82-12

United States
Department of
Agriculture

Statistical
Reporting
Service

Staff
Report No.
AGES 820526

Washington, D.C.

June 1982

Construction of a Remotely Sensed Area Sampling Frame for Southern Brazil

Ron Fecso

Wayne Gardner

Bob Hale

Van Johnson

Steve Pavlasek

CONSTRUCTION OF A REMOTELY SENSED AREA SAMPLING FRAME FOR SOUTHERN BRAZIL, by Ron Fecso, Wayne Gardner, Bob Hale, Van Johnson and Steve Pavlasek. Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. SRS Staff Report No. AGES820526. June 1982.

ABSTRACT

A remotely sensed area sampling frame was constructed for selected areas in southern Brazil. The sampling unit information was stored in digital form in a latitudinal/longitudinal characterized population. Computerized sampling procedures were developed which allow for flexibility in sample unit specifications and sampling designs.

Keywords: Area frame, land use stratification, remote sensing, sample unit size, sampling design.

ACKNOWLEDGEMENTS

We want to thank the participants from the following groups for their support and contributions:

1. Sampling Frame Development Section (SFDS), Statistical Research Division (SRD), Statistical Reporting Service (SRS), U.S. Department of Agriculture (USDA).

2. Remote Sensing Branch, SRD, SRS, USDA.

3. Earth Resources Group, National Aeronautics and Space Administration (NASA), - Johnson Space Center.

We are grateful to Bob Payne, Lockheed Corporation, and Dale King, SRS for their assistance in LANDSAT signature interpretation, Dave Frank for coordinating contacts with NASA officials, and Dave Hicks, Environmental Research Institute of Michigan (ERIM) for assistance in obtaining ground truth information. Dr. Frederick C. Westin, Remote Sensing Institute, South Dakota State University provided soils maps, while Ed Bulloch, Foreign Crop Condition Assessment Division, Foreign Agricultural Service, USDA and Horace Huckle, Soil Conservation Service, USDA provided additional interpretation of crop codes. Thanks also goes to Norm Beller and Bob Tortora for their overall management and technical support, and Karen Watkins and Kathy O'Donnell for their fine typing efforts.

CONTENTS

SUMMARY

The Statistical Reporting Service assumed the responsibility for construction of a remotely sensed area sampling frame for selected areas in Southern Brazil. Original specifications for frame development contained a highly manual process. A transparent grid overlay was to be used to determine sample unit location and stratification characteristics. The Sampling Frame Development Section was concerned with the following aspects of the original frame development specifications: (1) burden of manual imput, (2) restrictions on the sample unit size, (3) nonsampling errors, and (4) strata boundaries and definitions.

The paper describes an automated approach to frame construction which was developed by the section. Noteworthy differences, from the original procedures include: (1) a data base approach, (2) reduced manual input, (3) ability to change sample unit size, and (4) improved stratification ability.

Construction of a Remotely Sensed Area Sampling Frame for Southern Brazil

Ron Fecso
Wayne Gardner
Bob Hale
Van Johnson
Steve Pavlasek

INTRODUCTION

The Statistical Reporting Service (SRS) has assumed the responsibility for constructing the sampling frames for the Agriculture and Resources Inventory Surveys through Aerospace Remote Sensing (AgRISTARS) program.

The AgRISTARS program is a joint undertaking of the United States Departments of Agriculture (USDA), Commerce (USDC), and Interior (USDI), the National Aeronautics and Space Administration (NASA) and the Agency for International Development (AID). Development, planning, and execution of the frame construction was done by the Sampling Frame Development Section (SFDS), Sampling Frames and Survey Research Branch (SRB), Statistical Research Division (SRD) at the Fairfax location.

The task was to construct a remotely sensed area sampling frame for selected areas in southern Brazil (The states of Parana, Rio Grande do Sul, and Santa Catarina.). The frame and an associated computerized system were constructed in a manner which will allow digitally stored sample unit information and computerized sample selection. Sampling procedures and software were developed for latitudinal/longitudinal characterized populations in a way which permits variable sampling specifications.

ORIGINAL FRAME DEVELOPMENT PROCEDURES

The definition of the sample units prior to this work had been a manual process. A transparent grid overlay was made to the desired cluster size and placed over the associated base map (usually an Operational Navigation Chart). The sample units were then defined as the area enclosed by the individual grids. At this point, each grid was labeled for identification and assigned some auxiliary information for use in the creation of a stratified sampling design. The auxiliary information was based on visual interpretation of LANDSAT 1:1,000,000 transparencies which were "registered" to a base map by aligning observable physical features common to both products. The information was percentage of agriculture estimated to the nearest ten percent.

There are a number of problems associated with sample unit determination. These problems are now listed for reference when describing the benefits of the automated system.

1. Burden of Manual Input

Considerable manual input is necessary in the current procedures. Overlays need to be drawn according to the variable base map scales. Manual interpretation of the auxiliary information is necessary for a very large number of sample units. The accuracy of this control data may become suspect if the burden of time requirements forces quick estimation for many of the sample units. The relatively small cluster size may be difficult for image interpretation at a 1:1,000,000 scale. And not trivially, each of the many sample units will have to have a record manually created with identification and control information present.

2. Restriction on Sample Unit Size

The manual procedure produces clusters of a given size. If future research shows that considerable efficiency can be gained from the use of a different size, almost all the manual work will have to be redone to change to the new cluster size.

3. Nonsampling Errors

Potential causes of nonsampling error include overlay distortion and record keeping errors. Also changes in scale between base maps might make complete coverage of the population difficult.

4. Ineffiicient Strata Boundaries

In the manual procedure, the strata boundaries are defined by the land use of the clusters and the fixed edges of the clusters. If the cluster size is changed the boundaries may not be efficient or even useable.

SAMPLING CONCEPTS

The problem of AgRISTARS' frame development and sample selection is considerably different from the area sampling frame methodology used by SRS in its U.S. agricultural estimation program. The programs differ mostly in the amount of labor intensiveness. The SRS program requires manual collection of ground truth information which necessitates clearly recognizable boundaries for all sample units. As a result, the frame development and sample selection require a great deal of manual input. Since the AgRISTARS program will not require physically observable sample unit boundaries, such as roads, rivers, and so forth, the frame construction and sampling procedure has the potential for considerable automation. With large areas needing stratification, automation becomes an operational necessity.

A stratified design was determined to be best suited to meet the objectives of the AgRISTARS programs. This design allows for an efficient distribution of a limited sample size to areas which are homogeneous in crop density and, as much as possible, in yield potential of the various crops being estimated. Since there is a possiblity that sample unit sizes or stratification procedures may change with continuing research, the development of an automated system which would be used to classify sample units into strata is advantageous.

The first task to provide land use stratification was to determine which characteristics associated with the desired estimates can be quantified based on the available materials. We had to determine the characteristics which could be identified through inspection of LANDSAT imagery, in conjunction with topographic and soils maps. These characteristics should aid in the improvement of the sampling design, allocation, development of LANDSAT interpretation, signature extension, and yield model development.

The availability of LANDSAT imagery for the target region in Brazil was determined. Paper and film products with various scales were ordered for selected sites and acquisition dates were staggered through the growing season. In addition to the imagery, the following materials were acquired:

1. Operational Navigation Charts (ONC) at a 1:1,000,000 scale from the Aeronautical Charts and Information Command (ACIC) for use as a base map in registration of the frame and LANDSAT scenes.

2. Soils maps for use in construction of the land use strata (LUS).

3. Maps of the smallest political subdivisions (SPD).

The following materials were also used when available:

1. Topographic maps from the Defense Mapping Agency (DMA).

2. Land use maps.

3. Geological maps.

4. Meteorological (temperature and precipitation) data from the National Oceanic and Atmospheric Administration (NOAA).

5. Crop statistical data.

6. ONC base physical features overlays from ACIC.

Base map overlays when not available from ACIC, were made by the stratification group of SFDS. These overlays have major physical features such as rivers, lakes, and major road systems. The base map and its overlay contain enough information to register the LANDSAT imagery for the area being stratified.

The next step was to determine the useful auxiliary data which can be estimated from inspection of the materials. "Useful auxiliary data" in the statistical sense is information which can be used to reduce the error of the estimates through sampling efficiencies. This error reduction can be accomplished through the use of a stratified design, probability proportional to size sampling, or through more efficient estimators such as ratio or regression types. For this application, the search was restricted to information useful for stratification with respect to estimates and forecasts of corn and soybean production.

The Brazil area frame was to provide the basis to estimate and forecast the production of corn and soybeans. The estimator of production for either commodity is the product of an acreage estimate and a yield per acre estimate. Thus, the information necessary for an effective stratification of the area frame was the density of the crop in the sample unit and the yield potential for the crop in the area of the sample unit. The importance of items used for auxiliary information depends on the items' influence on the variability of the production forecast. Experience in the United States shows that the yield forecast contributes heavily to the variability of the production forecast. Research, conducted as part of the Large Acreage Crop Inventory Experiment (LACIE), suggested that this relationship holds globally.

The following questions were considered in trying to achieve an effective stratification:

1. What would the sample unit be?

2. Which characteristics could we use to stratify?

3. How could we stratify for more than one forecast item?

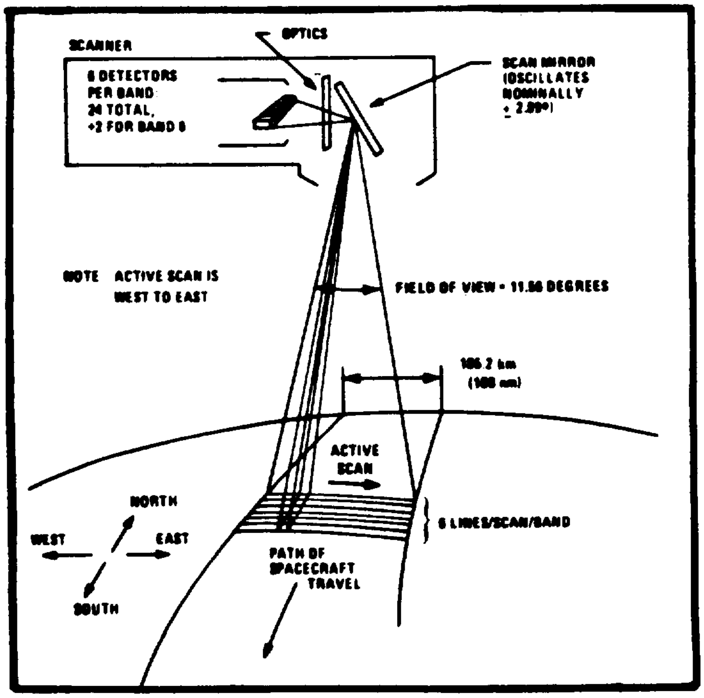
4. How could we use the auxiliary information to create strata?

The grouping of sample units into strata could only be accomplished with some knowledge of the sample unit content. Therefore, before pursuing stratification the sampling unit had to be defined. Then the stratification was merely a clustering of sample units in an attempt to reduce the heterogeneity within each stratum with respect to the overall population.

The reporting unit was predefined as the remotely sensed information corresponding to a LANDSAT "pixel." Thus, the elements of the population of interest consisted of the pixels for the areas within the Brazilian states of Parana, Santa Catarina, and Rio Grande Do Sul.

Since an area the size of a LANDSAT pixel (57 by 79 meters) can be considered flat, the elements of the population of interest was defined to be the land areas associated with each pixel. Thus, disregarding the roll, pitch, and yaw of the satellite, there is approximately a one-to-one correspondence between 57 by 79 meter disjoint pieces of the earth and LANDSAT pixels (see Figure 1).

Figure 1--Multispectral scanning arrangement



SOURCE: LANDSAT USERS GUIDE

The sampling frame for the area of interest was, thus, defined as 57 by 79 meter rectangular areas (elements) which do not overlap and which completely cover the area of interest. Note that the boundary of the frame will become altered to match the rectangular shape of the pixel or as later seen the rectangular shape of the sample units. The reporting unit for each element is the LANDSAT pixel with a centerpoint closest to the centerpoint of the element.

The actual frame from which samples will be drawn consists of the map corrected images now produced by NASA. Basically, the original pixel data is "resampled" by transforming the pixel locations to a 57 by 57 meter grid system. The technique used can be found in the papers by Williamson (24) and Donovan (5). 1/

The frame used for AgRISTARS sampling consists of the 57 meter square areas (elements) of land which correspond to the resampled frame pixels described above. Thus, all land is conceptually included in an element and no land is in two elements. The resampled MSS pixel information for the pixel closest to the land element is the reporting unit.

1/ Underscored numbers in parentheses refer to literature cited at the end of this report.

Current software limits the flexibility of processing the LANDSAT data. Therefore, the sampling design for Brazil was restricted to one in which the sampling units contain clusters of elements which are rectangular arrays of 193 by 117 elements, i.e. pixels. This limitation is expected to be eased to allow a rectangular array of any size for future countries. Thus, a frame permitting flexibility of sample unit size would allow the implementation of research findings on optimal cluster size. This research is being conducted by USDA/SRS and NASA.

The AgRISTARS sampling unit was defined as a general rectangular array L meters long by H meters high where the L and H are determined by

$$L = (N_l)(P_l), \text{ and}$$

$$H = (N_h)(P_h)$$

where

P_l = pixel length (in meters),

P_h = pixel height (in meters),

N_l = number of pixels along the length of the sample unit (S.U.), and

N_h = number of pixels along the height of the sample unit.

An array of elements of the earth's surface which had a corresponding array of reporting units (pixels) was defined as the sampling unit. these sample units must be defined in the area of estimation in such a way as to include all elements in a sample unit while not including any in more than one sample unit. The methodology which will uniquely define the sample units is described in the section on sampling frame development.

CONSTRUCTION OF THE LAND USE DATA BASE

The characteristics used to stratify the sampling units were kept independent of the sample units in a land use data base. To develop the data base, homogeneous land use blocks were recorded in digital form. These blocks were defined by a vector of information which contained variables for percentage of land cultivated, percentage of cultivated land which is utilized for corn and soybeans, field size, and soil type. The data for the vectors of information were developed by using LANDSAT color infrared imagery (CIR) with crop statistics for the smallest political subdivisions (SPD) used to help interpret the LANDSAT signatures.

Information for major crop items was available at the SPD (Municipio) level, but total land area was not. 2/ The estimates for

2/ Individual commodity estimates were available for corn, soybeans, rice, Kidney (dry) beans, cotton, and sugar. The remaining crops such as coffee, and wheat were included as other crops.

major crops were made by the Foreign Agricultural Service (FAS) and were obtained from the data base at the Laboratory for Applications of Remote Sensing (LARS). Total land area was estimated by digitizing the Municipio boundaries shown on maps supplied by NASA. 3/ The percentage of cultivated land was computed by dividing the sum of all crop estimates by the total land area. The percentage cultivated land devoted to corn and soybeans was computed by dividing areas planted to those crops by the sum of all crop area estimates.

Crop calendars, Operational Navigation Charts (ONC), soil data, and climatological data were also obtained to help interpret the LANDSAT signatures. Information on the geography and cultural practices of southern Brazil was used to provide a starting point for constructing the land use blocks.

Constructing Land Use Blocks

The land area of the three states was broken down into 77 "work units" or "one degree squares" to provide for an even workload flow. 4/ These 1° squares define an area of land within a "quadrangle" formed by the intersection of pairs of lines of latitude and longitude. Data base information for political subdivisions and land use was developed within each 1° work unit. The use of 1° work units increased the accuracy of registration and provided a data base which could be managed at reasonable cost with the automated system.

Base maps and soils maps for each 1° square were drafted using ONC's and information prepared by Dr. Westin of South Dakota and provided by NASA. A cartographer (carto) placed the base maps on the LANDSAT scenes using rivers or other visible landmarks for orientation. Then, using the ancillary data and crop calendars to support the interpretation of the LANDSAT signatures, areas having similar characteristics such as field patterns, bare soil, and so forth were delineated. The carto then coded the areas according to density of cultivation.

After the initial breakdown and coding were complete, an analyst reviewed the first carto's work, made necessary changes and additions to the percentages cultivated, and coded the percentages of corn and soybeans. This work unit then went back to a carto who made the corrections to land use boundaries, and modified boundaries due to varying soil types. Field size also was coded in this step.

3/ Several problems were encountered; 1) the maps showing Municipio boundaries had to be drafted; 2) the maps arrived late in the construction process, 3) boundaries had to be estimated where maps overlapped, 4) a few Municipios were miscoded, and 5) eight Municipios were omitted entirely. See Appendix 4 for additional information on data summarization and the associated problems.

4/ For a more complete description of the procedures for constructing the frame see the outline of the stratification procedures, the worksheets, and coding sheets included as appendices 1-3.

Following the second carto step the work unit was again reviewed by an analyst. This review not only checked for completeness and consistency within the work unit, but also checked for consistency between work units. (The analyst reviews will be discussed further in the section on quality control).

Once the review was complete, the final frame was drafted and all land use blocks coded. The frame was checked for accuracy and given a final review by a statistician. Once the final frame was approved it was ready for digitization.

<u>Creating The Data Base</u>

At this point all the necessary information was available to create the frame. However, this information needed to be converted into computer readable media. To accomplish this transformation graphic coordinate digitizers were used.

The digitizing equipment consists of:

1. A tablet with a wire grid underneath to determine the location of each point on the tablet.

2. A 16 key cursor to enter information.

3. A microprocessing controller.

4. A dual floppy disk system to record and store the digitized information.

Before each map or overlay was digitized, dots were placed on the boundaries of every polygon whenever there was a noticeable change of direction. Therefore, curved lines were represented by a group of straight line segments. Each municipio or land use area was represented by an N sided polygon where N was the number of dots needed to outline the area.

Each point on the boundary of the polygon was represented by its x-y coordinate system and was identified by polygon control and work unit identification numbers. Additionally, information for machine edits was recorded for each point. The polygon control number includes coding which defines the vector of information. Hence, the vector of information for each land use block was associated with every point determining that polygon.

The final frame was digitized by work unit and four files were created. These files, (a point file, area file, calibration file and calibration area file) were used to check the digitization and verify the correctness of the point file. 5/ Points in the point file could be associated with more than one land use polygon. Hence, these matching points would be recorded more than once. Matching points were averaged and the average value assigned to each polygon containing a matched point. The point averaging was done to ensure that no small

---

5/ Documentation for the software used to edit and convert the points is included as appendix 5.

areas were omitted or duplicated since very rarely will any point be recorded twice with exactly the same x and y coordinate.

The final step in creating the data base was to convert the x and y coordinates of each point into latitude and longitude. The four points in the calibration file were used to determine a linear relationship between the x and y coordinates and latitude and longitude. This relationship was then used to convert all the points in the point file to the corresponding latitude and longitude. The file was stored on magnetic tape for later use in classifying the sample units.

<u>Quality Control</u>

Quality control measures took on several forms throughout frame construction. Analysts and reviewers visually checked the construction of land use blocks to ensure consistency in interpreting the LANDSAT signatures. A data base of crop estimates for the political subdivisions (municipios) was used to help interpret LANDSAT signatures and to obtain characteristic values for land areas where scenes were missing or the imagery was obscured by cloud cover. Machine edits, check digits, and the point averaging program were used to ensure that the information in the digitized data files were complete and correct. Finally, the frame was evaluated using comparisons with ground truth data. This information was obtained from field observations made during travel through the three states in Brazil.

The analyst reviews were not only steps in constructing the frame, but also acted as quality control measures. This process uncovered two basic problems with the LANDSAT imagery. First, the quality of the paper products received from NASA was very poor. Colors and tints varied considerably from scene to scene. Often crop signatures were undiscernible, even when there was reasonable multitemporal coverage. Multitemporal coverage in itself was a problem because many areas had only scenes stretched over several years. Second, registration of the 1° work unit to the LANDSAT scenes was approximate. Both the work unit overlay and the LANDSAT scenes were on a scale of 1:250,000, but when the overlay was placed on the LANDSAT scene, the rivers, roads and so forth were not aligned except in small areas. Consequently, as the stratification work was being done, the overlay had to be moved around the LANDSAT scene to achieve correct registration for the small area being stratified. Discussions between the cartos and analysts in resolving these problems and other difficulties led to a more consistent interpretation of the signatures.

The second analyst review was used as a check for consistency among work units. In this step the land use blocks for adjacent work units were compared. Blocks which had identical codes for percentage of land under cultivation were expected to have similar signatures on the LANDSAT scenes. If the signatures were not consistent, codes were adjusted and/or boundaries redrawn to improve the homogeneity of the land use blocks.

To assess and improve staff capabilities of interpreting LANDSAT imagery outside expertise was sought. Bob Payne, with Lockheed in Houston, visited the Fairfax office and provided a critque of completed work and additional training in interpretation. Dale King, now the SRS, USDA, but formerly an analyst with Lockheed, also reviewed the work.

Machine edits were used for quality control in the digitizing process. One program added the areas of the land use polygons and compared this value to the total area of the work unit. If the two areas were not within a tolerance of $\pm.2$ percent, the digitized file was checked for accuracy and corrected. A check digit, consisting of the units value of the sum of the stratification codes, was included to check that codes were correctly entered into the data base. Finally, the point averaging program ensured that no land area was omitted from the frame or was duplicated.

Although one of the constraints of this frame construction project was that a visit to Brazil prior to the work was not possible, a trip was made to evaluate the work after it was completed. Wayne Gardner and Van Johnson made the evaluation trip to Brazil, (February 2-20, 1981). They were accompanied by Dave Hicks, a geographer with the Environmental Research Institute of Michigan (ERIM). Comprehensive results of that trip are available in an informal trip report by Gardner and Johnson (9) 6/. Basically, 49 locations were visited within the three Brazilian States and observations recorded. These observations were compared to the frame values for percentage of land cultivated, percentage of corn and soybeans and field size. These comparisons are shown in figures 2-4.

If the frame values and observed values matched exactly, only the blocks along the diagonal (bottom left to top right) of the charts would contain positive values. Since perfect correspondence of values is not expected, a high percentage of the observations falling in the blocks on the diagonal or one block either side of it would provide an indication that the frame stratification was "good." This assumes that an observation on the low end of a given range could fall in that range or in the preceding range or vice-versa. "Good" means that stratification should result in an overall increase in the efficiency of the sample design.

Notice in Figure 2, that 26 of the 49 observations are on the diagonal, and an additional 16 are on either side. Thus 86 percent of the observations are within the specified limits. The comparisons in Figure 2 suggest that stratified land use in the lower ranges of percentage cultivated more often matched the observed values than did the land use classification in the areas of more intensive cultivation. This would seem to indicate that stratifiers became more conservative in their interpretation of LANDSAT signatures as the intensity of cultivation increased.

6/ Additional comments can be found in (11)

FIGURE 2: COMPARISON OF OBSERVED TO FRAME PERCENTAGE CULTIVATED FOR BRAZIL

FREQUENCY BLOCK CHART

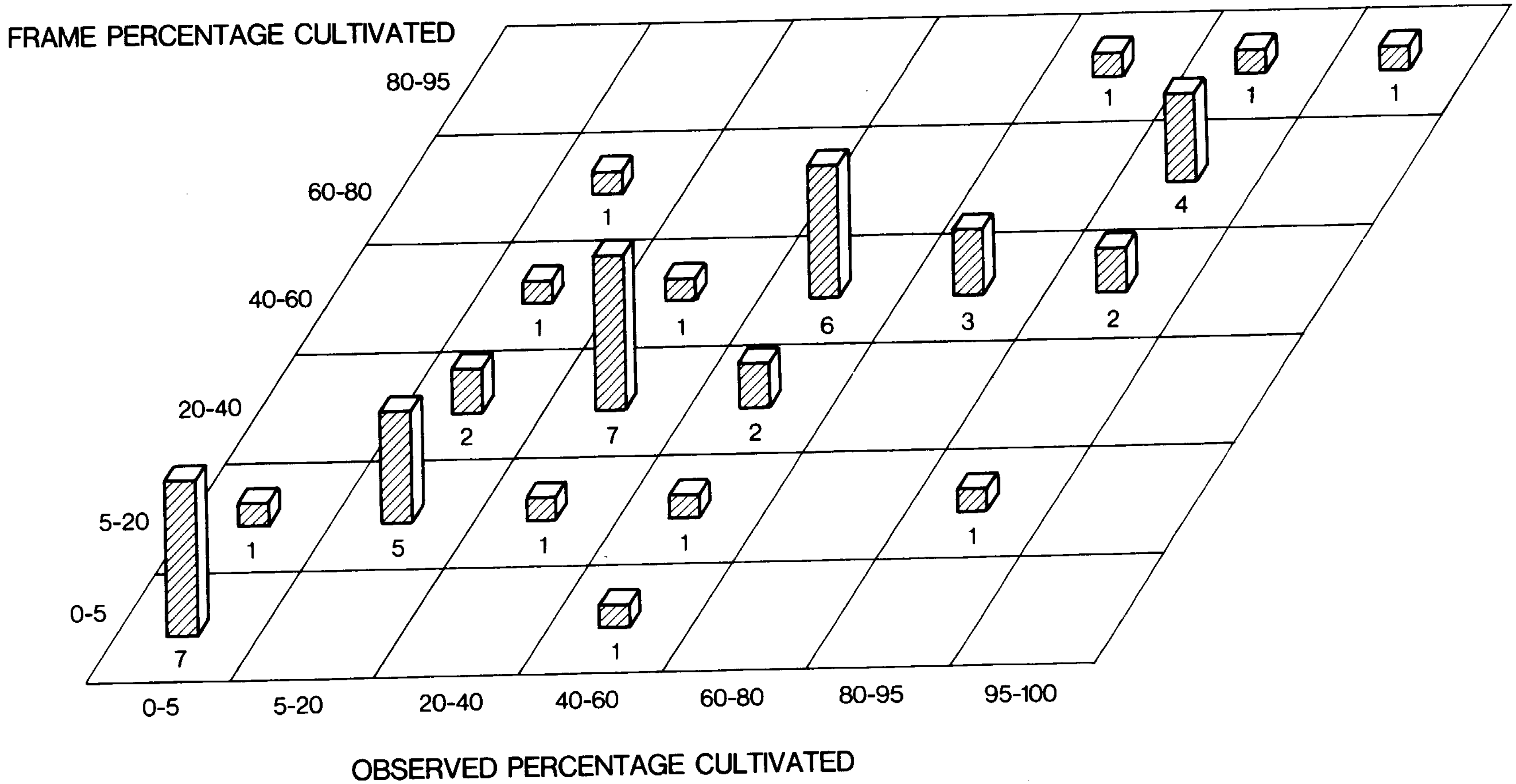


FIGURE 3: COMPARISON OF OBSERVED TO FRAME PERCENTAGE OF CORN/SOYBEANS FOR BRAZIL

FREQUENCY BLOCK CHART

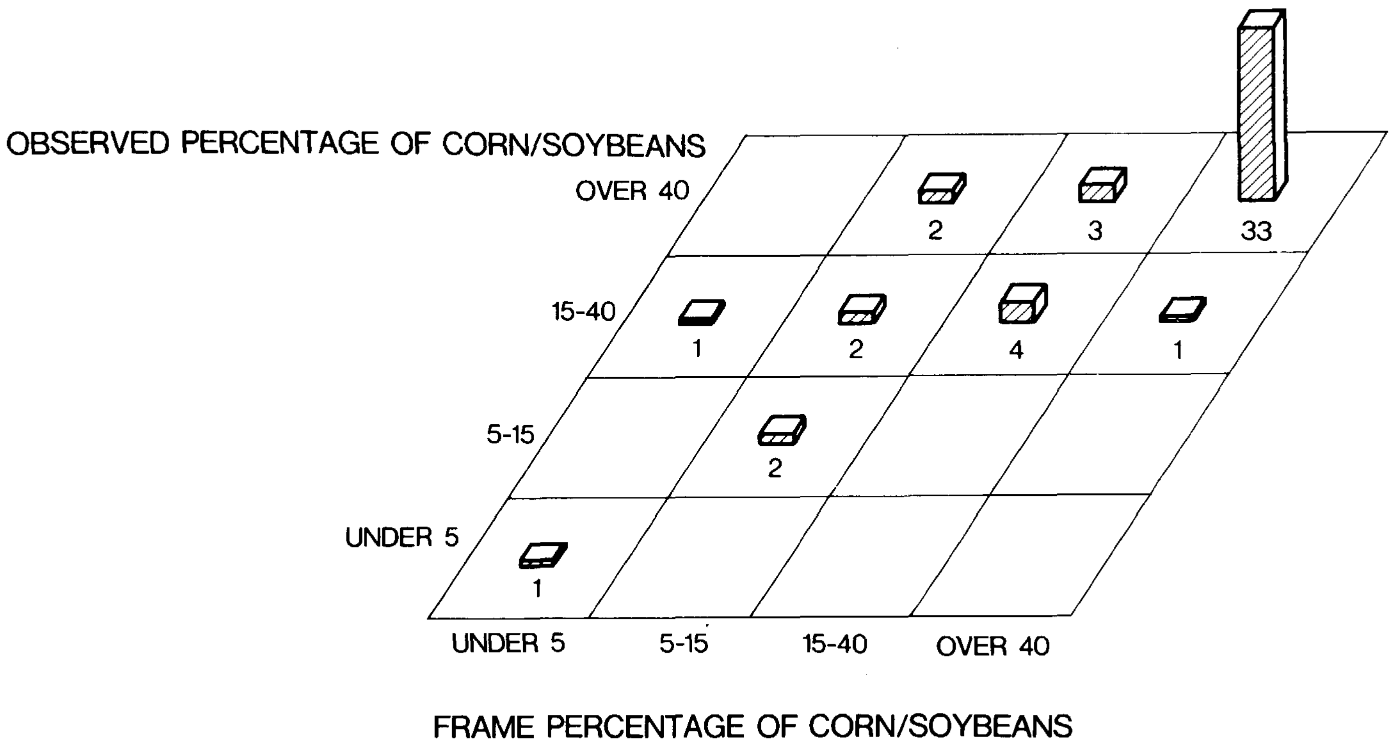


FIGURE 4: COMPARISON OF OBSERVED TO FRAME FIELD SIZES FOR BRAZIL

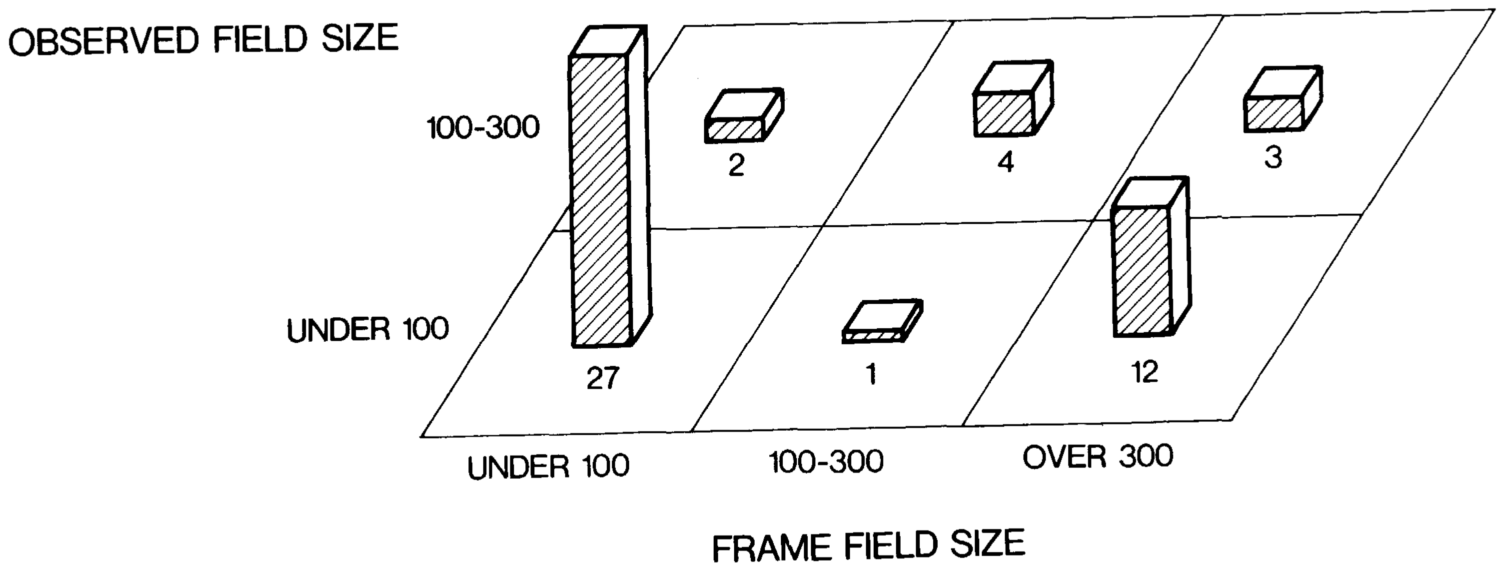


Figure 3 shows the comparison of percentage of corn and soybeans to total cultivated land. Forty of the 49 observations are in the blocks along the diagonal. The data in this table reflect the wide range of the percentage cultivated where almost 80 percent of the observations had values greater than 40 percent. Analysts could not determine the percentage of corn and soybeans from the LANDSAT signatures alone. Therefore, ancillary data was used as the major indicator of the percentage of corn and soybeans in each stratum block. These comparisons indicate that the use of ancillary data was generally accurate for coding the percentages of corn and soybeans.

Figure 4 compares field size. Fields larger than 300 acres were not observed in the selected sites. However, some fields which may have been in that range were seen while traveling from one location to another. Most of the fields observed were small (less than 100 acres). The data indicate that stratifiers had trouble in determining field size from the LANDSAT scenes.

The comparisons of observed to stratified values were generally very good. A few areas where the stratification was quite a bit off were noticed. Most of the gross misstratification was unavoidable because the signatures of some growing crops, such as sugarcane, could not be distinguished from natural vegetation.

Also, the random planting of fields in vast plains areas caused and will continue to cause problems in determining the percentage of land cultivated in these areas.

The interpretation of signatures which evolved from the stratification process and which is shared by the analysts and reviewers is consistent with actual conditions. For example, signatures which were coded as bare soil did represent bare soil; areas of cultivation were cultivated, and so forth. However, signatures of some growing crops will continue to be confused with those of native grasses.

The ancillary data for Brazil provided good indications of relative levels of percentages cultivated, but should not be relied on solely for absolute levels. The percentage of corn and soybeans to total crops, computed from the ancillary data, appeared to be a good indicator of the actual percentages. Hence, the ancillary data was reliable as an indicator of relative values.

SAMPLING FRAME DEVELOPMENT

Defining the sampling frame required that a program be developed which would compute the latitude and longitude of the center point of each sample unit in the target area. The geometric arguments used in this program are contained in Appendix 6. Basically the algorithm is an iterative procedure which takes into account the curvature of the earth and the size of the desired sample unit when positioning the rectangular sampling units over the target area.

Assigning Auxilliary Information to the Sample Units

A computer was used to define the sampling frame. This use of software to create the frame allows freedom in the choice of sample unit (cluster) size. In order to stratify the population, a method was devised to assign auxillary information to the sample units. This objective was accomplished through programming.

Software was written to assign the auxiliary information to each sampling unit in the frame. The location of the sampling unit was compared to the latitude/longitude polygon files which contain land use and political subdivision information. The information assigned to a sample unit was determined by the location of the center point of the sample unit. The program used to assign the auxiliary data is in Appendix 5.

Originally the auxiliary information for a sample unit was to be created by allocating polygon information to a sample unit in proportion to the amount of the polygon in the sample unit. However, there were many inherent problems with materials which would cause difficulties in registration. For these reasons the sample unit was assigned the auxiliary information which corresponds to the polygon in which the center point of the sample unit falls.

SAMPLING DESIGN

The Brazil sampling frame is a data file which contains the center points of the cluster sampling units in the three state area of interest. Each of these sampling units also contains political subdivision and land use information.

A number of sampling designs are possible with this type of file. These designs include:

1. Simple random sampling.

2. Stratified sampling by political subdivision and/or crop characteristics.

3. Sampling with probability proportional to size (PPS) based on a cropland or crop-specific estimate.

Additionally, the use of alternative estimators is a possibility. These estimators could include both ratio and regression types. Stratification of the sampling frame based on the auxiliary information is also being considered by a number of organizations, including the SFDS.

BIBLIOGRAPHY

(1) American University, Foreign Area Studies. Area Handbook for Brazil, 3rd. ed. GPO, 1975, pp. 13, 329-343.

(2) Butland, Gilbert J. Latin America: A regional Geography 3rd. ed. New York: John Wiley and Sons, Inc. 1972, pp. 406-419.

(3) Commissao Especial de Planejamento, Controle e Avaliacao das Estatisticas Agropecuarias. "Levantamento Sistematico da Producao Agricola," Instituto Brasileiro de Geografia e Estatistica. Rio de Janeiro, 1974-77 editions.

(4) Colvocoresses, Alden P. "Space Oblique Mercator," Photogrammetric Engineering and Remote Sensing, (1974), pp. 921-6.

(5) Donovan Walt. "Oblique Transformations of ERTS Images to Approximate North-South Orientation," Urbana, Illinois: Univ. of Illinois, Center for Advanced Computation technical memorandum No. 38, (1974).

(6) Fecso, Ron. "Stratification of a Remotely Sensed Area Sampling Frame," Proceedings of the Survey Research Section, American Statistical Assoc., (1981).

(7) Fecso, Ron and Van Johnson. The New California Area Frame: A Statistical Study. U.S. Department of Agriculture, Statistical Reporitng Service, No. 22, (1981).

(8) Fundacao Getulio Vargas, Instituto Brasileiro de Economia, Calendario Agricola Nacional, Rio de Janeiro, June 1975. p. 76.

(9) Gardner, Wayne and Van Johnson "Brazil AgRISTARS Area Frame Evaluation-Trip Report for February 2-20". Unpublished working paper March 1981.

(10) Gulliver, K.L., G.W. Williams, R.L. Thompson and M.E. Revelt, "The Brazilian Soybean Economy, An Econometric Analysis," Prepared for the Country Market Studies Program, Foreign Demand and Competition Division, Econimics, Statistics and Cooperatives Service, U.S. Department of Agriculture, (1979).

(11) Hicks, David R. "Notes for Brazil Sampling Frame Evaluation Trip." AgRISTARS Supporting Research, ERIM, Tech. Rept. SR-EJ-04138, Aug. 1981.

(12) Houseman, Earl E., Area Frame Sampling in Agriculture, U.S. Department of Agriculture, Statistical Reporting Service, No. 20, (1975).

(13) Huckle, Horace F. "Taxonomic Classification of World Soil Map Units Occurring in Selected Brazilian States with Representative U.S. Soil Series and Numberical Rating of Physical and Chemical Soil Properties Significant to Crop

Production." Unpublished working paper. U.S. Dept. Agr., Soil Conservation Service, 1979.

(14) INCRA, as reported in Ruy Miller Paiva, and others Brazil's Agricultural Sector, Economic Behavaior, Problems and Possibilities, Sao Paulo, 1973, p. 297.

(15) Instituto Brasileiro de Geografia e Estatistica. Censo Agricola de 1960-Brasil, Serie Nacional, vol. II, 2nd part, July 1970, p. 69.

(16) ________ Censo Agropecuria 1970, Rio de Janeiro 1974, Table 57 in each state volume.

(17) McBryde, F.W. and P.D. Thomas, Equal-Area Projections for World Statistical Maps, U.S. Department of Commerce, Special Publication No. 245, (1949).

(18) Ministerio da Agricultura, Subsecretaria de Planejamento e Orcamento. Serie Estatisticas Agropecuarias, volumes 14, Dec. 1977, pp. 17-22.

(19) Seltzer, Leon E. (Ed.). Columbia-Lippincott Gazeteer of the World. New York: Columbia Univ. Press, 1962.

(20) Sukhatme, P.V. and B.V. Sukhatme, Sampling Theory of Surveys with Applications, AMES, Iowa: Iowa State University Press, (1970).

(21) U.S. Department of Agriculture. "Brazil: Special Soybean Report." F.A.S. Agricultural Attache, Sao Paulo, various reports.

(22) U.S. Foreign Agricultural Service Estimates of Area Harvested of Soybeans, by State Brazil, 1971-72.

(23) U.S. Geological Survey, LANDSAT Data Users Handbook, 1979.

(24) Williamson, A.N., "Corrected Landsat Images Using a Small Computer," Photogrammetric Engineering and Remote Sensing, (1977), pp. 1153-9.

(25) Worldmark Encyclopedia of the Nations, New York: John Wiley and Sons, Inc. 1976., pp. 39-48.

(26) Zockon, Maria Helena G. P. A Expansao da Soja ne Brazil: Alguna Aspectos da Producao e Consumo, vol. I. (Sao Paulo: Fundacao Instituto de Pesquisas Economicas 1978). pp. 144, 148.

APPENDIX 1

STRATIFICATION PROCEDURES FOR AgRISTARS AREA FRAME

I. Project Information:

1. Country: Brazil

2. Area(s): States of Parana, Santa Catarina and Rio Grande do Sul

3. Requirements: Stratify the following variables:

   (1) a. States

       b. Political subdivisions (municipios)-to be entered internally by automated grid software.

   (2) Agricultural land (evidence of cropping pattern)

   (3) Percentage cultivated

   (4) Soils type

   (5) Density of corn and soybeans

   (6) Field size

4. Material: Geographic data, including cultural and climate data

   (1) Agricultural statistical data

   (2) Crop calendars

   (3) Soils maps

   (4) Satellite imagery

   (5) Maps and map transparencies

Primary work unit: A one degree latitude-longitude quadrangle outlined on prepared overlays. The overlays have an overlap on all sides which can be used for notes, observations, and so forth.

Restrictions: The area frames for AgRISTARS are to be constructed compatible with an automated grid sampling scheme.

Since no ground truth is available for the AgRISTARS frames, we will not be able to actually determine the accuracy of our stratification. Therefore, it is essential that we be consistent in all of the work. Even if we misclassify a particular variable, we can still have a frame that is efficient for sampling. The true identity of the characteristic will have to be determined through sample data collection or interpretation. Review of individual work to assure consistency throughout the frame is absolutely necessary.

II. Stratification Guidelines/Information Unique to AgRISTARS

1. Color-infrared paper scenes and map transparency overlays with a scale 1:250,000 will be used. The map transparencies will provide stratification variable 1-a (states) and also show main bodies of water and roads for general information.

2. Land use rather than physical boundaries will be emphasized.

3. Each 1^{o} work unit will have a worksheet (Stratification Procedure and Analysis Record) which contains space for identifying the work unit and the LANDSAT scenes used in the analysis and the persons responsible for each step in the process. Space is also provided for comments to aid in the analysis at each of the steps. A work number identifying the block should be entered on the overlay and the work sheet. When written comments refer to a specific block, the blocks should be numbered and the number circled on the overlay. These reference numbers will not be entered into any system in the first land use stratification, but will be used to point out questionable areas to the analyst or reviewer.

III. Stratification Work Procedures

1. First Carto Procedure

(1) Prepare the "Stratification Procedure and Analysis Record" work form for the one degree work unit. Enter all required information completely and legibly.

(2) Place the one degree base map transparency over the "best" (considering coloring, cloud cover and date flown) LANDSAT scene available for the area to be stratified and tape to the scene.

(3) Place the 1:250,000 soils map transparency over the base map overlay; align and tape to the base map overlay so that it does not slip and yet can be picked up to observe the LANDSAT scene more closely if required. Review with assistance other topographic maps if available. In some instances the magnification process may obscure a portion of the LANDSAT scene. If this is too drastic, new overlays should be drafted.

(4) Place clear overlay map transparency and tape to the base map and soils map transparencies.

(5) Review Operational Navigation Chart (ONC) and geographic research data (if available) and compare with imagery features to achieve good orientation and write the geographic characteristics of work unit on the work form. The land use map obtained from Dr. Westin may be used for supplementary information.

(6) As the first step in stratification, delineate definite crop land areas (field patterns visible) with a blue grease pencil on the clear overlay. Concentrate on broad pattern areas and color differences when delineating and also pick out small areas which are in striking contrast to the broad area. For example, a small densely cultivated valley inside a heavily wooded region. Code these cropland areas by percentage cultivated: less than 5, 5 to 20, 20 to 40, 40 to 60, 60 to 80, 80 to 95, and 95 plus (see coding sheet). All areas having less than 5 percent cultivated land are coded A. Bodies of water will also be drawn out and coded as "0" (not sampled). Minimum block size for land is 9 square miles (3/4 $inch^2$) and water is 4 square miles (1/2 $inch^2$). The latitude and longitude lines forming the boundary for the one degree block are the work unit boundaries and cultivated strata blocks should not cross these lines.

(7) A unique stratification problem involves classifying and coding land within the one degree work unit which is in a country other than Brazil (i.e., Argentina). Since the entire one degree work unit must be stratified and coded, the following guidelines apply:

A. "Foreign" land within 4 miles (approximately 1 inch) of the border must be stratified and coded as though it were part of Brazil.

B. "Foreign" land beyond the 4 mile limit can be lumped together and coded "0" (not sampled).

2. First Analyst Procedure

(1) Working with a purple grease pencil, review the basic cultivation stratification done by the first carto. Make working notes in the Carto note section of the worksheet and discuss any changes or modifications with the carto responsible for the cultivation stratification.

(2) Use ONC's or other topographic maps which may show land use such as swamps, brush, and trees, etc., to further subdivide the areas coded "A" into three general groupings: (1) <u>no potential</u> for agriculture (arid, steep, or mountainous terrain), (2) <u>little if any potential</u> for agriculture (woods, brush and salt marsh), and (3) <u>some potential</u> for agriculture (grassland and other).

(3) The blocks of land stratified by percentage of cultivated land in Carto step one are to be coded for the percent of the cultivated land that is potential corn and soybeans. Review the crop calendar and any

crop acreage data or supplemental crop maps available for the region being stratified. The crop calendar will show usual planting and harvesting dates for any competing crops (crops that may be planted on the same land at the same time). Arrange all available imagery by months--remember that crops are expected to change in specific fields between years as well as within years, if double or continuous cropping is normal practice. In Brazil the first month of the growing season will be August. There will be a varying number of scenes with potential dates across several years. Thus, on any one scene, the crop calendar will need to be interpreted for that specific date (interpretation consists of defining areas with potential corn and/or soybeans). Concurrently use all statistical data that may be available in determining occurrence and density of corn and soybeans. Adjoining scenes with varying amounts of overlap may provide additional multitemporal information and should be used whenever possible. Soil type groupings may provide additional information on delineating or extending areas of specific crops. Using the crop calendar, look for crops that are planted earliest in the year ahead of other crops. These should show up first on a sequence of images for the same footprint and you may be able to determine that crop's signature for the season. The same approach may be taken for harvest times when early crops are maturing and are being harvested. Once particular fields are identified as specific crops this information may be used to identify similar and nonsimilar crops.

(4) Classify the blocks by four groups based on percentage of corn and soybeans. The breakdowns are 40 percent or more of the area devoted to corn and soybeans, 15 to 40 percent, 5-15 and less than 5 percent (See appendix 3, Coding Sheet). If the density or potential density of corn and soybeans varies substantially within the block, additional blocks should be formed. The density of corn and soybeans is independent of the percentage agriculture, namely, of the cultivated land in the block 40 percent is potential corn and soybeans, and so forth.

(5) A short narrative should be developed by the analyst describing the cropping practice and calling attention to specific areas that will assist the cartographic personnel in subsequent steps. In addition to the narrative, the analyst may also write information about interpretation on the overlay materials. These comments may entail such items as saying the light pink colored areas are grassland, while the red field on Scene #___ and the brown fields on Scene #___ are thought to be corn and

soybeans and the white fields on Scene #___ are small grains. For notes referring to specific blocks of land, a temporary work number identifying the block should be entered on the overlay and used to reference the appropriate note or comment (see section II-3).

3. Second Carto Procedure

   (1) Place a new, clear overlay over existing transparencies. Align and label this second overlay.

   (2) Using green lumocolor, draw in stratum blocks using straight lines where possible, to facilitate digitizing. Code each block with the appropriate codes for percentage land cultivated and percentage of corn and soybeans.

   (3) The next step is to classify the blocks as to predominant field sizes. The three major breakdowns being fields less than 100 acres, fields 100-300 acres and fields greater than 300 acres. A template should be used for questionable or marginal blocks. Use appropriate codes from the AgRISTARS coding sheet.

   (4) Code the blocks of land according to major soils types and enter code on the overlay. We will code only the major soil types found in the block (see the coding sheet). Subdivide the block to improve the accuracy of the soil classification if possible to do so without violating the minimum block size (9 sq. miles, 3/4 inch^2). Each time a block is subdivided, all code data should be entered into each new block.

4. Second Analyst Review

   (1) Review the final carto stratification and make any necessary corrections in purple lumocolor. Base the review on all available materials, notes, and ancillary data.

   (2) Review for consistency in stratification and make any necessary corrections to ensure that consistency.

5. Final Frame Preparation

   (1) Place the final frame material over the base map and the reviewed and checked second carto overlay (remove all other overlays) and draft the final frame. Bring forward the appropriate codes for percentage cultivated, percentage corn/soybeans, field size and soil type. Assign a unique three digit code to each strata block by numbering the blocks starting with

001 in the northeast corner of the one degree work unit in a serpentine fashion until all blocks are numbered. The code data to be entered in each strata block is as follows:

XXX	- unique strata block number
X	- percent cultivated
X	- crop specific code
X	- field size
XX	- soil type
X	- check digit

(2) Compute the check digit and enter. The check digit is the unit portion of the sum of the individual digits. For example if the block had been coded 24721212 the check digit would be computed as

$$(2+4+7+2+1+2+1+2) = (21) = 1$$

Thus the complete code would be 247212121

6. Supervisor Review

Review for completeness and neatness.

7. Statistician Review

Formal acceptance of work unit.

APPENDIX 2

AgRISTARS

STRATIFICATION PROCEDURE AND ANALYSIS RECORD

Country ____________________

1st Political Subdivision(s) ____

2nd Political Subdivision(s) ____

____

Crop Variables (1) ______________ (2) ______________ (3) ____________

Work Unit Number ____________________ Coordinates ____________________

Footprint ________________

Date:

Footprint ________________

Date

Footprint ________________

Date:

Footprint ________________

Date:

Scene(s) used as base for stratification:

FIRST STRATA OVERLAY

Carto Name ______________________________

Date Started ________________ Date Finished ______________

Carto Notes: (Orientation, Alignment, Scene Quality)

Geographic Characteristics of Work Unit:

Climate of Work Unit:

Crop Calendar: Attached

Crop Data Sheet: Attached

Related Maps, Other Materials ______________________________

____

Carto Notes: (Related to land use analysis and agriculture)

Analyst Name: ______________________________

Date Started ______________ Date Completed ______________

Analyst Notes:

SECOND STRATA OVERLAY

Carto Name ______________________________

Date Started ______________ Date Finished ______________

Carto Notes:

STATISTICIAN REVIEW

Name ______________________________

Date Started ______________ Date Finished ______________

Notes:

FINAL FRAME

Carto Name ______________________________

Date Started ______________ Date Finished ______________

SUPERVISOR REVIEW

Name ______________________________

Date Started ______________ Date Finished ______________

APPENDIX 3

BRAZIL AgRISTARS CODING SHEET

1. Count Unit ID - enter sequential 001 to nnn digit(s) starting in northeast corner in east to west, west to east serpentine manner.

2. Percent Cultivated (Final Stratification)

Carto Code	Analyst Code	Description
0	0	not sampled (water, other countries).
	1	Less than 5% cultivated, no potential for agriculture, usually arid, steep mountainous terrain.
A	2	Less than 5% cultivated, little if any potential for agriculture, usually woods, brush and marsh.
	3	Less than 5% cultivated, has potential for agriculture, usually grassland.
4	4	5 to 20% cultivated.
5	5	20 to 40% cultivated.
6	6	40 to 60% cultivated.
7	7	60 to 80% cultivated.
8	8	80 to 95% cultivated.
9	9	95% or more cultivated.

3. Percent Corn and/or Soybeans

Code	Description
0	not sampled (water, other countries).
1	Less than 5% corn/soybeans.
2	5 to 15% corn/soybeans.
3	15 to 40% corn/soybeans.
4	40% or more corn/soybeans.

4. Field Size

Code	Description
0	not sampled (water, other countries).

1	Fields less than 100 acres.
2	Fields 100 to 300 acres.
3	Fields greater than 300 acres.

5. Soils Type

Code	Description
00	not sampled (water, other countries).
01	AH - Hydromorphic Alluvial Soils.
02	A_HA - Hydromorphic Alluvial Soils with well drained acid Alluvial Soils.
03	A_HRB T/A -(No narrative description provided)
04	CH R/L - Chernozem on loess-covered undulating to rolling plains.
05	CH_LT/A - Loamy Chernozem soils on alluvial plains.
06	G-HB - Gray Hydromorphic soils with half bog on nearly level plains.
07	LF_SR/Sd - Sandy Low - Humic Ferruginous Latosols from Sandstone on rolling terrain.
08	LF_S R/SL -(No narrative description provided)
09	LF_S R/T - Sandy Low-Humic Ferroginous Latosols from consolidated rocks on rolling terrain
10	LR H/C - Red Latosols from undifferentiated consolidated rocks on hilly terrain.
11	LR H/S - (LR H/TS) - Red Latosols from sedimentary rocks on hilly terrain.
12	LR M/ - Lithosolic Red Latosols on mountains of humid tropics.
13	LR R/C - Red Latosols from undifferentiated consolidated rocks on rolling uplands.

14	LR R/U - Red Latosols from unconsolidated materials on rolling terrain.
15	LTX R/B - Terra Roxa (Latosols) from basic rocks on rolling terrain.
16	M - Dominantly swampy land but locally includes islands of better-drained land.
17	PCH_{LA} SK - Planosolic Chernozem soils from lacustrine and materials with Solonchak.
18	PR R/S - Prairie from sedimentary rocks on undulating to rolling plains.
19	PR R/U - Prairie from unconsolidated materials on rolling plains.
20	RBu - Reddish brown from unconsolidated materials in level to rolling plains.
21	Reddish brown from unconsolidated materials on level to rolling plains, with Solonchak.
22	RG_{FU} - Regur soils fron fine-textured unconsolidated materials on level to gently sloping plains.
23	RL H/B - Reddish brown lateritic from basic igneous rocks on hilly terrain.
24	RL H/S - Reddish brown lateritic from sedimentary rocks on hilly terrain.
25	RL R/B - Reddish brown lateritic from basic igneous rocks on rolling plains
26	RYP H/C - Red and yellow Podzolic soils from consolidated rocks on hilly terrain.
27	RYP R/SdM - Red and yellow Podzolic from sandstone on undulating to rolling terrain, marshes and ponds.
28	RYP T/A M - Red and yellow Podzolic on terraces with marshes and ponds.

29	RZ H/B - Rubrozems and associated soils from basic rock on hilly terrain.
30	RZ H/g - (No narrative description provided)
31	RZ R/B - Rubrozems from basic rocks on rolling terrain.
32	RZ R/S - Rubrozems from sedimentary rocks or rolling topography.
33	RZ RH/C - Rubrozems from consolidated rocks on rolling and hilly terrain.
34	SB - Sandy Beaches
35	SM-A_{HS} -Salt water marsh with Hydromorphic Saline Alluvial soils.
36	W - Wiesenboden (flat to nearly level plains).
37	PPR Y/S - (No narrative description provided)
38	PR Y/S - (No narrative description provided)
39	RG R/B - (No narrative description provided)
40	AH-AA - (No narrative description provided)
41	RYP R/TS - (No narrative description provided)
42	RYP R/SD-M -(No narrative description provided)

APPENDIX 4

MUNICIPIO AND LAND USE DATA SUMMARIZATION

Municipio (a county sized political subdivision) boundaries were digitized to provide a file which could be used to assign municipio codes to the sampling unit file and to be used in the computation of crop densities for these regions.

Estimates of area planted to corn, soybeans, rice, kidney beans, cotton, and sugar. There was also a category for other unspecified crops such as wheat, coffee, and so forth. However, total land was not available at this level.

Total land area was estimated by digitizing maps supplied by NASA. These maps showed municipio boundaries and latitude and longitude lines. The original 1:1,000,000 maps were enlarged and reproduced on clear acetate at 1:250,000. In order to form our 1° square work units, mosiacs were made and the work units traced on paper copy. The tracing of work units was necessary because many 1° squares were on parts of two or more maps. Further, the dots used to indicate places to digitize were not clearly visible when placed on the acetate because of the thick black boundary lines which were caused by enlarging the small scale maps.

Estimating total area by municipio posed several problems which had to be resolved. The maps showing the municipio boundaries had to be drafted by NASA and were received late in the construction phase. These maps contained a few municipios that had been incorrectly coded, some boundaries had to be estimated where maps overlapped, and eight municipios were omitted entirely.

Errors in digitization also had to be resolved. The digitizer boards and microprocessor were recently purchased and software had to be developed for the system. Time was needed for training in the use of the equipment. The disk storage capacity of the minicomputer limited the amount of data which could be processed at any one time.

Municipio areas were digitized by work units. This procedure created special problems because many municipios had parts in more than one work unit. These parts had to be aggregated to obtain the total area for the municipio. Thus, a method was devised for determining when all parts of a municipio had been digitized. Only then could the areas be summed to obtain the total required.

The first step in aggregating municipio areas was to obtain computer listings of the digitized areas (in square inches) by work unit. Each work unit listing was manually matched with a map which showed municipio boundaries. If a municipio fell in more than one work unit, each part was numbered and coded so that the part could be associated with the particular work unit which contained that portion. The total number of parts and the specific part numbers for each municipio was recorded.

These values along with an identification (ID) variable and the digitized area were used as input for the update program. This update program created variables which identified each work unit containing a portion of a particular municipio and calculated the municipio area when all parts of a municipio had been processed. The master file was then updated. After each run, printouts showing the crop area estimates and the crop percentages of the total area were obtained. The printouts were checked against the map to verify that municipios were correctly shown as completed. Whenever discrepancies occurred, such as improper coding of the parts of the municipio falling in different work units, omission of municipio boundaries, duplication of municipio numbers, or errors in the identification variable, the problem was resolved.

All of the preceding data manipulations were required to obtain ancillary data in a form that was useful for developing the area frame. Since strata were defined using the percentage of agriculture and percentage of corn and soybeans, the ancillary data also needed to be in percentages. The analyst used the percentages to check signature interpretations for consistency within and among strata. The variable, percentage corn and soybeans, was computed by summing areas planted to those two crops and dividing by the sums of areas planted to all crops. After arriving at the estimated total area we computed the value for percentage of cultivation by dividing the sum of the crop areas by the total land area.

APPENDIX 5

SOFTWARE DOCUMENTATION	The programs necessary for the development of the sampling frame were written by Bob Hale and Ron Fecso of the SFDS, SRS, USDA in early 1980. These programs edit and manipulate the raw data in the digitized files and then convert the edited data into the format required for automated sampling of the frame. 7/

PROGRAM DESCRIPTIONS

Polygon Control Number Verification	The polygon control number identifies each land use block and is made up of a unique block number, the vector of information, and a check digit. The check is defined as the units value of the sum of the digits making up the block number and the vector of information. The digits of the polygon control number are summed and the units digit is compared to the check digit. If the values do not match a warning message is issued. The coding for each flagged polygon must be rechecked and corrected.
Area Check	The digitized area of the polygons is aggregated and compared to the total area of the work unit. If the areas differ by more than .2 percent an error message is issued. The points in the file must be verified and any corrections made by the person who digitized the work unit.
Calibration Check	The four corner points in the calibration file are compared to the four corner points in the point file. If either the X or Y coordinates of the matching corner points differ by more than .025 inch, the calibration file must be redone.
Point File Main Edit	Matching points are checked and the number of matching points is verified. Matching points are defined to be points coincident to contiguous polygons. For example, if matching points make up the coincident boundary of two contiguous polygons each of the matching points would have to be recorded twice in the digitizing process in order to define the boundaries of both polygons. If the X and Y coordinates of matching points differ by more than .025 inch they are considered to be nonmatches. All points not having the required number of matches are output in an error file. Corrections are made by the person who digitized the work unit.
Point Averaging	The X and Y coordinates of matching points are averaged and assigned the average value of all matching points in the point file.
Point File Conversion	The X and Y coordinates of the points in the point file are converted to latitude and longitude. The conversion uses a linear relationship determined by the four points in the calibration file.

7/ Copies of these programs are available upon request.

<u>Work Unit Gridding</u>

The center points of the work units are used to form a grid of sample units covering the area of interest. The auxillary information assigned to each sample unit, is determined by the location of the center points. Each sample unit is assigned the information corresponding to the polygon in which the center point falls.

APPENDIX 6

Geometric Arguments

Lines of longitude correspond to the intersection of the surface of the earth with planes which contain the polar axis. Lines of latitude are the intersection of planes with the surface of the earth such that the polar axis is perpendicular to the plane. Treating the earth as an ellipsoid of rotation (4) (Clark 1866 Ellipsoid also known as 1927 North American Datum), the following parameters are required:

1. Equatorial radius	$a = 6{,}378{,}206$ meters
2. Polar radius	$b = 6{,}356{,}584$ meters
3. First eccentricity	$e_a = (a^2-b^2)^{.5}/a$
4. Second eccentricity	$e_b = (a^2-b^2)^{.5}/b$

Then, the distance on the surface of the earth from the equator to a specified latitude, t, is (in series expansion)

$$S(t) = a_o t + a_1 \sin(2t) + a_2 \sin(4t) + \ldots$$

where

$$a_o = a(1-e^2)\ (1+(3/4)e^2 + (45/64)e^4 + (175/256)e^6 + (11025/16384)e^8 + \ldots)$$

$$a_1 = -(1/2)a(1-e^2)((3/4)e^2 + (15/16)e^4 + (525/512)e^6 + (2205/2048)e^8 + \ldots)$$

$$a_2 = (1/4)a(1-e^2)((15/64)e^4 + (105/256)e^6 + (2205/4090)e^8 + \ldots)$$

For the creation of AgRISTARS sampling units, the inverse of S(t) is necessary. Thus, using the following formula we can compute the latitude (in radians) of a specified distance, H, from point t_i along a line of longitude.

$$T(H,t_i) = t_i + (S-S_o)\ q_1 + (S-S_o)^2\ q_2 + \ldots$$

with $H = (S-S_o) = (S-S(t_i))$,

$$q_1 = 1/(a_o + 2a_1 \cos(2t_i) + 4a_2 \cos(4t_i)),$$

and

$$q_2 = (1/2)\ q_1^3\ (4a_1 \sin(2t_i) + 16a_2 \sin(4t_i)).$$

The distance along a line of latitude, t=T, between two lines of longitude is determined by

$$D((T,g_2), (T,g_1)) = (r \cos T)\ (g_2 - g_1)$$

where g_2 and g_1 are the lines of longitude and

$$r=(((\cos^2 o)/a^2) + ((\sin^2 o)/b^2))^{-2}.$$

Thus, the angle of longitude (in radians) subtended by an arc of length, L, along the line of latitude t_i is

$$G(L,t_i) = \frac{L}{(r\cos(t_i))}$$

Each of the distance functions has an inverse mapping in which, given a point, P_c, and a desired distance along a given latitude or longitude, the latitude and longitude of the point can be computed. To grid a spheriod using these distance measures and inverses, it is necessary to define starting points (6). The equator and the 0° line of longitude was used for convenience in computer applications.

Creating the Sampling Frame

In this section the approach for creating a "grid" system of sampling units is outlined. The grid system of sample units can be established by first defining planes of latitude which slice the surface of the sphere into strips which have latitude defined by t_i, such that,

$$S(t_{i+1}) - S(t_i) = H \quad i = 0, 1, 2, \ldots, I$$

where H = height of desired sample unit,

$t_o = 0$ (The equator),

t_{i+1} is greater than t_i, and

t_I is less than (2/9) radians (approximately 80°).

Note that the slicing algorithm for general application is done from the equator toward the North Pole. The appropriate reflections are used to define the slices in the Southern Hemisphere.

After the strips which will contain the sample units are defined, all that is required is to slice the strip into the appropriate length (L) sample units. To do this the algorithm starts the slicing by placing the westernmost edge of the first sample unit approximately along the 0° line of longitude. The center line of latitude is computed for the strip and increments of length equal to the sample unit length are made along the center line. Thus, the center points of this latitude/longitude slicing process define the sample units.

To understand the gridding algorithm, note that there are (I + 1) unique values for latitude, t_i, which slice the quadrant. For each of the j = 1, 2, . . ., I "center latitudes" of the slices there are N_j values of longitude in the quadrant which "chop" the "slice" into sampling units. Notice that N_j is nonincreasing as the center latitudes, Y_j, increase.

The Y_j are found incrementally by first defining $t_0 = 0$ and then incrementing along $g = 0$ to find $t_1, t_2, \ldots, t_I$. The latitude of the center points of the sample units, Y_j, is found by:

$$Y_j = (t_j + t_{j-1})/2, \; j = 1, 2, \ldots, I$$

where

$$t_j = T(H, t_{j-1}).$$

The N_j sample units along latitude Y_j are computed iteratively as follows. First let

$$g_0 = 0.$$

Then the line of latitude is chopped into lengths the size of the sample units by creating

$$g_k = g_{k-1} + G(L, Y_j), \quad k = 1, 2, \ldots, N_j.$$

Thus the center point longitude of the sample unit, X_k, are computed as follows:

$$X_k = (g_k + g_{k-1})/2, \quad k = 1, 2, \ldots N_j.$$

As a result, the iteratively created collection of ordered pairs

$$(X_k, Y_j)$$

define the center point expressed in terms of latitude and longitude for each sample unit in the quadrant. It should be noted that the actual program creates a file of sample units which consists of the intersection of these center points and the area for which the frame is being created. That is, only the sample units in the target area are created.

☆U.S. GOVERNMENT PRINTING OFFICE: 1982-360-931:SRS-688