LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)

NASA   NOAA   USDA

EFFECTS OF NON-RESPONSE INCLUDING CLOUD COVER

ON AGGREGATION OF WHEAT AREA

IN THE U.S. GREAT PLAINS
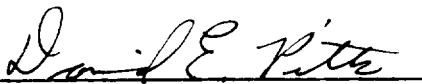
LARGE AREA CROP INVENTORY EXPERIMENT (LACIE)

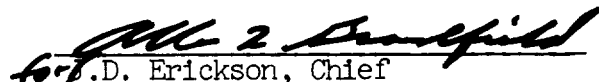ACCURACY ASSESSMENT UNSCHEDULED REPORT

Effects of Non-Response Including Cloud Cover on Aggregation of
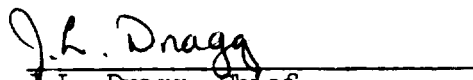Wheat Area in the U.S. Great Plains

PREPARED BY:
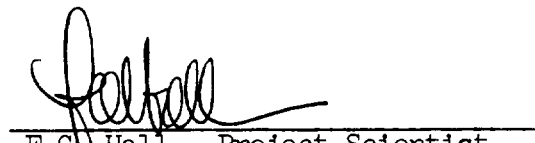
A. Glen Houston
A. Glen Houston

APPROVED BY:
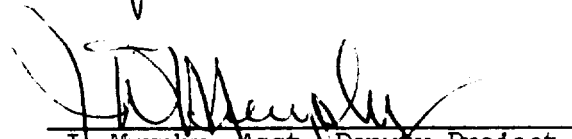
D.E. Pitts
D.E. Pitts, Manager
Accuracy Assessment

J.L. Dragg
J.L. Dragg, Chief
Applications Systems
 Verification Branch

J. Hill
J. Hill, Asst. Deputy Project
Manager, NOAA

R. B. Erb
R. B. Erb, Chief
Project Office
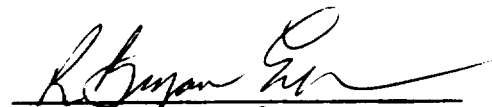
for J.D. Erickson
.D. Erickson, Chief
Research, Test, & Evaluation

F.G. Hall
F.G. Hall, Project Scientist

J. Murphy
J. Murphy, Asst. Deputy Project
Manager, USDA
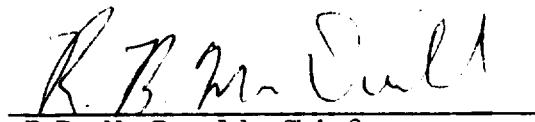
R.B. MacDonald
R.B. MacDonald, Chief
Earth Observations Division

Table of Contents

List of Figures

List of Tables

1.0 Introduction and Summary

In any sample survey from a countable population, potential for bias is introduced when certain elements of the population have probability zero of being selected. In many surveys, however, it is desirable to preclude the selection of certain areas of the population for sampling efficiency even though some bias is introduced. For example, excluding the state of Nevada improves the efficiency of the sampling design for a wheat survey, by not allowing samples to be taken from sparse wheat areas, and results in negligible bias. When the cost per sample is high or constraints are put on the number of samples that may be taken, it is desirable to exclude, from the population to be sampled, those areas which contain little or none of the information being sought. This process is known as 'screening' the population. In this case, for those areas containing small amounts of information, estimates based on historic data may be made with relatively small error contribution to the final population estimate.

The phenomena of "non-response" may also introduce bias into a sample survey. Generally, "non-response" refers to the unplanned or inadvertant failure to measure the response variable for certain members of the sample. In a mail survey, non-respondents would bias the sample if the cause of the non-response was related to the quantity being surveyed, e.g., farmers with a poor crop may be less likely to respond than farmers with a good crop.

In LACIE, non-response results from four causes:

a. The sample segment being obscured by cloud cover.

b. Landsat data quality being insufficient to permit processing.

c. Landsat data acquisition failing to register with the reference Landsat image.

d. Failure of acquisition/processing procedures to provide an acceptable estimate.

At the outset of LACIE, the plausible assumption was made that no causal relationship between these non-response phenomena and wheat acreage existed, i.e., the probability of non-response of a sampled segment within a stratum (county) was assumed to be uncorrelated with the wheat acreage of the stratum. Data has been monitored to verify this assumption, and the results corroborate that no significant correlation exists.

Empirical and theoretical investigations conducted in LACIE have indicated that the bias resulting from both the loss of segments to non-response and the nonexistence of segments in the sparse wheat regions is negligibly small. Therefore, these experiments verify that the intent of the LACIE sample design, i.e., to sample efficiently and cost effectively with negligible bias, is being satisfied. It is shown in this report that if, within a stratum, the probability of non-response of a sampled segment is not correlated with the wheat acreage of the sampled segment and there is at least one segment in the stratum not lost to non-response, then the resulting estimate of the wheat acreage for the stratum is not biased by non-response.

The equation of the bias resulting from the use of historic data to estimate the wheat acreage of certain wheat regions is also presented. In LACIE, those strata for which historic data are used to estimate the wheat acreage are called Group III strata. The Group III strata are made up of those nonsampled strata with sparse wheat acreage, and the resulting wheat acreage estimate for these strata is known as the Group III ratio estimate. However, the Group III ratio estimate is also applied to those strata which have lost all their sampled segments by non-response. A series of analyses has been conducted to determine (1) the magnitude of the bias resulting from using historic data to estimate wheat acreage for the nonsampled sparse wheat areas and (2) the magnitude of the bias resulting from applying the Group III ratio estimator to strata whose sampled segments have been lost to non-response. The results indicate that these biases are small.

2.0 Approach

In section 3.0, formulas for the biases due to nonresponse and Group III ratio estimation are presented. These formulas are developed for the strata level and the region (CRD, state, etc.) level. As pointed out earlier, it is seen that if, within a stratum, the probability of nonresponse of a sampled segment is not correlated with the wheat acreage of the segment and there is at least one segment in the stratum not lost to nonresponse, then the wheat acreage estimate for the stratum is not biased by nonresponse.

Data were obtained from 40 blind sites[1] randomly selected in the southern Great Plains [2] to determine the correlation between the probability of not acquiring a sample segment and the wheat acreage of the segment. The probability of not acquiring a sample segment was estimated empirically for each of the blind sites and was based on actual opportunities and acquisitions by Landsat 2 during LACIE Phase II. The wheat acreage for each of the blind sites was determined from ground truth data and photographs taken from NASA aircraft during the LACIE Phase II crop year. The correlation between the probability of not acquiring a sample segment and the wheat acreage in the segment was found to be negligible.

[1] A blind site is a regular LACIE segment selected by a stratified random sample to be photographed and ground-truthed over the entire 30 square nautical mile area.

[2] The southern Great Plains is defined as the 5 states: Colorado, Kansas, Nebraska, Oklahoma and Texas.

Therefore, by the result in section 3.0, it is concluded that non-response does not bias the wheat acreage estimate for those strata which have at least one segment not lost to nonresponse. A detailed description of this analysis is given in section 4.0.

In order to determine the magnitudes of the bias due to the use of historic data for estimating wheat acreage in the nonsampled sparse wheat areas and the bias due to the use of historic data for estimating wheat acreage in these sampled strata whose allocated segments were lost to nonresponse, selected LACIE segments were deleted from the data base and the resulting estimates were compared to estimates made with a complete data base. In crop year 1974-75, a particular set of LACIE segments in the U.S. Great Plains[3] were not processed. In crop year 1975-76 a different set of LACIE segments were not processed. The effects of the loss of these segments using both crop year acquisition histories were examined and found to be insignificant. In these analyses, LACIE estimates of wheat proportions of allocated segments were replaced by SRS wheat proportions for the counties in which the segments were located (adjusted by the LACIE county agricultural area). This essentially removes the within county sampling and classification errors. The resulting state wheat acreage estimates were then compared to the SRS state estimates. In some cases, all allocated segments were used,

[3]The United States Great Plains is defined as the 9 states: Colorado, Kansas, Minnesota, Montana, Nebraska, North Dakota, Oklahoma, South Dakota, and Texas.

and in other cases, only processed segments were used.

Any discrepancy between a state estimate and the SRS state estimate may

be attributed to either (a) the error due to not allocating segments to

all counties, i.e., error due to using historic data for estimating

the wheat acreage of Group III counties, and error due to the use of

PPS sampling for the Group II counties[4] or (b) in addition to the error

in (a), the error induced by the nonresponse of some of the allocated

segments, i.e., error due to applying the Group III ratio estimate

to those counties whose allocated segments were lost to nonresponse.

These analyses are summarized in the following tables:

Table I: Analysis based on those segments processed (271) during

LACIE Phase I (1975) using 1974 (SRS) planted wheat area.

Table II: Analysis based on those segments processed (271) during

LACIE Phase I (1975) using 1975 SRS harvested wheat area.

Table III: Analysis based on those segments allocated (431) during

LACIE Phase II (1976) using 1975 SRS harvested wheat area.

Table IV: Analysis based on those segments processed (394) during

LACIE Phase II (1976) using 1975 SRS harvested wheat area.

Actual LACIE Phase I results are also presented in Table V. A

detailed description of the appropriate comparisons of the tables

is given in section 5.0. The main result of these comparisons is that

_____

[4]Group II counties are counties which have more than a trace of wheat
in them historically, but not enough to receive one sample segment.
Hence, the Group II counties within a CRD have been considered as one
county altogether. The number of sample segments for this group
has been determined by the sampling allocation scheme and the sample
segments have been assigned to the counties in the group with probability
proportional to size (PPS). Size here includes historical wheat density as
well as the number of agricultural acres of a particular county in the
group.

the relative difference between the SRS estimates and the mock aggregation estimates is less than one percent at the Great Plains level whether or not segments are lost to nonresponse. This indicates that the bias incurred by Group III ratio estimation and Group II estimation with and without nonresponse is negligibly small at this level.

3.0 Formulas for the biases due to "nonresponse" and Group II ratioing.

In this section, formulas for the biases due to nonresponse and Group III ratioing will be developed at the stratum (county) level. Bias formulas will also be presented for the region (CRD, state or Great Plains).

3.1 Formula for the bias at the stratum level.

Consider a stratum (a county or collection of Group II counties) containing N segments, i.e., N segments cover the stratum. Suppose that a sample of size n segments is to be selected from the N segments. Suppose further that, due to nonresponse, only m of the n segments are actually obtained, e.g., (n-m) segments may have been lost due to cloud cover.

Define the indicator variable for this stratum:

$$v = \begin{cases} 1, & \text{if } m > 0 \\ 0, & \text{if } m = o \end{cases}$$

Where m is the number of segments actually received in this stratum. Then the LACIE estimate, $\hat{Y}$, of the total wheat acreage, Y, in this stratum may be written as:

$$\hat{Y} = v\,\hat{Y}_{(m>o)} + (1-v)\,\hat{Y}_{(m=o)}$$

where $\hat{Y}_{(m>o)}$ is the estimate when at least one segment is acquired in the stratum and $\hat{Y}_{(m=o)}$ is the estimate when no segment is acquired in the stratum. Note that $\hat{Y}_{(m=o)}$ is the estimate for the so-called

9

Group III counties, which have $v=o$ with probability one, as
determined by the sampling allocation. For the other counties that
have been allocated a segment (Group I or Group II counties) let

$Pr[v=1] = \gamma$ = probability that this stratum contains a segment
and the segment is not lost to nonresponse. Then the expected value,
or mean, of Y is given by

$$E[\hat{Y}] = \gamma E[\hat{Y}| m>o] + (1-\gamma)E[\hat{Y}|m=o] \tag{1}$$

where the symbol, $|$, means 'given that'. Let us now consider the
two cases separately.

Case(1): Suppose that $m>o$ and let

$$\hat{Y}_1 = \hat{Y}_{(m>o)},$$

that is, let $\hat{Y}_1$ be the estimate of the total wheat acreage, Y, in
the stratum when at least one segment has been acquired in this stratum.
This stratum contains N segments of which n have been selected for
classification but only m of these n have been acquired. The LACIE
estimate of the total wheat acreage of this stratum in this case is
given by
$$\hat{Y}_1 = \frac{N}{m} \sum_{j=1}^{m} y_j \tag{2}$$

where y_j is the wheat acreage of the jth sample segment in this stratum.
(This study is not addressing classification error, so y_j is assumed
to be the true value, i.e., no classification error). Note that

$$Y = \sum_{j=1}^{N} y_j .$$

Define the indicator variables:

$$a_j = \begin{cases} 1, & \text{if segment } j \text{ is selected to be in the sample} \\ 0, & \text{if not} \end{cases}$$

and

$$u_j = \begin{cases} 1, & \text{if } a_j = 1 \text{ and segment } j \text{ is not lost to nonresponse} \\ 0, & \text{if } a_j = 0 \text{ or, } a_j = 1 \text{ and segment } j \text{ is lost to nonresponse} \end{cases}$$

Then (2) may be rewritten as:

$$\hat{Y}_1 = \frac{N}{m} \sum_{j=1}^{N} a_j u_j y_j \tag{3}$$

Let $\Pr[a_j = 1] = p_j$ (For Group I counties, $p_j = \frac{n}{N}$. For a collection of Group II counties, p_j is the probability of selecting a particular county in the collection times the probability of selecting a segment in that county).

Also, let $\Pr[u_j = 1 \mid a_j = 1] = \pi_j$. With these definitions, the following joint probability table for a_j and u_j may be constructed.

$$
\begin{array}{c|cc|c}
 & 0 & 1 & \\
\hline
0 & 1-p_j & p_j(1-\pi_j) & 1-p_j\pi_j \\
\hline
1 & 0 & p_j\,\pi_j & p_j\,\pi_j \\
\hline
 & 1-p_j & p_j & 1
\end{array}
$$

(with u_j labeling the rows and a_j labeling the columns)

From the table, we see that the probability that an allocated segment is acquired, i.e., not lost to nonresponse, is $p_j\,\pi_j$. Since m of the segments have been acquired, it is clear that

$$
m = \sum_{j=1}^{N} a_j u_j
$$

Therefore,

$$
m = E(m) = E\left[\sum_{j=1}^{N} a_j u_j\right] = \sum_{j=1}^{N} E(a_j u_j) = \sum_{j=1}^{N} p_j\,\pi_j \qquad (4)
$$

Taking the expected value of (3),

$$
E[\hat{Y}_1] = E\left[\frac{N}{m}\sum_{j=1}^{N} a_j u_j y_j\right]
$$

$$
= \frac{N}{m}\sum_{j=1}^{N} y_j\, E[a_j u_j]
$$

$$
= \frac{N}{m}\sum_{j=1}^{N} y_j p_j\,\pi_j
$$

$$= \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right) + \frac{N}{m} \sum_{j=1}^{N} p_j \pi_j y_j - \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right)$$

$$- \sum_{j=1}^{N} y_{\cdot j} + \sum_{j=1}^{N} y_j$$

$$= \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right) + \frac{N}{m}\left[\sum_{j=1}^{N} p_j \pi_j\, y_j - \overline{Y}\sum_{j=1}^{N} p_j \pi_j - \frac{m}{N}\sum_{j=1}^{N} y_j + m\,\overline{Y}\right],$$

$$\text{where } \overline{Y} = \frac{1}{N}\sum_{j=1}^{N} y_j$$

$$= \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right) + \frac{N}{m}\sum_{j=1}^{N}\left[p_j \pi_j y_j - \overline{Y} p_j \pi_j - \frac{m}{N} y_j + \frac{m}{N}\overline{Y}\right]$$

$$= \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right) + \frac{N}{m}\sum_{j=1}^{N}\left[(y_j - \overline{Y})(p_j \pi_j - \frac{m}{N})\right]$$

$$= \frac{1}{m}\left(\sum_{j=1}^{N} y_j\right)\left(\sum_{j=1}^{N} p_j \pi_j\right) + \frac{N^2}{m}\left[\frac{1}{N}\sum_{j=1}^{N}(y_j - \overline{Y})(p_j \pi_j - \frac{m}{N})\right]$$

From (4), $\sum_{j=1}^{N} p_j \pi_j = m$, so that

$$E(\widehat{Y}_1) = \sum_{j=1}^{N} y_j + \frac{N^2}{m} \text{Cov}(y_j, p_j \pi_j).$$

$$= Y + \frac{N^2}{m} \text{Cov}(y_j, p_j \pi_j),$$

where $\text{Cov}(y_j, p_j \pi_j) = \frac{1}{N} \sum_{j=1}^{N} (y_j - \overline{Y})(p_j \pi_j - \frac{m}{N}).$

Therefore, $\text{Bias}(\hat{Y}_1) = E[\hat{Y}_1] - Y = \frac{N^2}{m} \text{Cov}(y_j, p_j \pi_j)$ \hfill (5)

So, the bias of the estimate of the total wheat acreage in a particular stratum, for which at least one segment has been acquired, is a function of the correlation between the wheat acreage in the segment and the probability of losing the segment to nonresponse. If this correlation is zero, of course, the estimate is unbiased. A study has been conducted which determined this correlation to be negligible. The details of this study are presented in section 4.0.

Case (ii): Suppose that m = 0 and let $\hat{Y}_2 = \hat{Y}_{(m=0)}$, that is, let $\hat{Y}_2$ be the estimate of the total wheat acreage, Y, in the stratum when no segments have been acquired in this stratum. As pointed out earlier, $\hat{Y}_2$ is the estimate for the Group III strata (counties) and is called the Group III ratio estimate. In this case, the LACIE estimate of the total wheat acreage in this stratum is

$$\hat{Y}_2 = \frac{Y_{69}}{X_{69}} \hat{X}_{current}$$

where

Y_{69} = 1969 total wheat acreage of this stratum,

X_{69} = 1969 total wheat acreage of the Group I and Group II strata for which segments have been acquired currently and are in the same region (CRD or state) as this stratum,

and $\hat{X}_{current}$ = current estimates of the total wheat acreage of these Group I and Group II strata.

Note that the true wheat acreage, Y, in the stratum may be written as

$$Y = \frac{Y_{current}}{X_{current}} \cdot X_{current}$$

where

$$Y_{current} = \text{current true total wheat acreage of this stratum}$$

and

$$X_{current} = \text{current true total wheat acreage of the above-}$$

mentioned Group I and Group II strata.

For convenience, let "current" = 76. Then,

$$\hat{Y}_2 - Y = \frac{Y_{69}}{X_{69}} \hat{X}_{76} - \frac{Y_{76}}{X_{76}} X_{76}$$

$$= R_{69} \hat{X}_{76} - R_{76} X_{76}$$

where R_{69} and R_{76} are fixed ratios. Taking the expected value,

$$\text{Bias } (\hat{Y}_2) = E[\hat{Y}_2 - Y]$$

$$= E[R_{69} \hat{X}_{76}] - E[R_{76} X_{76}]$$

$$= R_{69} E[\hat{X}_{76}] - R_{76} X_{76}$$

Now, $\hat{X}_{76}$ is the estimate of the total wheat acreage of the Group I and Group II strata for which segments have been acquired (in 1976). In Case (i) above, it has been established that this estimate is unbiased if the correlation between the wheat acreage in a segment and the probability of losing the segment to nonresponse is zero. Since a study has determined this correlation to be negligible (see section 4.0), it is assumed that

$$E[\hat{X}_{76}] = X_{76}.$$

Hence, Bias $(\hat{Y}_2) = (R_{69} - R_{76}) X_{76}.$ \hfill (6)

Clearly, if $R_{69} = R_{76}$, i.e., the ratio of the wheat acreage in the stratum to the wheat acreage in the Group I and Group II strata is the same for the years 1969 and 1976, this bias is zero.

Recall from (1) that

$$E[\hat{Y}] = \gamma \, E[\hat{Y} \mid m > o] + (1-\gamma)E[\hat{Y} \mid m = o]$$

where γ is the probability that the stratum contains a segment and the segment is not lost to non-response. For the Group III counties, those counties not allocated segments,

$$\gamma = \Pr[v=1] = 0$$

and $1 - \gamma = \Pr(v=o] = 1.$

Hence, by Case ii, the estimate of the total wheat acreage for a particular Group III county is given by

$$\hat{Y} = \hat{Y}_{(m=o)} = \hat{Y}_2$$
$$= \frac{Y_{76}}{X_{69}} \hat{X}_{76}$$

with bias

$$\text{Bias} (\hat{Y}) = (R_{69} - R_{76}) \, X_{76}$$

From Case (i), for the Group I and Group II strata,

$$\gamma = \Pr[v=1) \neq 0.$$

Hence, the estimate of the total wheat acreage for a particular Group I county or a particular collection of Group II counties is given by

$$\hat{Y} = v\hat{Y}_{(m>o)} + (1-v)\hat{Y}_{(m=o)}$$

$$= v\hat{Y}_1 + (1-v)\hat{Y}_2$$

$$= v \frac{N}{m} \sum_{j=1}^{N} a_j u_j y_j + (1-v) \frac{Y_{69}}{X_{69}} \hat{X}_{76}$$

with mean

$$E[\hat{Y}] = \gamma\, E[Y \mid m>o] + (1-\gamma)E[\hat{Y} \mid m=o]$$

$$= \gamma\, E[\hat{Y}_1] + (1-\gamma)E[\hat{Y}_2]$$

$$= \gamma[Y + \text{Bias}\,(\hat{Y}_1)] + (1-\gamma)[Y + \text{Bias}\,(\hat{Y}_2)]$$

where Bias $(\hat{Y}_1)$ is given by (5) and Bias $(\hat{Y}_2)$ is given by (6). This may be rewritten as

$$E[\hat{Y}] = Y + \gamma\, \text{Bias}\,(\hat{Y}_1) + (1-\gamma)\, \text{Bias}\,(\hat{Y}_2) \qquad (7)$$

so that

$$\text{Bias}\,(\hat{Y}) = \gamma\, \text{Bias}(\hat{Y}_1) + (1-\gamma)\, \text{Bias}\,(\hat{Y}_2)$$

$$= \gamma \cdot \frac{N^2}{m}\left[\frac{1}{N}\sum_{j=1}^{N}(y_j - \bar{Y})(p_j \pi_j - \frac{m}{N})\right] + (1-\gamma)\left[\frac{Y_{69}}{X_{69}} - \frac{Y_{76}}{X_{76}}\right]X_{76}$$

3.2 Formula for the bias at the region level.

Suppose there are M_k strata (counties) in the kth region. (If the region is a state, k goes from 1 to 9 in the U.S. Great Plains).

The estimate of the total wheat acreage for the ith stratum in this region will be denoted by $\hat{Y}_i$ ($i=1,2,\ldots,M_k$), and the true total wheat acreage for this stratum will be denoted by Y_i. We will also suppose that the ith stratum contains N_i segmtments, n_i segments are to be selected at random from the N_i segments, and only m_i segments of the n_i segments are actually acquired, the others being lost to non response. Defining, as before,

$$v_i = \begin{cases} 1, & \text{if } m_i > 0 \\ 0, & \text{if } m_i = 0 \end{cases}$$

the LACIE estimate, $\hat{Y}_K$ of the total wheat acreage in the kth region, Y_K, is given by

$$\hat{Y}_K = \sum_{i=1}^{M_k} \left[v_i \, \hat{Y}_{i\,(m>0)} + (1-v_i) \, \hat{Y}_{i(m=0)} \right]$$

where $\hat{Y}_{i(m>0)}$ and $\hat{Y}_{i(m=0)}$ are the same as in 3.1 but now they apply to the ith stratum.

Using the results in 3.1 and including the subscript i where necessary (results in 3.1 still hold, the subscript i just indicates that the results may be different for different strata),

$$\text{Bias}(\hat{Y}_K) = \sum_{i=1}^{M_k} \left\{ \gamma_i \, \frac{N_i^2}{m_i} \left[\frac{1}{N_i} \sum_{j=1}^{N_i} (y_{ij} - \overline{Y}_i)(p_{ij}\pi_{ij} - \frac{m_i}{N_i}) \right] \right.$$
$$\left. + (1-\gamma_i) \left[\frac{Y_{69_i}}{X_{69_i}} - \frac{Y_{76_i}}{X_{76_i}} \right] X_{76_i} \right\} \qquad (8)$$

The subscripting should be obvious, e.g., y_{ij} represents the true wheat acreage of the jth segment in the ith stratum of this region.

Let's suppose that the region is a state. Then the term in (8) multiplied by $(1-\gamma_i)$ is the bias due to the Group III ratioing in the state. The term in (8) multiplied by γ_i is the bias due to the loss of some of the allocated segments to non-response. For a particular year, an unbiased estimate of Bias $(\hat{Y}_k)$ may be obtained by replacing γ_i with v_i, where $v_i = o$ if the ith county has no segment that year and $v_i = 1$ if the ith county has at least one seg-ment that year. The results of a study doing just that are pre-sented in section 5.0, where the y_{ij} are taken as the SRS county wheat acreages. When all allocated segments are used, the resulting bias is the bias due to the Group III ratioing or Group II estimation of those counties not allocated segments. When only acquired segments are used, the resulting bias is due to the Group III ratioing of those counties whose allocated segments were lost to non-response in addition to the bias from Group II estimation and Group III ratioing of sparse wheat areas.

4.0 Empirical Correlation between Non-response and Wheat for LACIE Phase II.

In this study, 40 early season LACIE Phase II blind sites were selected at random from the 240 sample segments in the southern Great Plains states. These sites were photographed from aircraft and ground truthed within a few weeks of the photography. The probability of not acquiring these sites during LACIE Phase II was estimated by subtracting the number of times the segment was acquired from the number of acquisitions possible from Landsat II and dividing this by the number of acquisitions possible. The probability of not acquiring the segment was then plotted versus:

(1) the percent of wheat (planted) in the segment as determined by ground truth (Figure 1),

(2) 1975 SRS estimated wheat production of the county that the segment was located in (Figure 2),

(3) 1975 SRS estimated wheat density (percent) of the county that the segment was located in (Figure 3),

and (4) 1975 SRS estimated harvested wheat acreage of the county that the segment was located in (Figure 4).

A correlation coefficient was calculated between the probability of not acquiring a segment and each of the four wheat measurements and each was found to be less than 0.10. Hence, no significant correlations were found to exist. By the result in section 2.0,this implies that the event of nonresponse induces negligible bias to stratum (county) wheat acreage estimates if there is at least one segment in the stratum, i.e., the term in (8) multiplied by $\gamma_i \frac{N_i^2}{m_i}$ is essentially zero. This implies that the bias of the total wheat

acreage estimate for a region becomes

$$\text{Bias } (\hat{Y}_k) \approx \sum_{i=1}^{N_k} (1-\gamma_i) \left[\frac{Y_{69_i}}{X_{69_i}} - \frac{Y_{76_i}}{X_{76_i}} \right] X_{76_i} \qquad (9)$$

In the next section, a study has been conducted to determine the magnitude of the bias given in equation (9) when (1) the Group III ratioing is applied only to those counties that were not allocated segments, and (2), in addition to (1), the Group III ratioing is applied to those counties whose allocated segments were lost to nonresponse.
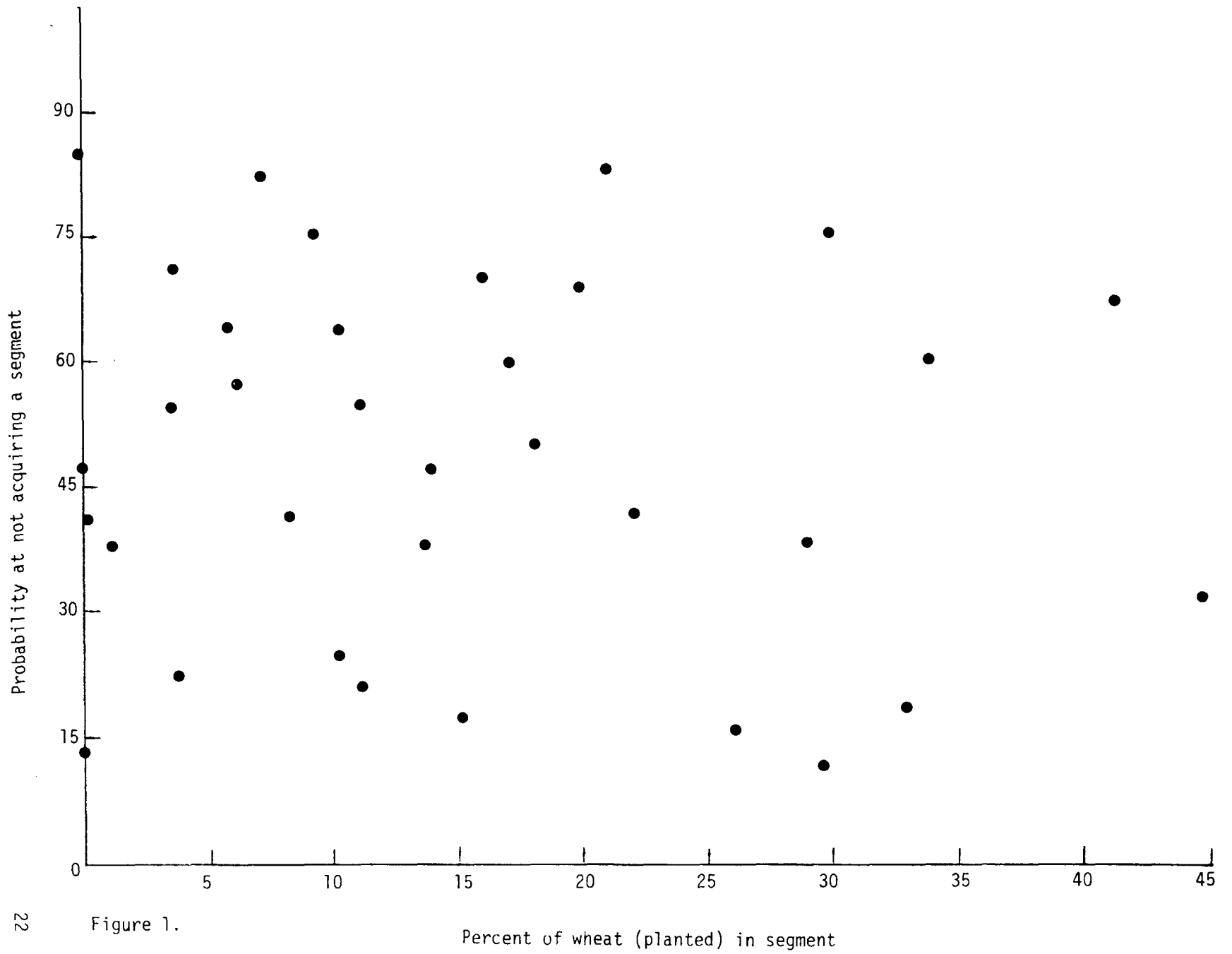
Figure 1.
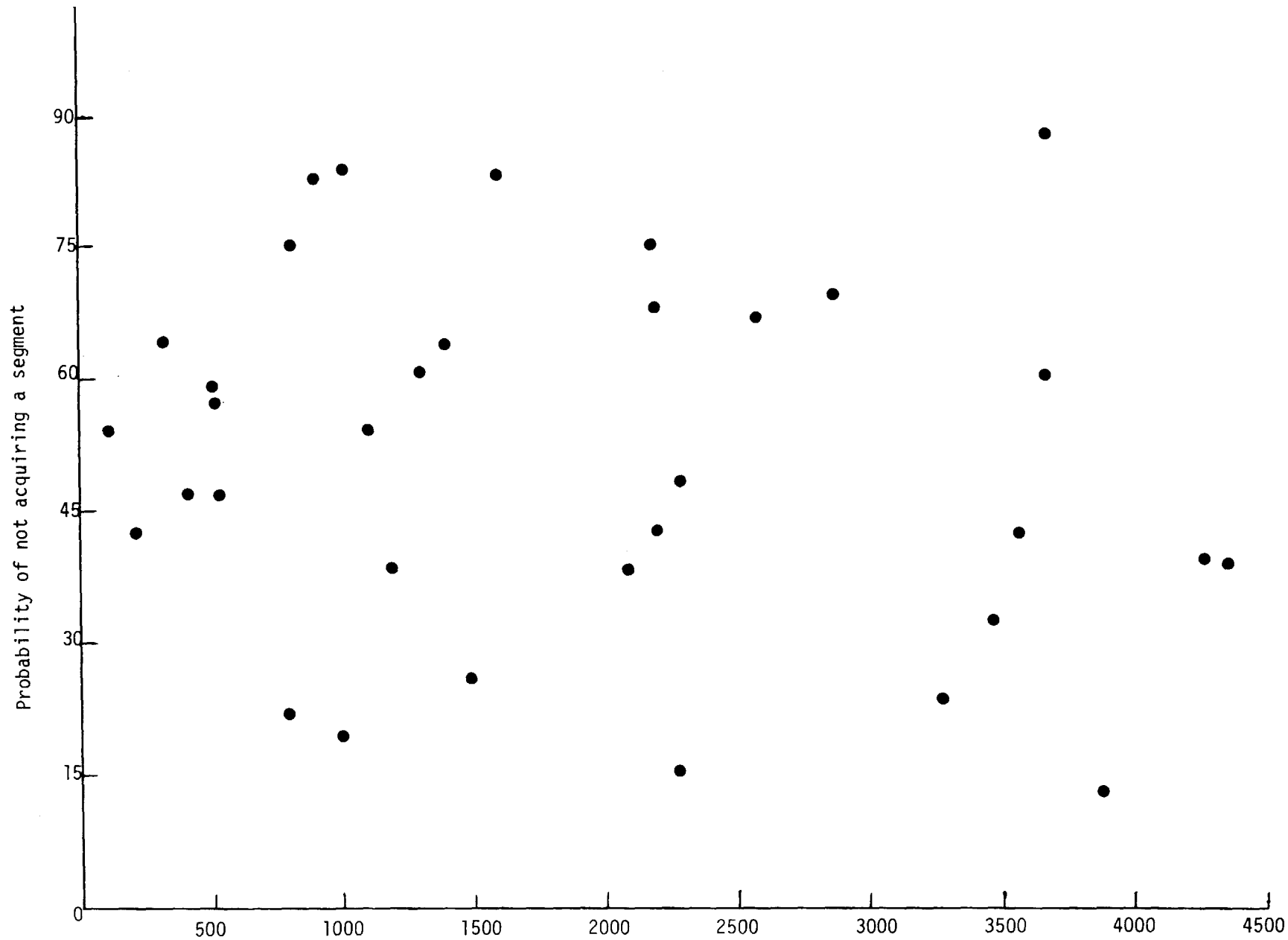
Percent of wheat (planted) in segment

Figure 2.   1975 SRS wheat Production (1000 BU) of the county containing the segment
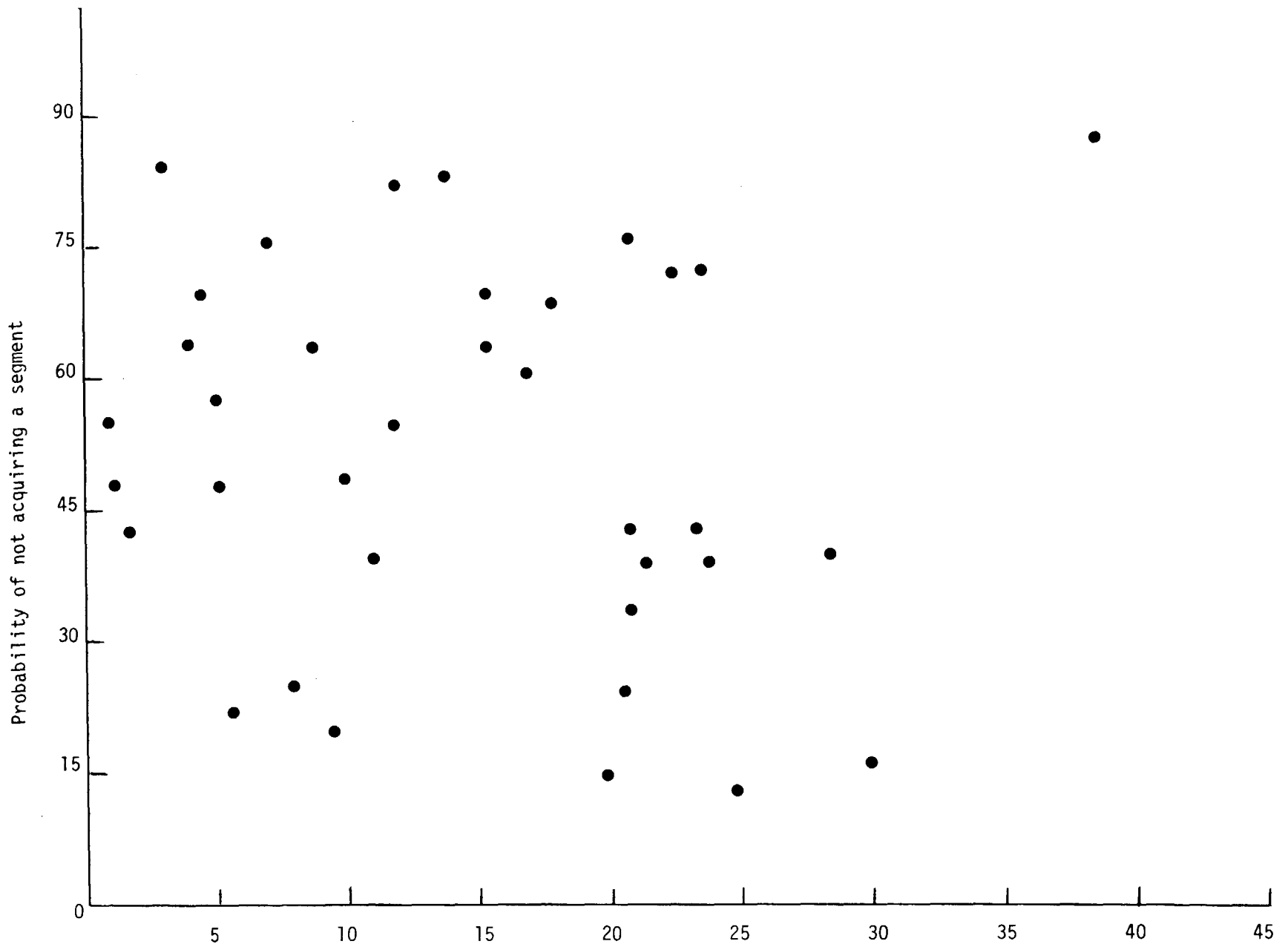
Figure 3.

1975 SRS wheat density (percent) of the county containing the segment
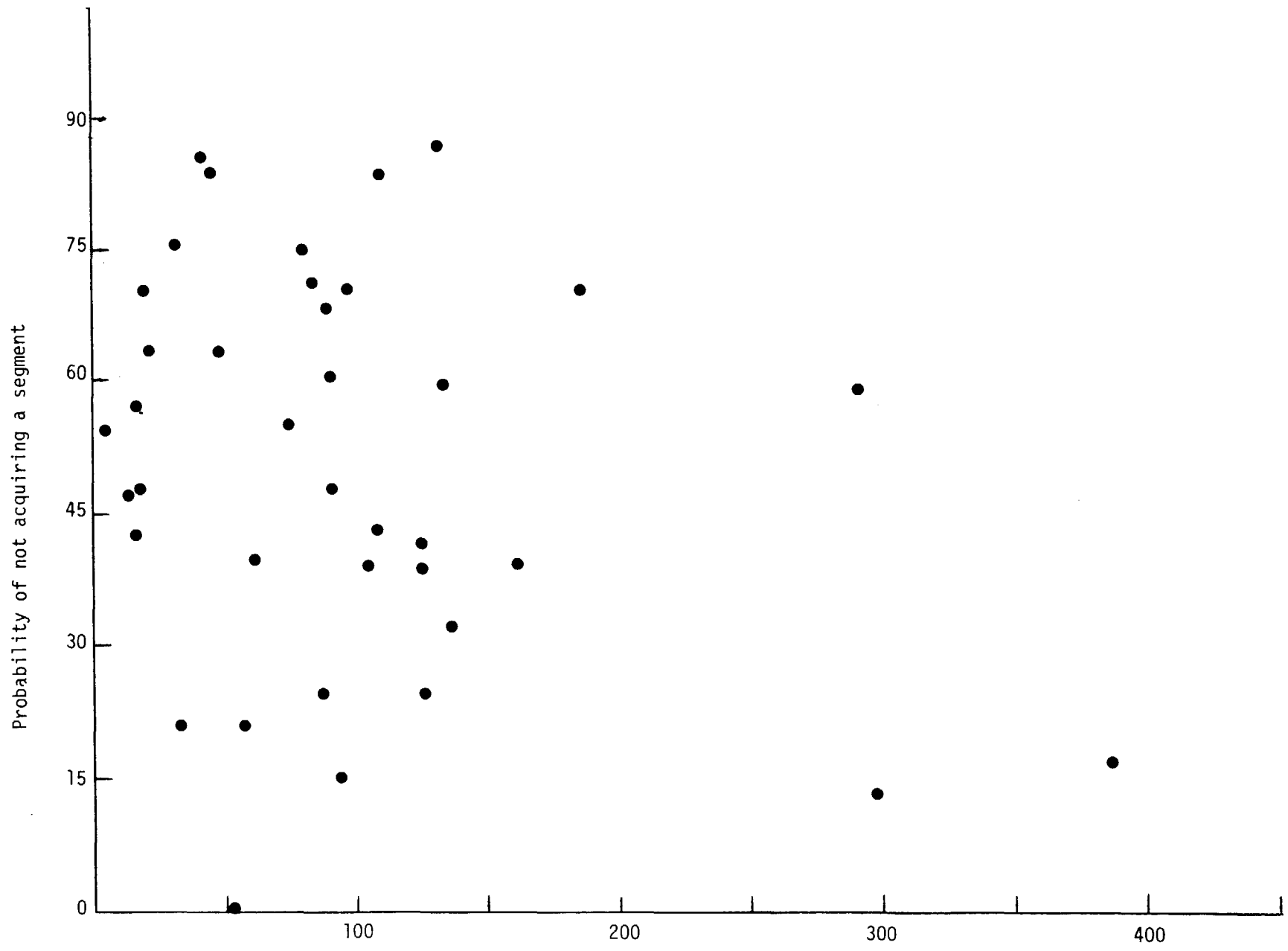
24

Figure 4   1975 SRS harvested wheat area (1000 acres) of the county containing the segment

5.0 Simulation Results

5.1 Previous Simulation Study - With and Without Nonresponse

An earlier study used the SRS estimate of planted wheat acreage for the 1973-74 crop year in each county contain- a ing a LACIE sample segment as a measure to determine wheat percent over the LACIE agricultural area in that county. In the study, two sets of aggregations were completed using the LACIE (CAS) software and substituting this wheat percent estimate in place of the estimate of wheat percent from the classification of Landsat data as determined by CAMS. The first aggregation was done for all LACIE Phase I segments (411), thus simulating the conditions for no nonresponse. The estimate for the U.S. Great Plains was 0.7 percent[3] lower than the SRS estimate. This is an estimate of the relative bias due to the Group III ratioing and Group II estimation of those counties not allocated segments. A second test was done in which only counties containing segments (271), which were acquired in LACIE Phase I (crop year 1974-75) with adequate acquisitions of biowindows to employ the Phase I rework procedure, were aggregated.

In this test the estimate was 0.8 percent higher than the SRS estimate at the Great Plains level. This is an estimate of the relative bias due to both the Group III ratioing and Group II estimation of counties not allocated segments and the Group III ratioing of those counties whose allocated segments were lost to nonresponse. Combining the results of the two aggregations yields an estimate of the relative bias due to the Group III ratioing of the counties whose allocated segments were lost to nonresponse to be approximately +1.5%.

---

[3]Relative difference, in percent, is defined as

$$\frac{\text{LACIE estimate} - \text{SRS estimate}}{\text{LACIE estimate}} \times 100$$

Rework of Previous Simulation Study - with Nonresponse.

Since the agricultural area determined by LACIE has recently been refined, it was decided to compare a current aggregation with the previous study. Table I displays the results of rerunning the original data base, i.e., 1974 SRS county data using the 1975 segments processed (271) through the current LACIE aggregation scheme. The results indicate only minor differences between this aggregation and the original study, this aggregation having a smaller relative difference (+0.07% compared to -0.7% from the original study). As pointed out in section 2.2, an estimate of the bias of the Great Plains level estimate due to Group III ratioing and Group II estimation of those counties not allocated segments and the Group III ratioing of those counties whose segments were lost to nonresponse is given by

LACIE-SRS = $(51,227 - 51,191)$ $(x10^3)$ acres for this aggregation.

This yields an estimate of the relative bias of $\frac{+36}{51,227}$ x 100 = +0.07%. This indicates that, for this aggregation, the error due to this Group III ratioing and Group II estimation is negligible (see below for explanation).

It should be pointed out here that the tabled c.v.'s do not represent the year-to-year variability of the historic ratio used in obtaining the Group III ratio estimates. Hence, these are not the c.v.'s to use to determine whether or not the corresponding tabled relative difference is significantly different from zero. The correct c.v.'s to use could be obtained by examining the true ratio (Group III county wheat proportion: Group I and II counties wheat proportion) for several years and calculating

the mean and variance of this distribution. This would have to be done for each CRD in each of the Great Plains states. Since this has not been done, only relative differences will be reported in the remainder of this section. For practical purposes, however, a relative difference less than, say, 2% at the Great Plains level will be considered negligible. At any rate, keep in mind that, for Tables I-IV, the c.v. corresponding to a particular relative difference is not the correct c.v. to use to determine the significance of the relative difference. For example, in Table I, as pointed out earlier, the relative difference at the Great Plains level is +0.07%. Practically (not statistically) speaking, at this level, this relative difference due to Group III ratioing and Group II estimation is negligible.

5.2   LACIE Phase I Simulation – With Nonresponse.

Table II shows the aggregation of county SRS estimates for crop year 1974-75 for those segments processed (271) during LACIE Phase I (1974-75). This aggregation differs from that in Table I in that different years SRS estimates were used and harvested wheat acreages were used rather than planted wheat acreages. Note that the relative difference at the Great Plains level is – 0.80% . As in Table I, this is an estimate of the relative bias due to Group II estimation and Group III ratioing of both the counties with no allocated segments and the counties whose allocated segments (140) were lost to nonresponse. As before, practically speaking, this relative difference is negligible.

5.3   LACIE Phase II Simulation – No Nonresponse.

Table III shows the aggregation of county SRS estimates for crop year 1974-75 for all segments allocated (431) during LACIE Phase II (1975-76). Since these same SRS estimates were used in Table II, Table III

can be compared with the results in Table II. The relative difference
at the Great Plains level for this aggregation is + 0.80%. This is an
estimate of the relative bias due to Group II estimation and Group III
ratioing of those counties not allocated segments. Note that for
practical purposes this relative difference is negligible. Combining
this with the results in Table II, an approximation of the relative bias
due to the Group III ratioing of those counties whose allocated seg-
ments were lost to nonresponse is -1.6% at the Great Plains level.
From a practical viewpoint, this estimate is negligible and indicates
that the nonresponse of the 160 allocated segments introduced negligible
bias at the Great Plains level.

The results of Table II and Table III indicate that the relative
difference of the LACIE estimate and the SRS estimate at the Great
Plains level is negligible whether all allocated segments are used
in the aggregation or only those segments not lost to nonresponse are
used in the aggregation.

Upon examination of the individual state relative differences,
Texas is seen to have a relative difference of +10.2%. From a practical
viewpoint, this indicates that the relative bias due to Group II esti-
mation and Group III ratioing of counties not allocated segments in
Texas is probably non-negligible. This indicates possibly a problem
in aggregation logic or a problem in sampling allocation exists in Texas.
The latter is more likely to be the problem.

5.4 LACIE Phase II Simulation with Nonresponse.

Table IV shows the aggregation of county SRS estimates for crop
year 1974-75 for all segments processed (394) during LACIE Phase II
(1975-76). Since 91.4% of the allocated segments were processed in

Phase II, Table IV differs only slightly from Table III. The relative difference at the Great Plains level is 0.10%. Therefore, practically speaking, the relative bias due to Group II estimation and Group III ratioing of both those counties whose allocated segments were lost to nonresponse and those counties not allocated segments is negligible. Combining the results of Table III with those of Table IV, an estimate of the relative bias due to the Group III ratioing of those counties whose allocated segments were lost to nonresponse (37) is − 0.70%,which for practical purposes is negligible.

As in section 5.3, when the results of Table IV are compared with the results of Table III, the relative difference between the LACIE estimate of wheat acreage at the Great Plains level and the SRS estimate is negligible whether or not segments are lost to nonresponse. Also, the estimate of bias due solely to segments lost to nonresponse is negligible indicating that nonresponse is introducing negligible bias.

5.5 Results of LACIE Phase I.

Table V shows the aggregation of the LACIE estimates of county wheat proportions as determined by CAMS in Phase I. Hence, the relative difference is due not only to (a) the bias due to Group II estimation and Group III ratioing of both those counties not allocated segments and those counties whose allocated segments were lost to nonresponse, but also to (b) the bias induced by within county classification and sampling errors. Table V shows a relative difference of -11.0% with a c.v. of 4.7% at the Great Plains level. This indicates

that the relative bias is <u>not</u> zero. (Assuming that the historic ratio used in obtaining the Group III ratio estimates is constant from year to year, this c.v. is the correct c.v. for determining whether or not the relative difference is significant.) Since the study in section 4.0 and the results in section 5.0 indicate that the relative difference due to (a) is negligible, one must conclude that the difference is due to within county classification and sampling errors and not nonresponse, Group II estimation or Group III ratioing.

Table I

1975 SEGMENTS PROCESSED

74 SRS COUNTY W% DATA

# SEGS	STATE	WHEAT AREA PLANTED SRS 1,000 ACRES 2	ESTIMATED WHEAT AREA (1,000 ACRES) USING LACIE ACQUISITIONS 3	$\frac{(3-2)}{3}$ X 100	CV%
24	CO	2,800	2,897	+ 3.3	6
55	KS	12,000	12,122	÷ 1.0	2
23	NE	2,941	2,865	- 2.7	6
29	OK	7,000	7,085	+ 1.2	3
28	TX	5,600	5,600	0.0	9
9	MN*	2,794	2,807	0.5	12
41	ND	9,787	9,581	- 2.2	5
39	MT	4,964	4,983	+ 3.8	3
23	SD	3,305	3,287	-0.5	5
271		51,191	51,227	+ .07%	1.8

Table II

1975 SEGMENTS PROCESSED

75 SRS COUNTY W% DATA

# SEG	STATE	WHEAT AREA* (1,000 ACRES) SRS 2	ESTIMATED WHEAT AREA (1,000 ACRES) USING LACIE ACQUISITIONS 3	$\frac{3-2}{3}$ X 100	CV%
24	CO	2,240	2,215	– 1.1	7
55	KS	12,100	12,278	+ 1.4	3
23	NE	3,008	2,876	– 4.6	8
29	OK	6,700	6,733	+ .5	4
28	TX	5,700	5,579	– 2.2	10
9	MN	2,844	2,995	+ 5.0	15
41	ND	10,213	9,963	– 2.5	4
39	MT	4,975	4,950	– .5	4
23	SD	3,003	2,807	– 7.0	7
271	9	50,783	50,396	– .8%	2.0

* Harvested acres

Table III

1976 SEGMENTS ALLOCATED

75 SRS COUNTY W% DATA

# SEGS	STATE	WHEAT AREA* (1,000 ACRES) SRS 75 2	ESTIMATED WHEAT AREA USING LACIE ALLOCATIONS 3	$\frac{3-2}{3}$ X 100 % RD	CV%
32	CO	2,240	2,163	- 3.6	7
84	KS	12,100	12,064	- 0.3	2
35	NE	3,008	3,071	2.1	5
40	OK	6,700	6,776	+ 1.1	3
49	TX	5,700	6,348	+ 10.2	5
13	MN	2,844	2,957	+ 3.8	11
85	ND	10,213	10,123	- .89	3
60	MT	4,975	4,948	- .55	3
33	SD	3,003	2,740	-9.6	7
431		50,783	51,190	+ .80	1.4

* Harvested acres

Table IV

1976 SEGMENTS PROCESSED

75 SRS COUNTY W% DATA

# SEG	STATE	WHEAT AREA * (1,000 ACRES) SRS 75 2	ESTIMATED WHEAT AREA 1,000 ACRES) USING LACIE ACQUISITIONS 3	$\frac{3-2}{3}$ X 100	CV%
32	CO	2,240	2,163	-3.6	7
81	KS	12,100	12,130	+0.2	2
33	NE	3,008	2,999	-0.3	6
40	OK	6,700	6,776	+1.1	3
47	TX	5,700	6,140	+7.2	5
10	MN	2,844	2,980	+4.6	13
67	ND	10,213	10,014	-1.3	3
54	MT	4,975	4,996	+0.4	2
28	SD	3,003	2,647	-13.4	7
394	9	50,783	50,845	+0.1%	1.4

* Harvested Acres

Table V

LACIE PHASE I RESULTS

State	Number of Segments Acq.	Number of Segments Allocated	SRS Est.[a] Acres (10^3)	LACIE Est. Acres (10^3)	$\%$ Relative[c] Difference	LACIE Estimated CV$\%$
COLORADO	24	32	2,260	3,058	26.1	20
KANSAS	55	84	.12,100	12,942	6.5	6
NEBRASKA	23	35	3,070	2,657	-15.5	31
OKLAHOMA	29	40	6,700	6,864	2.4	11
TEXAS	28	49	5,700	4,219	-35.1	21
MINNESOTA	9	13	2,844	2,150	-32.3	19
NO. DAKOTA	42	65	10,213	5,849	-74.5	10
MONTANA	39	60	4,975	3,947	-26.0	23
SO. DAKOTA	23	33	3,003	4,126	27.2	13
9 STATES	272	411	50,865	45,812	-11.0	4.7
U.S.[b]	--	637	69,656	___	___	3.7

a. January 1976 SRS Estimate of Wheat Area for the Crop Year 1974-75.

b. Projected.

c. $\dfrac{\text{LACIE-SRS}}{\text{LACIE}}$ x 100; This is done so that the $\%$ relative difference and CV are consistent.

d. CV: coefficient of variation = $\dfrac{\text{standard deviation}}{\text{LACIE}}$ x 100.