SAMPLING UNITED STATES FORESTS
USING LANDSAT DATA

By

Virgil L. Anderson, Louis J. Cote and K.C.S. Pillai

Technical Report #85-22

Department of Statistics
Purdue University

December, 1985

SAMPLING UNITED STATES FORESTS USING LANDSAT DATA

By

Virgil L. Anderson, Louis J. Cote and K.C.S. Pillai

I. Background (Including the Population)

Landsat data are available over all of North America and may be used to estimate the areal extent of vegetation types. NASA (Goddard) is interested in obtaining areal estimates of conifer and hardwood cover types in the 48 contiguous United States. Of particular interest is the estimate of the total number of hectares of conifers in these 48 states. The basic sampling concepts developed to assess conifer acreage may be generalized to assess other land cover categories over continental or subcontinental areas.

Various steps were taken to arrive at approximate sampling rates for the different stages. For example, Ross Nelson of Goddard was able to show that it was better to take fewer large blocks (200 x 200 pixel blocks) of MSS (Multispectral Scanner) data than more small 50 x 50 pixel blocks to obtain good estimates of the number of hectares of conifers. (There are 0.803 acres or 0.3249 hectares in a geometrically corrected pixel from MSS data). In general, in this study, we utilized sampling rates suggested by Goddard personnel. The actual method to obtain sample sizes is given later in this paper.

The usual sampling design for these investigations consists of a stratified random sample with two stages within each stratum. The subcontinental area (the 48 contiguous United States) being studied is stratified into major vegetation categories. Within each stratum at least two MSS scenes are randomly selected. The MSS scenes are natural "clusters" of spectral observations, and at least two scenes must be selected in order to calculate between-scenes variability. Within each MSS scene, a systematic subsample of discrete blocks of pixels are selected. These blocks are classified into the cover types of interest. Portions of the blocks are assessed using aerial photos or Thematic Mapper (TM) photo imagery. Regressions are derived which, for each MSS scene, correct the MSS areal estimates to the photointerpreted estimates.

The population of interest (in this study, the continental United States) is divided into strata. The strata are divided into scenes, the scenes into 497 x 500 pixel blocks and the 497 x 500 pixel blocks into small blocks of approximately 100 x 100 pixels. Let us describe each group as follows:

A. Strata

The 48 contiguous United States are divided into seven forest regions. These strata, derived from a United States Forest Service map of the major forest cover types of these United States, are:

1. Western Conifer
2. Northern Hardwood
3. Northern Conifer
4. Southern Conifer
5. Pinyon-Juniper
6. Central Hardwood
7. Nonforest

B. Scenes (Multispectral Scanner, MSS)

Each stratum is divided into as many scenes as the entire stratum demands. One scene is made up of 2983 lines by 3548 columns of pixels. Twenty-four columns from the left and right sides of the image and the last line are dropped to produce the MSS scene; illustrated in Figure 1. The area enclosed by one scene is 170 kilometers x 185 kilometers or about 105 miles x 114 miles. Since the land areas of the strata are different, each stratum contains a different number of scenes.

C. Large Blocks (497 x 500 pixel blocks)

Each MSS scene contains 30 "usable" large blocks. Figure 1 indicates the peculiarities encountered in a given scene.

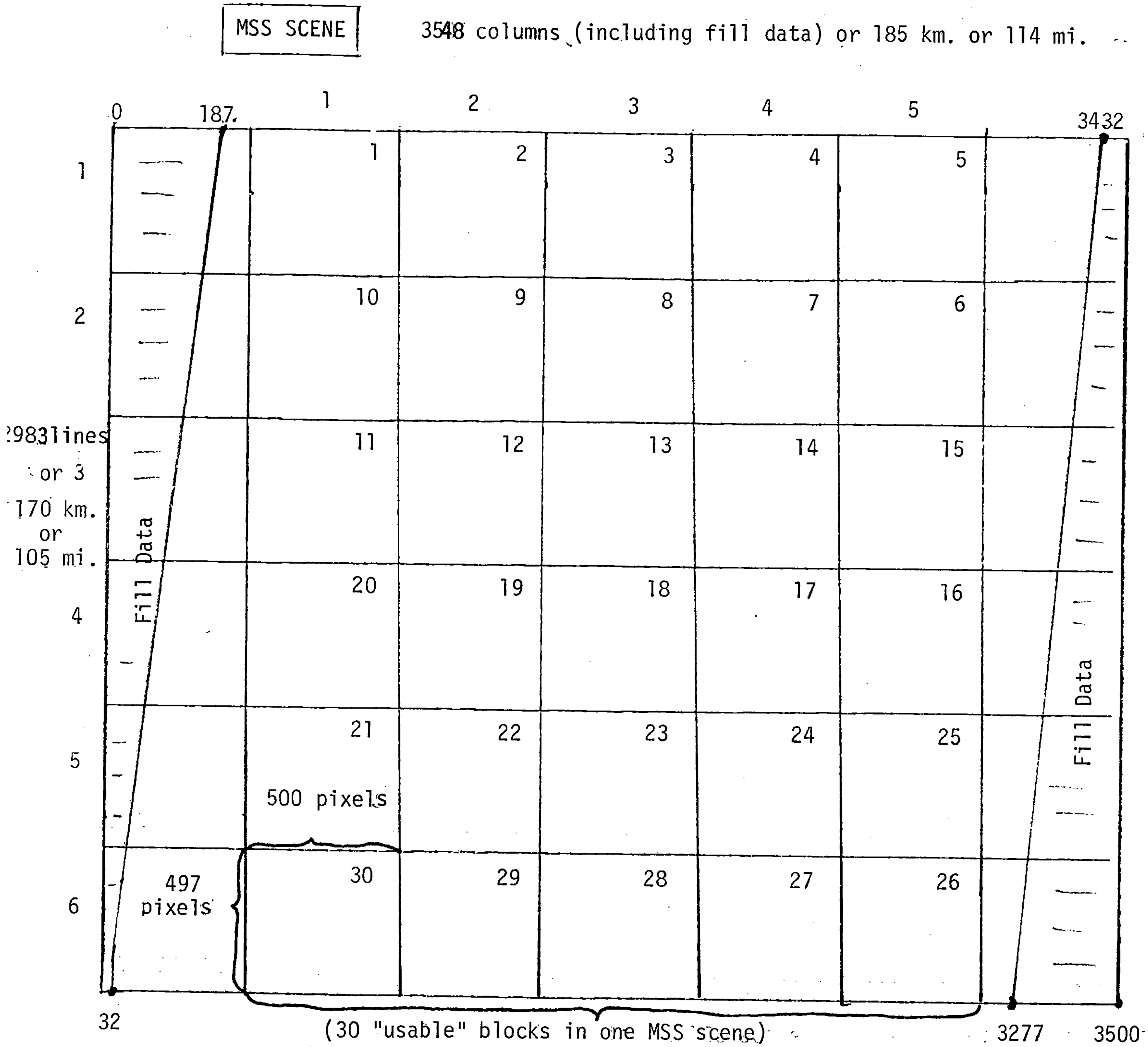


Figure 1.

ONE MSS SCENE (with 30 - 497 x 500 pixel blocks)

LARGE BLOCK

500 columns

100

100

497 lines

1	2	3	4	5
10	9	8	7	6
11	12	13	14	15
20	19	18	17	16
21	22	23	24	25

Figure 2.

One Large Block (with 25 small blocks)

II. Design of the Study (Sampling)

The main objective of this study is to show an efficient sampling plan to allow an estimate of the total number of hectares for forest cover types in the 48 contiguous United States which may be used as a model for sampling forests in other nations. In addition, an estimation procedure for obtaining sample sizes at the scene and large block levels is provided.

One assumption that is used throughout this paper is that the variance between small blocks within a large block is much smaller than the variance between large blocks. In the example given later, it is readily seen that this tends to be true.

Utilizing the concept of small variance between small blocks, we have adopted the procedure of sampling only <u>one</u> small block, #13 (the center small block in Figure 2.) to represent a large block that will be sampled. Hence in all cases when describing the sampling methodology, we are assuming that scenes will be sampled and large blocks will be sampled, but that only one small block (#13) will be used in the estimation procedure from a large block drawn in the sample. The variance components of the large blocks and small blocks are estimated together. In the example given later, there are two small blocks per large block, but it would have been much more efficient (that is, we could have obtained more information or had a smaller variance estimate of the total if only one small block were taken per large block and twice as many large blocks were included in the sample) to have sampled as we suggest in this part of the paper.

To show the effectiveness of using only one small block within a large block and use more large blocks, let us use the data from one scene on conifers given in detail in the example later.

Using Scene 1 $\quad$ y = number of pixels of conifers

Small Blocks (8) (100 x 100 pixel blocks) (random)

(4) Large Blocks (497 x 500 pixel block) (random)

1	2	3	4
1) 8571	3) 2899	5) 5050	7) 3206
2) 8010	4) 2459	6) 5460	8) 1623

Analysis of Variance [Reference: Ostle and Mensing (p. 310, 1975)]

Source	df	SS	MS	EMS
Large Blocks	3	45,002,373	15,000,791	$\sigma_s^2 + 2\sigma_L^2$
Small Blocks in large Blocks	4	1,591,155	397,789	σ_s^2
Total	7	46,593,528		

$$\hat{\sigma}_s^2 = 397{,}789 \text{ and } \hat{\sigma}_L^2 = \frac{15{,}000{,}791-397{,}789}{2} = 7{,}301{,}501$$

The estimate of the variance of the mean for Scene 1 is

$$\hat{\sigma}^2_{\bar{y}_{hi}} = \frac{\hat{\sigma}_s + 2\hat{\sigma}^2_L}{8} = \frac{15,000,791}{8} = 1,875,099$$

If only one small block is used per large block, assuming the variance components remain the same and allowing the variance of the scene mean to be

$$\hat{\sigma}^2_{\bar{y}_{hi}} = \frac{\hat{\sigma}^2_s + \hat{\sigma}^2_L}{n} = 1,875,099$$

then

$$1,875,099\ n = 397,789, + 7,301,501$$

$$\therefore \quad n = 4.1 \quad \begin{pmatrix}\text{number of large blocks with} \\ \text{one small block per large block}\end{pmatrix}.$$

Hence if one takes 5 large blocks with 1 small block per large block the variance of the scene mean will be smaller than taking 4 large blocks with 2 small blocks per large block.

In sampling the population of interest, we will follow the pattern set forth in Section I of this paper by using the following:

A. Strata

The sampling procedure for strata is to take all seven strata in the sample.

B. Scenes (MSS scenes within strata)

For a given stratum one may number the scenes in a serpentine manner similar to the numbering system given in Figure 1 for large blocks within a scene. If this numbering system is used on all the strata, a systematic sample with a random start is an excellent sampling method. The estimation procedure assumes random sampling under those conditions.

The overall sampling procedure for MSS scenes assumes that the MSS scenes are allocated randomly within each forest stratum proportional to the area within that stratum. The random allocation is constrained as follows: 1) At least two MSS scenes are located in each stratum. 2) An MSS scene is a candidate for selection within a stratum if its Worldwide Reference System (WRS) nominal scene centerpoint is within the stratum boundaries. 3) MSS scenes within the nonforest stratum may be purposefully allocated. A small number of MSS scenes may be adequate to characterize an extensive nonforest stratum.

The total number of scenes allocated will most likely be driven by the cost of the digital data, currently $730/scene. In this study 20 MSS scenes are available for allocation. Three are purposefully allocated to the nonforest stratum, the remaining 17 scenes are allocated proportionally to the remaining area in the strata.

Next we must show the procedure for obtaining the sample of scenes within a stratum. Assume there are 25 scenes numbered in a serpentine manner within the chosen stratum and we want 5 scenes to represent that stratum. The following method of sampling may be used:

1) Draw a random number between one and five, say 2.

2) Then take #2, 7, 12, 17 and 22 as the scenes to be investigated.

C. Large Blocks (497 x 500 pixel blocks with scenes)

Results from preliminary studies [(Horning et al., (1985) and Nelson et al., (1985)] have shown that the number of sample blocks required within a given scene 1) varies between strata; 2) is a function of the MSS scene sampling intensity within strata; and 3) is a function of the proportion of the cover type of interest within the MSS scene. With so many diverse factors affecting sample block allocation, no generic estimate of the number of sample blocks needed per

MSS scene can be derived.

In addition, the same samples are used both to estimate the regression adjustment and the mean hectares of conifers. The allocation must reduce both the variance of the adjustment and of the estimate of the mean. This is discussed later in the report.

Next we need to describe the sampling procedure within a scene. Assume we want to obtain 8 blocks from a scene described in Figure 1. To run a systematic sample with a random start so that eight (8) blocks are used for the sampled blocks, one should take a random number between 1 and 3 and take every 3rd block (30/8 equals 3 plus a remainder) starting from the upper left hand corner. Then advance through the blocks in a serpentine manner (e.g. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29 if one obtains "2" as the random start number. This, of course, gives 10 possible blocks.

Next,(do all of this before going to the MSS scene) order the 10 blocks at random by picking numbers from 1 to 30 and using only the 10 admissible ones, namely 2,5,... ,29 in this case. If the order of draw was

11, 17, 2, 29, 23, 5, 14, 26, 8, 20

do not use 8 and 20 unless there is a cloud cover on one or more of the first eight blocks, then use the #8 block first to replace the first clouded block (or a block that could not be used for some reason) and 20 for the next replaced block. If there are more than two blocks to be replaced, add one to the block number to be replaced and use it. Of course if 30 is to be replaced, take the #1 block. A "clouded" block is defined as one that has over 30% of its area clouded.

After the eight sampled blocks have been selected, classify the pixels in each block (497 x 500 pixels) into conifer hardwood, water and other using MSS imagery. We will develop the estimation formulas for conifer only.

If, however, 15 blocks are taken, pick a random number, one or two. If it turns out to be one, take all odd numbered blocks, and of course if it is two, take all even numbered blocks.

D. Small Blocks (100 x 100 pixel blocks within large blocks)

Using the population description given in Figure 2, we may use the following sampling procedure to obtain small blocks within large blocks. If a random sample of two 100 x 100 pixel blocks is to be taken from the 25, draw a random number between 1 and 13 for the first 100 x 100 pixel block to be sampled. Then, add 12 to that number for the second 100 x 100 pixel block to be included in the sample. In this sampling procedure, #13 has the largest probability of being drawn. This characteristic is desirable because the middle of the large block is thought to better represent the block than the outside small blocks. If only one 100 x 100 pixel block is to be chosen from a selected large block, take the #13 small block only.

This method of sampling is considered a random sample of 100 x 100 pixel blocks within the large block.

III Analysis

It must be understood at the onset of examining the analyses of data from the sampling plan set up in previous sections of this paper that we are assuming random sampling throughout the various stages, and the analyses are reasonably good approximations. However the correlation between the estimate from the large 497 x 500 pixel blocks and that from the 100 x 100 pixel block taken from the same large block is ignored, the variance of the estimated mean of the large block is ignored, and many covariances that may be associated with the terms for the adjusted means are ignored. In short, this presentation provides simplified estimates of the variance of the adjusted means, following the versions given in Scheaffer et al. (1979, henceforth SMO) and Cochran (1977).

In a personal conversation with Dr. James R. Chromy of the Research Triangle Institute in North Carolina, we learned that the exact variance for this type of sampling (taking full account of the design) is being investigated by their group at the present time.

Thematic Mapper (TM) photo products and/or National High Altitude Photography (NHAP) products are available over much of the United States. These photo products may be used to make aerial estimates of the cover type of interest on selected areas.

In our sampling scheme we are using only one small 100 x 100 pixel block per large 497 x 500 pixel block sampled. The data used in this scheme are of three types:

1. Conifer acreage, 497 x 500 pixel block, from MSS digital data.
2. Conifer acreage, 100 x 100 pixel block within the large block, MSS data.
3. Conifer acreage, 100 x 100 pixel block, TM or NHAP photo data.

The data on the 100 x 100 pixel blocks (both photo and MSS) are used to derive a regression equation for a given scene that adjusts the MSS estimate to the photo estimate. The 497 x 500 pixel block conifer acreage estimate is transformed using the equation to correct the MSS estimate to the more accurate photo estimate. To do this the acreage estimate derived from the 497 x 500 pixel block is scaled to the 100 x 100 pixel block size (divided by 24.85) and the regression equation is applied. The exact procedure for doing these calculations is shown later.

To show how the estimates for the entire 48 contiguous states are made, we suggest analyzing the data from the regression estimate for the scene to the stratum level and finally to the National level as follows:

A. Regression Estimate (Scene)

It is necessary to utilize the information on the MSS scenes to estimate the population total number of hectares of conifers in the 48 contiguous United States. Since the TM data are relatively expensive to obtain and process, only the TM information on the 100 x 100 pixel blocks is utilized. MSS scene data are not as accurate as photointerpreted TM data, but are quite inexpensive to obtain and process. The entire large block (497 x 500) is classified into the cover types of interest. These large blocks contain the 100 x 100 pixel block data. The small block data and the corresponding TM photo data are used in regression estimation.

Let us assume that there exists a regression coefficient appropriate for each scene. The next step could be that one may want to pool certain scenes that have the same slope. For the general case, however, we will assume that there is a separate adjustment (slope) for each scene.

To show how the regression estimator is to be used, we define population and sample indices as follows:

1) Population:

$h = 1,2,3,4,5,6,7$; Strata

$i = 1,2,3,\ldots,N_h$; Scenes in stratum h,

$j = 1,2,\ldots,M_{hi}$; Large blocks in scene i, stratum h.

2) Sample:

$h = 1,2,3,4,5,6,7$; Strata

$i = 1,2,\ldots,n_h$; number of MSS scenes sampled in stratum h

$j = 1,2,\ldots,m_{hi}$; number of 100 x 100 pixel blocks sampled in scene i, stratum h.

To begin the estimating procedure, we will find the adjusted estimate of hectares of conifers in the i^{th} Scene by using the MSS information from the small blocks as x and TM information from the small blocks as y.

Using the results on page 123 of Schaeffer, Mendenhall and Ott (SMO) (1979)

$$\hat{\mu}_{hi} = \bar{y}_{hi} + b_{hi}(\hat{\mu}_{x_{hi}} - \bar{x}_{hi}) \tag{1}$$

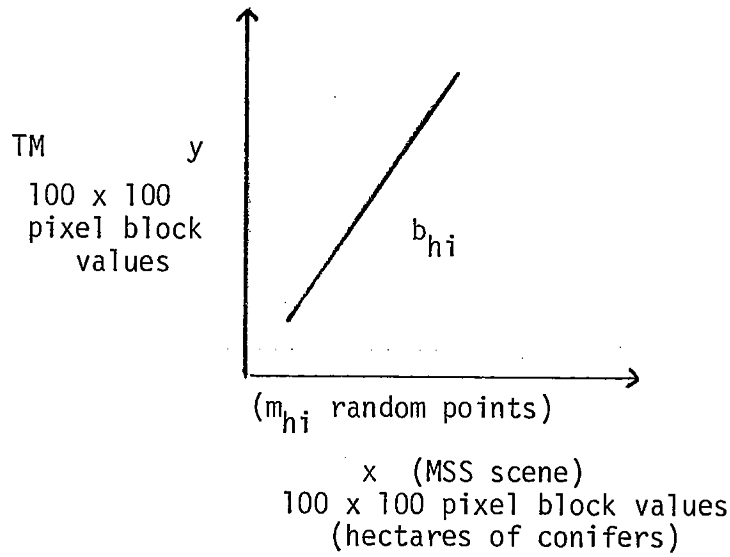
where $\hat{\mu}_{hi}$ = adjusted estimated population mean hectares of conifers per 100 x 100 pixel block in the i^{th} scene in the h^{th} stratum,

$\bar{y}_{hi}$ = mean hectares of conifers from the sampled 100 x 100 pixel blocks Thematic Mapper (TM) photo data in the i^{th} scene sampled in the h^{th} stratum = $\dfrac{\sum_{j=1}^{m_{hi}} y_{hij}}{m_{hi}}$,

b_{hi} = linear regression coefficient found in the i^{th} scene of the h^{th} stratum using TM photo data as y and MSS data as x in the 100 x 100 pixel blocks sampled,

$\hat{\mu}_{x_{hi}}$ = estimated population mean hectares of conifers found in the 497 x 500 pixel blocks in the i^{th} scene in the h^{th} stratum. This mean value is scaled to a 100 x 100 pixel block

$\bar{x}_{hi}$ = sample mean of the i^{th} MSS scene in the h^{th} stratum for the 100 x 100 pixel blocks.



If we ignore the variance between 100 x 100 pixel blocks within large blocks and consider the m_{hi} observations random over the MSS scene, the estimated variance of the estimated mean for the $i^{\underline{th}}$ scene in the $h^{\underline{th}}$ stratum is:

$$\hat{V}(\hat{\mu}_{hi}) = \left(\frac{M_{hi} - m_{hi}}{M_{hi}\, m_{hi}}\right)\left(\frac{1}{m_{hi} - 2}\right)\left[\sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2 - b_{hi}^2 \sum_{j=1}^{m_{hi}} (x_{hij} - \bar{x}_{hi})^2\right] \quad (2)$$

where:

M_{hi} = Population number of 100 x 100 pixel blocks in the $i^{\underline{th}}$ scene in the $h^{\underline{th}}$ stratum,

m_{hi} = sample number of 100 x 100 pixel blocks in the $i^{\underline{th}}$ scene in the $h^{\underline{th}}$ stratum,

y_{hij} = number of hectares of conifers estimated from the TM photo corresponding to the $j^{\underline{th}}$ 100 x 100 pixel block in the $i^{\underline{th}}$ scene of the $h^{\underline{th}}$ stratum,

b_{hi} = estimated linear regression coefficient of y on x (TM on MSS) in the $i^{\underline{th}}$ scene in the $h^{\underline{th}}$ stratum

X_{hij} = number of hectares of conifers estimated from the MSS digital data in the $j\underline{th}$ 100 x 100 pixel block in the $i\underline{th}$ scene in the $h\underline{th}$ stratum.

B. Stratum Estimate

To obtain the estimate of the stratum mean on a per 100 x 100 pixel block, we will follow the procedure given by SMO (1979) page 203 by changing all of the symbols to the same ones with "h" as a subscript to represent the $h\underline{th}$ stratum, but using the regression estimated variance, eq (2), for the within variance. It follows that

N_h = the population number of scenes in the $h\underline{th}$ stratum

n_h = the sampled number of scenes in the $h\underline{th}$ stratum

M_{hi} = the population number of small blocks in the $i\underline{th}$ scene of the $h\underline{th}$ stratum

m_{hi} = the sampled number of small blocks in the $i\underline{th}$ scene of the $h\underline{th}$ stratum

$M_h = \sum_i M_{hi}$ is the total number of small blocks in the $h\underline{th}$ stratum

$\overline{M}_h = \frac{M_h}{N_h}$ = average number of small blocks in the $h\underline{th}$ stratum

y_{hij} = the number of hectares of conifers in the $j\underline{th}$ block (100 x 100 pixel block) in the $i\underline{th}$ scene of the $h\underline{th}$ stratum

$\bar{y}_{hi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$ = the sample mean number of hectares of conifers in the $i\underline{th}$ scene of the $h\underline{th}$ stratum

and the unbiased estimator of the population mean number of hectares of conifers in the $h\underline{th}$ stratum, μ_h:

$$\hat{\mu}_h = \frac{N_h}{M_h} \quad \frac{\sum_{i=1}^{n_h} M_{hi} \; \hat{\mu}_{hi}}{n_h} \tag{3}$$

and the estimated variance of $\hat{\mu}_h$:

$$\hat{V}(\hat{\mu}_h) = \left[\left(\frac{N_h - n_h}{N_h}\right) \frac{1}{n_h \overline{M}_h^2} s_{hb}^2 + \right. \tag{4}$$

$$\left. \frac{1}{n_h N_h \overline{M}_h^2} \sum_{i=1}^{n_h} M_{hi}^2 \left(\frac{M_{hi} - m_{hi}}{M_{hi}}\right)\left(\frac{s_{hi}^2}{m_{hi}}\right)\right],$$

where $s_{hb}^2 = \dfrac{\sum_{i=1}^{n_{hi}} (M_{hi}\,\bar{y}_{hi} - \overline{M}_h \hat{\mu}_h)^2}{n_h - 1}$

and

$$s_{hi}^2 = \left[\frac{\sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2 - b_{hi}^2 \sum_{j=1}^{m_{hi}} (x_{hij} - \bar{x}_{hi})^2}{m_{hi} - 2}\right]$$

C. National Estimate

Since there are seven strata in the 48 contiguous United States and each stratum is sampled in a similar manner, the overall estimation procedure for the population of the 48 states is comparable to the methodology given on page 65 of SMO (1979). The estimator of the population total τ is:

$$\hat{\tau} = N\hat{\mu}_h = N_1\hat{\mu}_1 + \ldots + N_7\hat{\mu}_7 .$$

The estimated variance of $\hat{\tau}$ is:

$$\hat{V}(\hat{\tau}) = N^2\hat{V}(\hat{\mu}_h) = \sum_{h=1}^{7} N_h^2 \frac{N_h - n_h}{N_h} \hat{V}(\hat{\mu}_h).$$

where N = total number of MSS scenes in the U.S. or

$$N = \frac{\text{total area of U.S.}}{\text{area of one MSS scene}} .$$

IV. Example

A preliminary study was designed to compare Thematic Mapper (TM) photointerpreted areal estimates of number of hectares of conifers in two scenes with Landsat Multispectral Scanner (MSS) areal estimates for the same scenes. Two 100 x 100 pixel blocks were sampled from eight 497 x 500 pixel blocks in scene 1 and from four 497 x 500 pixel blocks in scene 2.

The data were recorded in pixels and analyzed as pixels here. To convert the results to hectares, use the relationship:

1 pixel = .3249 hectares.

Also to be more accurate, one should use the fact that there are not quite 25, 100 x 100 pixel blocks in a 497 x 500 pixel block. For demonstration of the method to use the estimation procedure, this example is adequate.

Southern Conifer Stratum (h)

Scene 1 (i = 1)

497 x 500 Block*	100 x 100 Block(j)	MSS X	TM Y
1	1	8571	7100
	2	8010	6000
2	3	2899	3600
	4	2459	3900
3	5	5050	5800
	6	5460	4400
4	7	3206	5100
	8	1623	3400

*Subscript ignored

$$\sum_{j=1}^{8} x_{h1j} = 37{,}278$$

$$\bar{x}_{h1} \cong 4660$$

$$\sum_{j=1}^{8} y_{h1j} = 39{,}300$$

$$\bar{y}_{h1} \cong 4912$$

$$\sum_{j=1}^{8} x_{h1j} y_{h1j} = 204{,}123{,}400$$

$$\sum_{j=1}^{8} x_{h1j}^2 = 220{,}299{,}688$$

$$\sum_{j=1}^{8} y_{h1j}^2 = 205{,}150{,}000$$

$$\sum_{j=1}^{8} (y_{h1j} - \bar{y}_{h1})^2 = 12{,}088{,}750$$

$$\sum_{j=1}^{8} (x_{h1j} - \bar{x}_{h1})^2 = 46{,}593{,}528$$

$$b_{h1} = \left[\frac{\sum_{j=1}^{8} x_{h1j} y_{h1j} - \frac{\left(\sum_{j=1}^{8} x_{h1j}\right)\left(\sum_{j=1}^{8} y_{h1j}\right)}{n}}{\sum_{j=1}^{8} x_{h1j}^2 - \frac{\left(\sum_{j=1}^{8} x_{h1j}\right)^2}{n}} \right]$$

$$b_{h1} = \left\{ \frac{204{,}123{,}400 - \frac{(37{,}278)(39{,}300)}{8}}{220{,}299{,}688 - \frac{(37{,}278)^2}{8}} \right\} \cong .45$$

From the full 497 x 500 pixel blocks

497 x 500 pixel block	MSS x
1	146,838
2	82,336
3	101,127
4	55,258

we obtain the estimated mean x for scene 1 as

$$\hat{\mu}_{x_{h1}} = \frac{146,838 + 82,336 + 101,127 + 55,258}{4 \times 25}$$

where 25 = approximate number of 100 x 100 pixel blocks in a 497 x 500 pixel block.

Note: Actually should be $\frac{497 \times 500}{100 \times 100} = 24.85$ instead of 25.

$$\hat{\mu}_{x_{h1}} \cong 3856$$

We know that the adjusted mean pixels of conifers in the 1st scene of the hth stratum is then

$$\hat{\mu}_{h1} = \bar{y}_{h1} + b_{h1}(\mu_{x_{h1}} - \bar{x}_{h1})$$

$$\hat{\mu}_{h1} = 4912 + .45(3856\text{-}4660)$$

$$\underline{\underline{\hat{\mu}_{h1} \cong 4550}}$$

The estimated variance of $\hat{\mu}_{h1}$ is:

$$\hat{V}(\hat{\mu}_{h1}) = \left(\frac{M_{h1} - m_{h1}}{M_{h1}m_{h1}}\right)\left(\frac{1}{m_{h1}-2}\right)\left[\sum_{j=1}^{m_{h1}} (y_{h1j} - \bar{y}_{h1})^2 - b_{h1}^2 \sum_{j=1}^{m_{h1}} (x_{h1j} - \bar{x}_{h1})^2\right]$$

$$\hat{V}(\hat{\mu}_{h1}) = \frac{750-8}{(750)(8)} \left(\frac{1}{8-2}\right) [12,088,750 - (.45)^2(46,593,528)]$$

$\hat{V}(\hat{\mu}_{h1}) \cong 54,170$ where:

M_{h1} = 30 x 25 = (# of large blocks in MSS scene)(# of small blocks/large block)
= 750.

Note: Actually it should be 30(24.85) = 745.5 instead of 750.

<u>Scene 2 (i = 2)</u>

497 x 500 Block*	100 x 100 Block j	MSS x	TM y
5	7	2816	2600
	8	2628	2700
6	9	1700	2800
	10	2225	3000

*Subscript Ignored)

$$\sum_{j=1}^{4} x_{h2j} = 9369$$

$$\bar{x}_{h2} = 2342$$

$$\sum_{j=1}^{4} y_{h2j} = 11100$$

$$\bar{y}_{h2} = 2775$$

$$\sum_{j=1}^{4} x_{h2j}y_{h2j} = 25,852,200$$

$$\sum_{j=1}^{4} x_{h2j}^2 = 22,676,865$$

$$\sum_{j=1}^{4} y_{h2j}^2 = 30,890,000$$

$$\sum_{j=1}^{4} (y_{h1j}-\bar{y}_{h2j})^2 = 87,500$$

$$\sum_{j=1}^{4} (x_{h2j}-\bar{x}_{h2j})^2 = 732,325$$

$$b_{h2} = \left\{ \frac{25,852,200 - \frac{(9369)(11100)}{4}}{22,676,865 - \frac{(9369)^2}{4}} \right\} = -.20$$

From the full 497 x 500 pixel blocks

497 x 500 Block	MSS X
5	52,395
6	55,311

$$\hat{\mu}_{x_{h2}} = \frac{52,395 + 55,311}{(2)(25)} = 2154$$

It follows that the adjusted mean pixels of conifers is:

$$\hat{\mu}_{h2} = \bar{y}_{h2} + b_{h2}(\hat{\mu}_{x_{h2}} - \bar{x}_{h2})$$

$$\hat{\mu}_{h2} = 2775 - .2(2154 - 2342)$$

$$\underline{\underline{\hat{\mu}_{h2} \cong 2813}}$$

The estimated variance of $\hat{\mu}_{h2}$ is

$$\hat{V}(\hat{\mu}_{h2}) = \left[\frac{750 - 4}{(750)(4)}\right]\left[\frac{1}{4-2}\right]\left[87,500 - (-.2)^2(732,325)\right]$$

$$\underline{\underline{\hat{V}(\hat{\mu}_{h2}) \cong 7230}}$$

Finally, we must find the estimated total number of pixels of conifers in the $h^{\underline{th}}$ stratum. To do this, we will find the adjusted estimated mean number of pixels of conifers per 100 x 100 pixel block in the $h^{\underline{th}}$ stratum, $\hat{\mu}_h$,

$$\hat{\mu}_h = \left(\frac{8\hat{\mu}_{h1} + 4\hat{\mu}_{h2}}{12}\right)$$

$$\hat{\mu}_h = \left(\frac{8(4550) + 4(2813)}{12}\right)$$

$$\underline{\underline{\hat{\mu}_h = 3971}}$$

The estimated total number of pixels of conifers in the $h^{\underline{th}}$ stratum, $\tilde{\tau}_h$, is

$$\tilde{\tau}_h = M_h\hat{\mu}_h$$

where: M_h = total number of 100 x 100 pixels in the $h^{\underline{th}}$ stratum. There are 30 large blocks (497 x 500 pixel blocks) per scene and we are assuming there are 20 scenes in the $h^{\underline{th}}$ stratum. Since we assume there are 25 small blocks (100 x 100 pixel blocks) in a large block, the total number of 100 x 100 pixel blocks in the $h^{\underline{th}}$ stratum is

20 x 30 x 25 = 15,000.

Hence

$$\tau_h = (15{,}000)(3971)$$

$$= 59{,}565{,}000 \text{ pixels of conifers in stratum h.}$$

The estimated variance of that total is:

$$\hat{V}(\hat{\tau}_h) = \frac{N_h - n_h}{N_h} \frac{N_h^2}{n_h} s_{bh}^2 + \frac{N_h}{n_h} \sum_{i=1}^{n_h} M_{hi}^2 \frac{M_{hi} - m_{hi}}{M_{hi}} \frac{s_{hi}^2}{m_{hi}}$$

where:

$$s_{bh}^2 = \frac{\sum_{i=1}^{n_h} \left(M_{hi} \hat{\mu}_{hi} - \frac{M_h}{N_h} \hat{\mu}_h \right)^2}{n_h - 1} ,$$

$$s_{hi}^2 = \frac{1}{m_{hi} - 2} \left[\sum_{j=1}^{m_{hi}} (y_{hij} - \bar{y}_{hi})^2 - b_{hi}^2 \sum_{j=1}^{m_{hi}} (x_{hij} - \bar{x}_{hi})^2 \right] ,$$

N_h = 20, population number of scenes in stratum h,

n_h = 2, sample number of scenes in stratum h,

$M_{h1} = M_{h2} \cong$ 750 population number of 100 x 100 pixel blocks in each scene of the h$^{\underline{th}}$ stratum. The actual number is

$$\frac{30 \times 497 \times 500}{100 \times 100} = 745.5,$$

m_{h1} = 8 , 100 x 100 pixel blocks in the sampled first scene,

m_{h2} = 4, 100 x 100 pixel blocks in the sampled second scene,

M_h = 750 x 20 = 15,000 = population number of 100 x 100 pixel blocks in stratum h.

Due to the unequal number of blocks (8 and 4) per scene we are using the approximation,

$$s_{bh}^2 \cong 750^2 \left\{ \frac{8(4550-3971)^2 + 4(2813-3971)^2}{12\ (2-1)} \right\}$$

$$s_{bh}^2 \cong 377,146,130,000.$$

Since

$$s_{h1}^2 = V(\hat{\mu}_{h1}) \cong 54,170$$

and

$$s_{h2}^2 = V(\hat{\mu}_{h1}) \cong 7,230,$$

$$\hat{V}(\hat{\tau}_h) \cong \left(\frac{20-2}{20}\right) \frac{(20)^2}{2} \Big[377,146,130,000$$

$$+ \left(\frac{20}{2}\right) (750)^2 \left(\frac{750-8}{750}\right) \left(\frac{54,170}{8}\right) + (750)^2 \left(\frac{750-4}{750}\right)\left(\frac{7,230}{4}\right)\Big]$$

$$\hat{V}(\hat{\tau}_h) \cong 67,886,303,000,000.$$

$$\sqrt{\hat{V}(\hat{\tau}_h)} \cong 8,240,000$$

Standard deviation = $\sqrt{\hat{V}(\hat{\tau}_h)} \cong 8,240,000$

$\hat{\tau}_h \pm \sqrt{\hat{V}(\hat{\tau}_h)} =$ 59,565,000 ± 8,240,000 pixels of conifers in the $h^{\underline{th}}$ stratum

or $\hat{\tau}_h \cong$ .3249(59,565,000 ± 8,240,000 $\cong$ 19,352,700 ± 2,678,000 hectares of conifers in the $h^{\underline{th}}$ stratum.

V. Optimal Allocation of Samples

The basic MSS data is taken from the large (497 x 500) blocks. The total number of pixels of conifers in a block is corrected to mean hectares of conifers per 100 x 100 pixel block or, equivalently, mean hectares of conifers per 3249 hectares of ground area. Denoting this mean by $\hat{\mu}_{xhij}$ for the $j^{\underline{th}}$ large block in scene i of stratum h, a mean for the scene is calculated

$$\hat{\mu}_{xhi} = \frac{1}{m_{hi}} \sum_{j=1}^{m_{hi}} \hat{\mu}_{x_{hij}} .$$

The scene mean is then adjusted to TM data by the regression formula

$$\hat{\mu}_{hi} = \bar{y} + b_{hi}(\hat{\mu}_{x_{hi}} - \bar{x}_{hi}) .$$

The two means $\bar{y}_{hi}$ and $\bar{x}_{hi}$ are, respectively the means of the TM and the MSS data measured in the small blocks within each large block in scene i, stratum h. These are used to estimate b_{hi} to adjust the MSS data to TM data.

Then for the stratum, these are averaged to give the estimated adjusted stratum means in hectares of conifers per 3249 hectares of ground area

$$\hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{\mu}_{hi} ,$$

The variance of this estimate is

$$\mathrm{Var}(\hat{\mu}_h) = \frac{N_h - n_h}{N_h n_h} \sigma_B^2 + \frac{1}{N_h n_h} \sum_{i=1}^{n_h} \frac{M_{hi} - m_{hi}}{M_{hi} m_{hi}} \sigma_W^2$$

The population variances in this formula can be thought of as the variance between scenes in stratum h

$$\sigma_B^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (\hat{\mu}_{hi} - \hat{\mu}_h)^2$$

and the variance of blocks within scenes,

$$\sigma_W^2 = \frac{1}{M_{hi} - 1} \sum_{j=1}^{M_{hi}} (\hat{\mu}_{hij} - \hat{\mu}_{hi})^2 .$$

We will assume in what follows that the latter variance is the same for every scene in the stratum.

We will assume, in addition, that because the number of large blocks in a scene is the same for every scene, that the sample number of blocks per scene m_{hi}

is constant for all strata.

Suppose the costs of selecting scenes for a sample is the same for each scene in stratum h and equals C_S per scene. The cost of selecting large blocks, observing data in them and calculating the adjustments, is assumed to be the same for each block and equal to C_B per block. We assume there is no other cost. Then the cost of sampling in stratum h is

$$C_h = C_S n_h + C_B n_h m_h \; .$$

(Since the number of large blocks is the same for each scene, the cost does not depend on the scenes sampled and is therefore not random.)

Using a result found in Cochran, (p. 313, 1977) the allocation of n_h and m_h which minimizes the variance for a fixed cost is obtained by taking

$$m_h = \sqrt{\frac{C_S}{C_B} \quad \frac{\sigma_W^2}{\sigma_B^2 - \frac{1}{M_h}\sigma_W^2}}$$

One may obtain an estimate of m_h by using the data on page 16, 18 and 20 to estimate σ_W^2 and σ_B^2. Using analysis of variance, one obtains

$$\hat{\sigma}_B^2 \cong 1{,}700{,}000$$

and $\hat{\sigma}_W^2 \cong 880{,}000$.

Next assuming that

$C_S = \$730$, $C_B = \$100$

$\sigma_W^2 = 880{,}000$ $\sigma_B^2 = 1{,}700{,}00$

and there are $30 = M_h$ large blocks in a total scene, we obtain

$$m_h = \sqrt{\frac{C_S}{C_B}\left(\frac{\sigma_W^2}{\sigma_B^2 - \frac{1}{M_h}\sigma_W^2}\right)}$$

$$= \sqrt{\left(\frac{730}{100}\right)\left(\frac{880,000}{1,700,000 - \frac{1}{30}(880,000)}\right)}$$

$$m_h \cong 1.94 \text{ or } \underline{\underline{2}}$$

The value of n_h is found by setting the desired total cost, C_h, and we obtain

$$n_h = \frac{C_h}{C_S + C_B m_h} \quad .$$

It must be stressed here that sample size calculations (for this problem dealing with adjusted means using regression estimates) should consider the estimate of the regression coefficients within scenes. In other words, the number of large blocks, with the center small block representing it, should be approximately 12 per scene even though the optimum allocation shows fewer large blocks per scene. The reason for suggesting 12 is that the degrees of freedom for residual is 10 and the theoretical F with 1 and 10 degrees of freedom with the probability of Type I error at 5% is just under 5 (Anderson and McLean p. 394, 1974). It is most desirable to have a significant linear effect if one is to use the linear regression adjustment on the estimated mean.

VI. References

1. Anderson, V. L. and McLean, R. A., "Design of Experiments: A Realistic Approach," Marcel Dekker, Inc. (1974).

2. Cochran, W. G., "Sampling Techniques," John Wiley and Sons, Inc. (3rd edition 1977).

3. Horning, N., Nelson, R. and Case, D., "A Determinination of Sampling Intensity to Characterize a Landsat MSS Scene Using Two Block Sizes," Internal Review NASA/GSFC (1985).

4. Nelson, R., Case, D. and Horning, N. "Forest Inventory of the Continental United States Using Landsat Data," Internal Review NASA/GSFC (1985).

5. Ostle, B. and Mensing, R. W. "Statistics in Research," Iowa State University Press(3rd edition, 1975).

6. Schaeffer, R. L., Mendenhall, W. and Ott, L., "Elementary Survey Sampling," Duxbury Press (1979).