for these estimates has always been a large non-probability survey of U.S. farmers, ranchers, and agribusinesses who voluntarily provide information on a confidential basis (see Chapter 7). In addition, the Census of Agriculture, conducted by the Bureau of the Census every five years, serves as a valuable benchmark for the NASS county estimates.

Earth resources satellite data, particularly from the Landsat series of satellites, provide another useful ancillary data source for county estimates of crop acreage. The potential for improved estimation accuracy using satellite data is based on the fact that, with adequate coverage, all of the area within a county can be classified to a crop or ground cover type. The accuracy of the estimates is then dependent on how accurately the satellite data are classified to each crop type based on the "ground truth" data obtained from the annual June Agricultural Survey (JAS) conducted by NASS. Through the use of aerial photographs, this survey identifies the crop type of individual fields within randomly selected land segments. Segments in major agricultural areas are approximately one square mile in area and normally contain 10 to 20 fields. The satellite spectral data are matched to the corresponding fields for use in classifying all individual imaged areas, known as pixels, to a particular crop type. Recent studies (Bellow 1991; Bellow and Graham 1992) have shown that, for certain crops, approximately 80 percent of the pixels are classified correctly. This correct classification level is high enough to provide improved estimation accuracy.

NASS has been a user of remote sensing products since the 1950's when it began using mid-altitude aerial photography to construct area sampling frames (ASF's) for the 48 states of the continental United States. A new era in remote sensing began in 1972 with the launch of the Landsat I earth-resource monitoring satellite. Four additional Landsats have been launched since 1972, with Landsat IV and V still in operation in 1993. The polar-orbiting Landsat satellites contain a multi-spectral scanner (MSS) that measures reflected energy in four bands of the electromagnetic spectrum for an area of just under one acre. The spectral bands were selected to be responsive to vegetation characteristics. In addition to the MSS sensor, Landsats IV and V have a Thematic Mapper (TM) sensor which measures seven energy bands and has increased spatial resolution. The large area (185 by 170 km) and repeat (16 day per satellite) coverage of these satellites opened new areas of remote sensing research: large area crop inventories, crop yields, land cover mapping, area frame stratification, and small area crop cover estimation.

Research from 1972 to 1978 led to the creation of an operational procedure for large area crop acreage estimation. A regression estimator was developed which related the ground-gathered area frame data to the computer classification of Landsat MSS images. The basic regression approach used to produce State estimates does not produce reliable county estimates. Domain indirect regression estimators were developed for this purpose. In the 1978 crop season, corn and soybean acreage State and county estimates based on remotely sensed data were produced for Iowa. One to two States were added to the project through 1984. For the 1984-1987 crop seasons, this project covered an eight-State area in the central United States and produced regression estimates of corn, winter wheat, soybeans, rice, and cotton acreages. These regression estimates were combined with other survey indications and administrative data to provide final published county estimates. Estimation based on data from Landsat MSS sensors was discontinued in 1988 in order to implement the increased capabilities of higher resolution sensors.

France entered the field of earth resources satellites in 1986 with the launch of SPOT I, which carries an improved multi-spectral scanner. This scanner images an even smaller area than the TM sensor but only measures three energy bands. Several NASS research projects compared the SPOT MSS and Landsat TM sensors with respect to crop estimation. This research led to the selection of Landsat TM as the preferred sensor for crop area estimation based on its superior spectral characteristics. The spatial characteristics of the SPOT MSS sensor provide a benefit only in areas with mostly small fields.

Regression estimation of crop acreages for large and small areas based on computer classification was reinstated in 1991 with the Delta Remote Sensing Project using Landsat Thematic Mapper data imaged over the Mississippi Delta region, which is a major rice and cotton area. Results from the operational eight-State program in 1987 and from sensor comparison experiments showed that the regression approach was most effective for rice and cotton estimation. State and county estimates of rice, cotton, and soybean acreages were produced for Arkansas and Mississippi in 1991, with Louisiana added in 1992. The project only covers Arkansas in 1993 due to budgetary constraints.

Three domain indirect regression estimators have been used or considered for producing small area county estimates using ancillary satellite data. From 1976 to 1982, the Huddleston-Ray estimator was used (Appendix B). In 1978, the Cardenas family of estimators was considered but not implemented (Appendix C). Beginning in 1982, the Battese-Fuller family of estimators was used for calculating county crop acreage estimates using Landsat MSS data. Since 1991, the Battese-Fuller model has been used to produce county estimates with Landsat TM data. Currently, this is the preferred model. However, non-regression estimation procedures based on total pixel counts are being evaluated.

6.2 Program Description, Policies, and Practices

The basic element of Landsat 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 several bands of the electromagnetic spectrum. The individual imaged areas, known as pixels, are arrayed along east-west rows within the 185 kilometer wide north-to-south pass (swath) of the satellite. For purposes of easy data storage, the data within a swath are subdivided into overlapping square blocks, called scenes. The two satellites currently in operation (Landsats IV and V) image a given point on the earth's surface once every 16 days. The MSS sensor, formerly used for crop area estimation, contained four spectral bands with 80 meter spatial resolution. The more advanced TM sensor has seven bands (three visible and four infrared) with 30 meter resolution.

Several Landsat scenes may be required to cover an entire region of interest within a given State. It is not always possible to have the same image date for all such scenes due to schedule, cloud cover, and image quality factors. Consequently, analysis districts are created. An analysis district is a collection of counties or parts of counties contained in one or more Landsat scenes that have the same image date, or in areas for which usable Landsat data is not available to the analyst. To obtain State level crop acreage estimates, NASS sums all analysis district level estimates within the State. County level estimates are obtained using domain indirect regression and synthetic estimation methods, to be discussed later.

The area sampling frame for each State is stratified based on land use such as percentage cultivation, forest, and rangeland. NASS uses the regression estimator described by Cochran (1977, pp. 189-204) to compute crop acreage estimates for each land use stratum within an analysis district that has satellite coverage for an adequate number of JAS segments. These regression estimates are more precise than the direct expansion estimates obtained from JAS data alone. A detailed description of the procedure involved is provided by Allen (1990). Briefly, the steps required are as follows:

1. A graphics oriented registration process associates Landsat pixels with JAS sampled segments.

2. JAS data for sampled segments are used to label each pixel within the segments to a crop or other cover type.

3. Labelled pixels are clustered based on their Landsat data values to develop discriminant functions (signatures) for each cover type.

4. The discriminant functions are used to classify each pixel within the sampled segments to a cover type.

5. The segment level classification results are used to develop regression relationships for each crop between the ground and satellite data within each land use stratum. For each stratum, the independent (regressor) variable is the number of pixels classified to that crop per segment, and the dependent variable is the JAS segment reported crop acreage.

6. All pixels within the analysis district are classified, using the discriminant functions developed in Step 3.

7. For each stratum, the mean number of pixels per segment classified for a given crop over all segments in the population is substituted into the corresponding regression equation to obtain the stratum level mean crop acreage per segment. This mean is multiplied by the known total number of segments in the stratum to obtain the stratum level crop acreage estimate.

8. The stratum level estimates are summed to obtain the analysis district level crop acreage estimate for the portion of the analysis district covered by satellite data.

For land use strata lacking satellite coverage of an adequate number of JAS segments to develop the regression relationship, the direct expansion of JAS data is used to obtain estimates. These stratum level JAS estimates are also summed to obtain analysis district estimates for each crop representing the area <u>not</u> covered by satellite data. The total analysis district estimate for a particular crop is then:

$$\hat{T}_{a} = \hat{T}_{(REG),a.} + \hat{T}_{(DE),a.}$$

where:

$\hat{T}_{(REG),a.}$ = sum of the regression estimates over the land use strata with satellite coverage for analysis district a

$\hat{T}_{(DE),a.}$ = sum of the JAS direct expansion estimates over the land use strata without satellite coverage for analysis district a.

The final estimate of total crop acreage in the State is made by adding the analysis district estimates:

$$\hat{T}_{s} = \Sigma\, \hat{T}_{a}\ .$$

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 then use the above procedure to obtain county level estimates. Instead, the Huddleston-Ray, Cardenas, and Battese-Fuller domain indirect regression estimators have been developed and investigated for providing county estimates of crop acreage. The Battese-Fuller approach is currently favored by NASS, and is described in detail in Section 6.3.

The NASS County Estimates system, described in Chapter 7, is designed to accept the Battese-Fuller values as a separate set of county crop acreage estimates. Within this system, the Battese-Fuller county estimates are first scaled to be additive to the official NASS State estimate for each commodity. The scaled Battese-Fuller values are then composited with scaled values from other NASS surveys and administrative data sources. Thus the Battese-Fuller estimates serve as an additional input to the County Estimates system in States where they are available. Currently, the composite weights are subjectively set by the statisticians in the State office to provide satisfactory and reliable estimates. Each NASS State Statistical Office (SSO) prepares their own annual publication of the final county estimates. Although sampling variances are calculated for the Battese-Fuller estimates, no variances or error information are published for the final county estimates. Mean squared error information is only published for major agricultural items at the U.S. level.

6.3 Estimator Documentation

The Battese-Fuller family of estimators was first developed in the general framework of linear models with nested error structure (Fuller and Battese 1973), and later applied to the special case of county crop area estimation (Battese, Harter, and Fuller 1988). The method has been used for all Landsat county estimation done by NASS since 1982.

Similar to the State level estimation, land use strata are separated into those that have adequate satellite coverage and those that do not. The Battese-Fuller model can be applied within an analysis district for all strata where classification and regression have been performed. The analyst computes stratum level Battese-Fuller acreage 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 is used to obtain county estimates.

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

6.3.1 Application of Battese-Fuller Model within Regression Strata

The Battese-Fuller model for county level estimation assumes that segments grouped by county admit the same slope relationship as the analysis district but that a different intercept is required. For each sample segment within regression stratum h in county c, the Battese-Fuller model proposes the following relation at the analysis district level:

$$\begin{aligned} y_{hci} &= \beta_{0h} + \beta_{1h}\, x_{hci} + \omega_{hci} \\ &= \beta_{0h} + \beta_{1h}\, x_{hci} + \nu_{hc} + \epsilon_{hci}, \quad i=1,\ldots,n_{hc} \end{aligned}$$

where:

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

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

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

ω_{hci} = regression error for stratum h, county c, segment i.

The two components of the error term ω_{hci} are the county effect ν_{hc} and the random error ϵ_{hci}. They are assumed to be independent and normally distributed, with mean 0 and variances σ^2_{vh} and σ^2_{eh}, respectively. The covariance structure of the regression error terms is then:

$$
\begin{aligned}
\mathrm{cov}(\omega_{hci}, \omega_{hkm}) &= 0, && \text{if } c \neq k \\
&= \sigma^2_{vh} && \text{if } c=k,\ i \neq m \\
&= \sigma^2_{vh} + \sigma^2_{eh}, && \text{if } c=k,\ i=m
\end{aligned}
$$

The parameter σ^2_{vh} is both a within county co-variance and a between county component of the variance of any residual, while σ^2_{eh} 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 estimates for stratum h.

For a given county the stratum level mean crop area per 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 segment classified to crop in stratum h, county c

= X_{hc} / N_{hc}

X_{hc} = number of pixels classified to crop in stratum h, county c

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

δ_{hc} = weighting factor selected by user ($0 \leq \delta_{hc} \leq 1$).

The total number of pixels classified to the crop of interest in stratum h, county c (X_{hc}) is provided by the analysis software. The total number of segments in stratum h, county c is obtained during the development of the area sampling frame.

The mean square error of this estimator is:

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

This expression is conditional on known values of σ^2_{vh}, σ^2_{eh}, and δ_{hc}. The estimator of the mean square error is known to have a downward bias due to estimation of the variance components, and a correction due to Prasad and Rao (1990) is currently under investigation. In general, these parameters are unknown and will be estimated as discussed below.

Conditioned on the county effects, the mean squared bias is:

$$MSCB\ (\bar{y}_{(BF),hc.}) = (1-\delta_{hc})^2\sigma_{vh}^{\ 2}$$

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

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

The range of allowed values of δ_{hc} defines a family of Battese-Fuller estimators. If $\delta_{hc} = 0$, then the estimate of $\bar{y}_{(BF),hc.}$ lies on the analysis district regression line for the stratum. The mean square error for stratum h in county c is minimized by:

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

In general, the variance components σ^2_{vh} and σ^2_{eh} are unknown, so they must be estimated to approximate δ_{hc}^* using the above formula. The estimators currently used (Appendix A) 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. If county estimates are generated for an entire State, adjustment terms are used to ensure the county estimate sums match the corresponding district and State estimates. If county estimates are only provided for a region of the State, the unadjusted estimates are considered the best estimates. The resulting adjusted Battese-Fuller estimator is:

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

where:

N_h = number of segments in stratum h.

The adjusted estimates sum to the appropriate analysis district totals. The estimate of total crop area in the regression strata of county c is:

$$\hat{T}_{(BF),c} = \sum_{h=1}^{H_r} \hat{T}_{(uBF),hc.} \quad \textit{(unadjusted form)}$$

or:

$$\hat{T}_{(BF),c} = \sum_{h=1}^{H_r} \hat{T}_{(aBF),hc.} \quad \textit{(adjusted form)}$$

Estimation of the mean square error, bias, and variance of the unadjusted Battese-Fuller estimator can be accomplished via substitution of the selected values of δ_{hc} and estimates of σ^2_{vh} and σ^2_{eh} into the formulas for mean square error and mean square conditional bias given above (Walker and Sigman, 1982).

6.3.2 Application of Synthetic Estimator within Synthetic Strata

As mentioned earlier, domain indirect synthetic estimation is used to obtain crop area estimates for land use strata where regression is not viable. Since each county usually contains very few segments if any for a given stratum, the stratum level mean crop acreage per segment over the entire analysis district is used to compute the synthetic estimates. For synthetic stratum h, the estimate of crop area in county c is:

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

where:

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

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

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

When doing the synthetic estimation, it is important to realize that the use of the analysis district level average to estimate county totals ignores county effects. Therefore, the synthetic component of a county estimate is often biased. In general, the larger and more heterogeneous the synthetic stratum, the greater will be this bias. It is therefore advantageous to use synthetic strata that have consistent agricultural intensity for the crops of interest.

6.3.3 Total County Estimate

The final county estimate is obtained by adding the regression and synthetic components together:

$$\hat{T}_c = \hat{T}_{(BF),c} + \hat{T}_{(SYN),c} \ .$$

If a county is split between two analysis districts, the estimates, $\hat{T}_{c,}$ from each district will be added to provide the total estimate.

6.4 County Estimates Example

This section describes an example of the computation of county estimates, using 1988 Iowa data. The example is taken from a research project that compared the TM sensor with the French SPOT multi-spectral scanner with regard to estimation effectiveness (Bellow 1991). However, county estimation was only done with the TM data. The research site was a nine county region in western Iowa, where corn and soybeans are the major crops. Ground data from the 1988 JAS were used, involving a sample of 30 segments. Of the 30 segments used for the study, 28 came from stratum A (agricultural) and the other two from stratum B (agri-urban). Only 23 of the 28 segments in stratum A were located in both the Landsat TM and SPOT scenes. Data from these 23 segments were used in the Battese-Fuller estimator.

TM based estimates of corn and soybean acreage were computed for all nine counties in the study area. Three counties (Calhoun, Crawford, and Ida) were not completely contained within the TM scene. Table 1 shows the total number of segments in the population (N_{hc}) and the number of sample segments (n_{hc}) in each county, broken down by stratum. Synthetic estimation

Table 1: Population (N_{hc}) and Sample (n_{hc}) Number of Segments by County, Stratum

County	Stratum A N_{hc}	Stratum A n_{hc}	Stratum B N_{hc}	Stratum B n_{hc}
Audubon	436	3	19	0
Calhoun	562	3	22	0
Carroll	566	1	39	0
Crawford	709	6	50	0
Greene	566	4	23	0
Guthrie	586	2	34	0
Ida	432	2	20	0
Sac	573	4	44	1
Shelby	579	3	31	1
Total	5,009	28	282	2

was used within stratum A for the parts of counties outside the scene, and in stratum B for all nine counties.

Table 2 gives the computed county estimates by stratum and estimation method. Table 3 contains the official county estimates issued by the Iowa Agricultural Statistics Service. These published estimates are based on additional survey and administrative data (see Chapter 7), and are considered as the standard for evaluating the Battese-Fuller model values. The tables show that the computed county estimates for corn were more efficient overall than those for soybeans. For eight of the nine counties, the C.V. for corn was less than 4 percent. No county had a C.V. of less than 4 percent for soybeans. The percent difference ranged from 0.2 to 9.2 for corn, and from 0.8 to 17.8 for soybeans.

Table 2: Iowa 1988 County Estimates of Crop Acreage by Stratum and Estimation Method

County	Stratum A Battese-Fuller	Stratum A Synthetic	Stratum B Synthetic	Total	C.V.
Corn	acres (000)	acres (000)	acres (000)	acres (000)	percent
Audubon	91.9	-	.3	92.2	3.5
Calhoun	130.3	2.6	.4	133.2	2.9
Carroll	140.7	-	.7	141.4	3.2
Crawford	128.4	23.4	.9	152.7	3.1
Greene	129.6	-	.4	130.0	3.0
Guthrie	105.7	-	.6	106.3	4.9
Ida	43.4	63.2	.4	107.0	3.7
Sac	137.5	-	.8	138.3	2.9
Shelby	140.2	-	.5	140.7	2.9
Total	1047.7	89.2	5.0	1141.8	
Soybeans					
Audubon	69.8	-	.1	69.9	6.6
Calhoun	143.2	1.7	.1	145.0	4.0
Carroll	106.6	-	.1	106.7	9.0
Crawford	91.3	15.5	.2	106.9	5.4
Greene	117.4	-	.1	117.5	4.6
Guthrie	64.3	-	.1	64.4	10.9
Ida	34.6	41.7	.1	76.4	6.9
Sac	112.8	-	.1	112.9	4.9
Shelby	80.9	-	.1	81.0	7.4
Total	820.9	58.8	1.0	880.7	

Table 3: County Estimates for 1988 Iowa Study

County	Corn Official	Corn Computed	Corn % Diff.*	Soybeans Official	Soybeans Computed	Soybeans % Diff.*
	acres (000)	acres (000)		acres (000)	acres (000)	
Audubon	100	92.2	7.8	70.7	69.9	1.1
Calhoun	133	133.2	0.2	150.0	145.0	3.3
Carroll	141	141.4	0.3	117.0	106.7	8.8
Crawford	147	152.7	3.9	106.0	106.9	0.8
Greene	125	130.0	4.0	143.0	117.5	17.8
Guthrie	98	106.3	8.5	77.5	64.4	16.9
Ida	112	107.0	4.5	75.2	76.4	1.6
Sac	136	138.3	1.7	124.0	112.9	9.0
Shelby	155	140.7	9.2	94.9	81.0	14.6
Total	1,147	1,141.8		958.3	880.7	

* $\%Diff. = \frac{|\,Official - Computed\,|}{Official} * 100$

6.5 Evaluation Practices

NASS first began to address the problem of applying satellite data to small area estimation in the mid 1970's. In 1976, Huddleston and Ray (1976) proposed that within each stratum, the mean pixels per segment calculated by classifying all segments within an entire analysis district be replaced by the mean pixels per segment computed by classifying all segments within a given county. This county pixel mean is substituted into the corresponding stratum regression equation for the crop of interest. Amis, Martin, McGuire, and Shen (1982) describe the Huddleston-Ray estimator as an analysis district regression estimator applied to a subarea of the analysis district. The regression coefficients are estimated from sampled segments located throughout the analysis district, while the mean being estimated is from a subpopulation of the analysis district. The Huddleston-Ray estimator is simple and intuitively appealing, but Walker and Sigman (1982) point out two major drawbacks. First, it is unclear how to accurately compute the variance of the estimator. Second, the estimator lumps together a term attributable to sampling error within a given county and another term that measures the inherent distinction between a county and the analysis district. Amis et al. (1982) empirically demonstrate that the Huddleston-Ray method can generate biased estimates and that the variance estimation formula can overestimate the variability for a given county. The mathematical formulas for the Huddleston-Ray estimator and its variance estimator are provided in Appendix B.

The problems with the Huddleston-Ray estimator documented by Walker and Sigman (1982) and by Amis et.al (1982) were recognized soon after its development and prompted Cardenas, Blanchard, and Craig (1978) to devise a different type of estimator. The Cardenas family of estimators has three forms, each of which uses auxiliary Landsat data through a regression type estimator. However, the versions use different methods of estimating the slope term. The three forms are the ratio estimator, the separate regression estimator, and the combined regression estimator. (Appendix C gives the mathematical formulation for the Cardenas family of estimators.) As with the Huddleston-Ray method, within each stratum the Cardenas method compares the analysis district level mean pixels per segment classified to a crop to the corresponding county level mean for that crop. However, the Cardenas methods uses all segments in the analysis district to calculate the analysis district mean, where the Huddleston-Ray approach only uses sample segments. The estimate of average crop area per segment is adjusted by an amount proportional to this difference between the county and analysis district means. Amis et al. (1982) examined the ratio and separate regression Cardenas estimators, and compared them with the Huddleston-Ray estimator. Cardenas et al. (1978) stated that none of the estimators they presented were shown to be "best" in any sense, nor did they demonstrate any optimum properties. They did show that each of these estimators, when summed over counties, provides an unbiased stratum level estimate for the State. Also, assuming that the within county variance is the same for all counties, the method enables unbiased estimation of the State-wide variance. Amis et al. (1982) emphasized that an unbiased estimate of the county mean crop area per segment may not be possible when there are few sample segments in a county. Whenever there are significant differences in county variances, the Cardenas estimators appear to have higher variances than the Huddleston-Ray estimator. Amis et al. (1982) concluded that there appears to be no difference between the Cardenas ratio estimator and the separate regression estimator, and that the Cardenas estimators do not perform better than the Huddleston-Ray estimator. Both Cardenas estimators studied appeared to be biased, with larger variances than the Huddleston-Ray estimator.

The Cardenas method was never used in an operational remote sensing program since it did not provide sufficient improvement over the Huddleston-Ray estimator. The Huddleston-Ray estimator was used to generate county estimates for use by the NASS State Statistical Offices (SSO's) until 1982. At that time, Walker and Sigman (1982) advised that calculation of county estimates using the Huddleston-Ray method be discontinued, and that the Battese-Fuller method be used instead.

Walker and Sigman (1982) studied the Battese-Fuller model using Landsat MSS data over a six county region in eastern South Dakota. They found a modest lack of fit of the model, with larger model departure corresponding to low correlation between classified pixel counts and ground survey observations. A key feature of the Battese-Fuller model is the county effect parameter and this effect was found to be highly significant for corn, the most prevalent of the four crops considered in the study. Furthermore, this effect manifested itself within several strata but was negligible across strata. The study nonetheless indicated 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

Table 4: County Estimates for Mississippi 1991

County	Official	Computed	% Diff[a]	CV
Cotton	acres (000)	acres (000)	percent	percent
Bolivar	65.5	61.6	6.0	9.9
Coahoma	105.7	88.3	16.5	4.8
Humphreys	61.6	57.3	7.0	5.9
Issaquena	38.0	34.6	9.0	11.3
Leflore	79.2	87.8	10.9	4.0
Quitman	31.0	46.4	49.7	8.6
Sharkey	47.0	48.6	3.4	7.0
Sunflower	100.0	79.3	20.7	6.9
Tallahatchie	64.2	67.9	5.8	7.2
Tunica	45.6	38.0	16.7	6.6
Washington	95.7	102.4	7.0	3.9
Yazoo	94.5	93.9	.6	8.0
Total	828.0	806.1		
Rice				
Bolivar	74.0	66.2	10.5	5.4
Coahoma	15.8	10.4	34.2	24.0
Humphreys	3.6	7.1	97.2	32.4
Leflore	16.6	19.4	16.9	18.6
Sharkey	5.0	7.8	56.0	21.8
Sunflower	36.0	37.8	5.0	9.3
Tallahatchie	9.6	8.5	11.5	35.3
Tunica	17.5	9.9	43.4	26.3
Washington	30.5	22.6	25.9	15.5
Total	208.6	189.7		

[a] $\%Diff. = \frac{|\,Official - Computed\,|}{Official} * 100$

than 20 percent of the estimate was attributable to root mean square error. These members were the estimators 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 replacing the Huddleston-Ray estimator with the Battese-Fuller family.

The reliability of the county estimates based on the Battese-Fuller model has been closely watched since its implementation in 1982. As mentioned previously, these estimates are only one of possibly four or more indications that are composited to provide the final published crop acreage values. The reliability of the Battese-Fuller estimates can vary between years, between crops, between counties and between States depending on the stage of the crop at the time of the Landsat imagery, the amount of crop acreage within the county, the number of segments within the county, and cloud cover. The results presented in Tables 2 and 3 for corn are relatively good with all CVs less than 5 percent and over half of the percentage differences from the published value less than 4 percent. The soybean results are slightly poorer, with CVs ranging from 4 percent to 11 percent and percentage differences ranging up to 18 percent. Table 4 presents more recent results for a set of counties covered by Landsat in Mississippi for 1991. A review of the CVs and percentage differences indicate that the Battese-Fuller estimates can have relatively large CVs and percentage differences when the county crop acreage is less than 30,000 acres. Some summary statistics of the differences for the four crop examples discussed are presented in Table 5. The mean average difference is typically less than 10,000 acres, but

Table 5: Summary Statistics on Accuracy of Battese-Fuller Estimates (1000 acres)

Crop/State/Year	MD*	RMSD*	MAD*	LAD*
Corn Iowa 1988	-0.6	6.8	5.4	14.3
Soybeans Iowa 1988	-8.6	11.9	9.1	25.5
Cotton Mississippi 1991	-1.8	10.0	7.8	20.7
Rice Mississippi 1991	-2.1	5.2	4.5	7.9

* MD = mean difference between Battese-Fuller and published value

RMSD = root mean squared deviation

MAD = mean absolute difference

LAD = largest absolute difference

for small county acreages such as rice in Mississippi, large percentage differences may still occur. Consequently, NASS SSO's still use additional survey and administrative data to help set the published values.

6.6 Current Problems and Activities

As technology improves, new sensors produce satellite data that can be more accurately classified to a given crop than ever before. Consequently, the overall count of pixels classified to a given crop within a county can possibly be used directly to estimate crop acreage. The overall pixel count represents a census of pixels covering the county and therefore is not subject to sampling error. However, a nonsampling error is introduced due to inaccuracies in the classification. 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

Two adjustment terms that are being investigated are:

1) a conversion factor representing the area on the ground corresponding to one pixel for the specific sensor used, and

2) a combined ratio of the form:

$$\frac{\sum_h N_h \bar{y}_{h..}}{\sum_h N_h \bar{x}_{h..}}$$

where:

N_h = number of segments in stratum h

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

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

Both adjustment terms are conceptually simple. The combined ratio uses stratum level survey information to compute the adjustment term that may provide a more accurate conversion of pixel counts to crop area than the set conversion factor. Also, the ratio has a readily available formula for estimating the variance.

Research continues to focus on identifying new geographic areas and crops where this estimator would be applicable. Also, possible benefits of remotely sensed data from alternative sources, such as radar satellites, will be investigated as the newer sources are available. In recent years TM sensor data have been used to produce county estimates in the Delta region. County estimates of rice, cotton, and soybeans were produced for Arkansas and Mississippi in 1991, with Louisiana added in 1992. In 1993 satellite data are only being used in Arkansas due to budgetary constraints. To date, the satellite based estimates have only been produced on a limited scale. The NASS SSO's continue to rely on other data series for helping set the published county estimates of crop acreages. They conduct a large non-probability county estimates survey (see Chapter 7) that serves a dual purpose of also providing updated control data for the list sampling frame. This is an integral part of the NASS survey program and so will continue in some form for the foreseeable future. Fairly reliable administrative data sources are also available. NASS is continuing to investigate the benefits of satellite based county estimates in relation to these other available data sources. One by-product of the satellite data process that is attractive to the State offices is color coded land use maps at the county level. These maps provide a pictorial view of the distribution of the crops within each county. Identifying alternative uses of satellite data such as this is an important research objective of NASS.

REFERENCES

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

Amis, M.L., Martin, M.V., McGuire, W.G., and Shen, S.S. (1982), "Evaluation of Small Area Crop Estimation Techniques Using Landsat and Ground-Derived Data," LEMSCO-17597, Houston, TX: Lockhead Engineering and Management Services Company, Inc.

Angelici, G., Slye, R., Ozga, M., and Ritter, P. (1986), "PEDITOR - A Portable Image Processing System," Proceedings of the IGARSS '86 Symposium, Zurich, Switzerland, pp. 265-269.

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

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

Bellow, M.E. and Graham, M.L. (1992), "Improved Crop Area Estimation in the Mississippi Delta Region using Landsat TM Data," Proceedings of the ASPRS/ACSM Convention, Washington, D.C., pp. 423-432.

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

Cochran, W.G. (1977), "Sampling Techniques," New York, N.Y.: John Wiley & Sons.

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

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

Ozga, M. (1985), "USDA/SRS Software of Landsat MSS-Based Crop Acreage Estimation," Proceedings of the IGARSS '85 Symposium, Amherst, MA, pp. 762-772.

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

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

Appendix A: Estimators of Battese-Fuller Variance Components

The estimators of the Battese-Fuller variance components, σ^2_{vh} and σ^2_{eh}, at the analysis district level are expressed as:

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

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

where:

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

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

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

n_h = number of sample segments in stratum h

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

C = number of counties in analysis district

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

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

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

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

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

$\hat{\beta}_{oh}, \hat{\beta}_{1h}$ = least squares regression parameters in stratum h for regression of y_{hci} on x_{hci} .

Appendix B: Huddleston-Ray Estimator

The Huddleston-Ray estimator 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 frame unit. Within the analysis district, the overall mean crop area in regression stratum h is estimated by:

$$\bar{y}_{(REG),h..} = \bar{y}_{h..} + \hat{\beta}_h(\bar{X}_{h..} - \bar{x}_{h..})$$

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

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

where:

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

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

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

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

$\hat{\beta}_h$ = least squares regression slope parameter estimate in stratum h for regression of y_{hci} on x_{hci} .

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

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

where:

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

H_r = number of regression strata.

Walker and Sigman (1982) pointed out two problems with the Huddleston-Ray estimator. They stated that the variance calculation was unclear, and use of the difference:

$$\bar{X}_{hc.}-\bar{x}_{h..}=(\bar{X}_{hc.}-\bar{x}_{hc.})+(\bar{x}_{hc.}-\bar{x}_{h..})$$

combined within-county sampling error with the county analysis district difference. However, Amis et al. (1982) gave the following estimator for the variance:

$$V\,(\hat{T}_{(HR),c}) = \sum_{h=1}^{H_r} N^2{}_{hc}\ \frac{N_h-n_h}{n_h}\ s_h^2\ \frac{n_h-1}{n_h-2}(1-r_h^2)T_{hc}$$

where:

N_h = number of segments in stratum h

n_h = number of sample segments in stratum h

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

s_h^2 = sample variance for the reported area in stratum h

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

$r_h{}^2$ = sample coefficient of determination for stratum h

C = number of counties in analysis district

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

Appendix C: Cardenas Family of Estimators

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

$$\bar{y}_{(CAR),hc.} = \bar{y}_{h..} + \hat{\beta}_h(\bar{X}_{hc.} - \bar{X}_{h..})$$

where:

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

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

(See Appendix B for definitions of all terms not defined here). The estimate of total crop area in the regression strata of county c is:

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

The parameter $\hat{\beta}_h$ relates classified pixel counts to reported crop area. Cardenas et al. (1978) proposed three alternative estimators for $\hat{\beta}_h$:

(1) Ratio estimator

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

(2) Separate regression estimator

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

(3) Combined regression estimator

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

The combined regression estimator is applicable only when the $\hat{\beta}_h$'s are assumed to be constant over strata. Cardenas et al. (1978) also provided formulas for estimating the variance of their county estimators.