Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

LANDSAT ESTIMATION WITH CLOUD COVER

George A. Hanuschak, Statistical Reporting Service
U.S. Department of Agriculture, Washington, D.C. 20250

I. ABSTRACT

The problem addressed in this paper is crop acreage estimation techniques which utilize LANDSAT imagery that is not cloud free. Several statistical techniques are proposed that would allow inferences about the population even if cloud cover is an extensive problem. These techniques are entirely dependent upon a random sample of ground data from the total population of interest.

II. INTRODUCTION

One serious problem in utilizing satellite imagery for crop acreage estimation is cloud covered areas on the imagery. Cloud covered areas are essentially nonresponse areas for crop acreage estimation. However, a well designed random sample of ground data allows for statistically sound crop acreage estimation techniques utilizing satellite imagery.

The proposed approaches to this problem are post stratification or domain estimation depending on the particular set of assumptions adopted. The first set of assumptions would be that a given single date of imagery could be used and that the total number of frame units in the cloud cover post stratum are known. The second set of assumptions would be a given complete set of imagery for a state and the total number of frame units in the cloud cover domain is not known.

Under either set of assumptions in the cloud free (post stratum or domain), a regression estimator from a stratified (land use) random sample will be used. Also, in the cloud covered (post stratum or domain), a direct expansion estimate of a total from a stratified (land use) random sample will be used.

In application, the second set of assumptions would be more realistic for several reasons. First, single date imagery is not available for an entire state and secondly the determination of only sample segments into the appropriate domains instead of all frame units would considerably reduce the resources involved.

III. DOMAIN ESTIMATION

Thus, an example of the second technique will be presented using a specific crop and geographic area such as corn in the state of Illinois.

Area Frame - Aerial photography is photo interpreted to define land use strata.

Let $h = 1, 2, \ldots, L$ be the L land use strata

Let $j = \begin{cases} 1 & \text{Cloud free domain (all area frame units not covered by clouds)} \\ 2 & \text{Cloud covered domain (all area frame units covered by clouds)} \end{cases}$

Let $i = 1, 2, \ldots, n_{hj}$ be the number of area frame units in the sample (n) in the h^{th} land use stratum and j^{th} domain. The n_{hj}'s are random variables.

N_{hj} = The total number of area frame units in the h^{th} land use stratum and j^{th} post stratum and is not known.

Y = Total corn acres for the state of Illinois

$\hat{Y}$ = Estimated total corn acres for Illinois

$$\hat{Y} = \sum_{j=1}^{2} {}_j\hat{Y} = {}_1\hat{Y} + {}_2\hat{Y} \quad (1)$$

$$\mathrm{Var}(\hat{Y}) = \mathrm{Var}({}_1\hat{Y}) + \mathrm{Var}({}_2\hat{Y}) + 2\,\mathrm{Cov}({}_1\hat{Y},{}_2\hat{Y}) \quad (2)$$

If N_{hj} were known, then ${}_j\hat{Y} = \sum_{h=1}^{L} N_{hj}\,\bar{y}_{hj}$

However, N_{hj} are not known and the following device is used to derive the estimates and the variances of the estimates (Cochran 1963, Huddleston 1966).

Define for every unit in the population:

$$ {}_jy'_{hi} = \begin{cases} {}_jy_{hi} & \text{total corn acres in the } i^{th} \text{ segment in the } h^{th} \text{ land use stratum and } j^{th} \text{ domain.} \\ 0 & \text{otherwise} \end{cases} $$

So, for the cloud cover domain the total estimator is:

$$ {}_2\hat{Y} = \sum_{h=1}^{L} N_h \left(\sum_{i=1}^{n_h} {}_2y'_{hi} \right) / n_h $$

The sample estimate of the variance of the total is:

$$ v({}_2\hat{Y}) = \sum_{h=1}^{L} \frac{N_h^2}{n_h(n_{h-1})} \left(\frac{N_h - n_h}{N_h} \right) \cdot \left[\sum_{i=1}^{n_h} {}_2y'^2_{hi} - \frac{\left(\sum_{i=1}^{n_h} {}_2y'_{hi} \right)^2}{n_h} \right] $$

For $j = 1$, the sample estimate of the variance of the regression is:

$$ v({}_1\hat{Y}) = \sum_{h=1}^{L} \left(\frac{N_h}{N} \right)^2 \left(\frac{N_h - n_h}{N_h} \right) \cdot \left[\sum_{i=1}^{n_h} {}_1y'^2_{hi} - \frac{\left(\sum_{i=1}^{n_h} {}_1y'_{hi} \right)^2}{n_h} \right] \cdot \left[\frac{1 - r_h^2}{n_h - 2} \right] $$

where the estimator is presented as (Von Steen and Wigton 1976) the following:

$$ {}_1\hat{Y} = \sum_{h=1}^{L} N_h \, {}_1y'_{h\cdot} $$

and

$$ {}_1y'_{h\cdot} = {}_1\bar{y}_h + \hat{b} \left({}_1\bar{X}_h - {}_1\bar{x}_h \right) $$

Term definitions are:

- ${}_1\bar{y}_h$ - the average corn acres per sample segment from the ground survey for the h^{th} land use stratum in the cloud free domain.
- $\hat{b}$ - the estimated regression coefficient.
- ${}_1\bar{X}_h$ - the average number of pixels of corn per sample segment over the entire h^{th} stratum in the cloud free domain.
- ${}_1\bar{x}_h$ - the average number of pixels of corn per sample segment for the selected sample units in the h^{th} stratum in the cloud free domain.

To estimate total corn acres for the state of Illinois and the variance of the estimator, we return to equations (1) and (2).

$\hat{Y} = {}_1\hat{Y} + {}_2\hat{Y}$ where ${}_1\hat{Y}$ and ${}_2\hat{Y}$ are estimated as shown previously.

$$ v(\hat{Y}) = v({}_1\hat{Y}) + v({}_2\hat{Y}) + 2\text{cov}({}_1\hat{Y}, {}_2\hat{Y}) $$

where $v({}_1\hat{Y})$ and $v({}_2\hat{Y})$ are as shown previously.

To derive the $\text{cov}({}_1\hat{Y}, {}_2\hat{Y})$ consider for a particular land use stratum the allocation of corn acres in the domains.

Sample i unit	${}_1y'_h$	${}_2y'_h$
1	40	0
2	0	50
3	65	0
.	.	.
.	.	.
.	.	.
n_h	48	0

Then $\text{cov}({}_1\hat{Y}_h, {}_2\hat{Y}_h) = \text{cov}[N_h({}_1y'_h), N_h({}_2y'_h)]$

$$ (N_h)^2 \, \text{cov}({}_1y'_h, {}_2y'_h) = $$

$$ (N_h)^2 \cdot \left[\frac{\sum_{i=1}^{n_h} {}_1y'_{hi} \, {}_2y'_{hi} - \frac{\left(\sum_{i=1}^{n_h} {}_1y'_{hi} \right)^2 \left(\sum_{i=1}^{n_h} {}_2y'_{hi} \right)^2}{n_h}}{(n_h - 1)} \right] $$

All cross product terms are obviously zero.

Thus,

$$ \text{cov}({}_1\hat{Y}_h, {}_2\hat{Y}_h) = \frac{-N_h^2 \left(\sum_{i=1}^{n_h} {}_1y'_{hi} \right)^2 \left(\sum_{i=1}^{n_h} {}_2y'_{hi} \right)^2}{(n_h)(n_h - 1)} $$

and $\text{cov}({}_1\hat{Y}, {}_2\hat{Y}) = \sum_{h=1}^{L} W_h^2 \cdot \text{cov}({}_1\hat{Y}_h, {}_2\hat{Y}_h)$

where W_h is the appropriate stratum weighting factor.

In application, the n_{hj} may not be large enough for precise estimation of hj^{th} stratum-domain combination. It may then be necessary to collapse the number of strata to get estimates with an acceptable precision for each domain. A set of rules regarding the amount of cloud cover could be established to assign each "sample segment" to one of the two domains.

Under the first set of assumptions and using the post stratification technique, the variance formulas for the post stratum include N_{hj} and only the observations in that domain are used. That is, the variate ${}_jy'_{hi}$ is not necessary.

IV. CONCLUDING REMARKS

The theory of domain estimation proposed in this paper has not yet been applied to crop acreage estimation. The task of applying this theory to crop acreage estimation is a formidable one. The Statistical Reporting Service has a LANDSAT project in Texas for the 1975 crop year that will provide the opportunity to test the feasibility of these methods. The essential component of the methodology proposed is the random sample of ground data for the entire population of interest, which is provided by the operational systems of the Statistical Reporting Service.

ACKNOWLEDGMENT

The author would like to express his thanks to Harold F. Huddleston, Principal Research Statistician, USDA-SRS, Earl E. Houseman, Mathematical Statistician, USDA-SRS, and fellow members of the New Survey Techniques Section, USDA-SRS, for their guidance and stimulating discussions in this application area of domain theory.

REFERENCES

1. Cochran, W. G., Sampling Techniques, John Wiley and Sons, 2nd Edition, 1963.

2. Huddleston, H. F., Analytic Studies in Sample Surveys, United States Department of Agriculture, Statistical Reporting Service, 1966.

3. Von Steen, D. H. and W. H. Wigton, Crop Identification and Acreage Measurement Utilizing LANDSAT Imagery, United States Department of Agriculture, Statistical Reporting Service, 1976.