Comparison of
Systematic Unaligned Sampling Designs
For Estimating Land Uses

by

SITI ZABEDAH AMRI

A Research Paper

submitted to

THE DEPARTMENT OF GEOGRAPHY

in partial fulfillment of the
requirements for the
degree of

MASTER OF SCIENCE

April, 1989

Directed by
Dr. JAMES R. PEASE

Contents

I. Introduction

Increasing demands for more intensive land-use have led to greater conflicts of interests on the issue of how to manage land resources. This trend is expected to continue due to mounting population pressure and increasing demand for higher standards of living. As a result, more competition for land as a resource can be expected among land users (Zeimetz et al, 1976).

In an effort to address these concerns, information on land use and land-use dynamics is essential, particularly in areas where the population is growing rapidly. Generally, there are two ways to obtain this information; a complete census or sampling a small fraction of the area of interest. The choice between complete enumeration and sampling depends on the types of information required and the cost and time involved. If accurate information is desired for many categories of land use, the size of the sample needed is sometimes so large that a complete census offers the best solution. On the other hand, if there is more scope and flexibility regarding the information to be obtained, sampling may well be the practical method to utilize.

Realization that the greatest pressure for land conversion occurs in the areas surrounding large urban places and that there is the need to protect prime agricultural lands led to completion of two pilot studies. Both studies were conducted in Oregon to test methodology, procedures, and data sources for a national level examination of resource use and land use change in fast-growth counties (Vesterby, 1987). This paper is a further attempt to improve the technique of obtaining reliable and efficient land information through sampling methods.

I.1. Problem Statement

The direction and scope of this research paper follows closely an earlier sampling study conducted by Behm and Pease (1985). In this pilot project, the authors employed a two-stage random sampling method in determining land use estimates. Anderson's land use classification system was used for data categorization (Table 1). The study shows random sampling does not give adequate coverage of the population of land-use types. Primary sampling units were concentrated in just a few areas of the map and in other areas no primary sampling units were chosen. Representative coverage of the area was important because all parts were of equal interest. In addition, the random sampling scheme was not sensitive in detecting sporadic or concentrated types of land use distribution.

I.2. Study Objectives

This paper focuses on sampling the same study area using seven different variations of the stratified systematic unaligned sampling method. This is a hybrid design incorporating elements of the simple random sample and systematic methods. It is probably more accurate in determining land use estimates because it is more geared to provide representative coverage of the study area. The two main objectives of this study are:

1. To determine the most effective sample point density and allocation which yield improved estimates of land use proportions.
2. To evaluate whether stratification based on a Land Use District Map will provide better estimates.

The population parameters in this study are the proportions of the different

Table 1. Anderson's Data Classification
(adapted from the pilot study Behm and Pease,1983)

1. Urban or Built-up Land		
	P1	Residential
	P2	Commercial and Services, Industrial, Industrial Complexes, Mixed Urban
	P3	Transportation, Communications, and Utilities
	P4	Other Built-up Land
2. Agricultural Land		
	P5	Cropland and Pasture
	P6	Orchards, Groves, Vineyards, Nurseries and Ornamental Horticultural areas
	P7	Other Agricultural Land and Confined Feeding Operations
3. Others		
	**	Rangeland
	P8	Forest Land
	P9	Water
	P10	Wetland
	P11	Barrenland

land use classes. The sample data estimate the means of these population parameters. The primary use of these data is for paired comparison of land use between two dates to detect changes. Presumably, improved accuracy of estimates of land use distributions will, in turn, increase the accuracy of estimating land use change.

II. Literature Review

Urban use of land will continue to expand in the future. Formulation and advocacy of particular policies concerning land will require up-to-date and reliable land use statistics. Improved data about land resources are necessary to avoid mismanagement and alert the resource planner to serious losses or stress (Heimlich and Anderson, 1987).

Early efforts to aid development of effective land use policies put emphasis on mapping existing land resources (Marschner, 1940; Wood, 1955; USDA, 1958). Research activities were also concentrated on finding sampling methods which provided rapid, cost-effective, and reliable information (Berry, 1962; Holmes, 1967; Stobbs, 1968). Currently, data obtained from aerial survey are used to inventory/map, plan, manage, and monitor various types of land use. Research findings from various studies seem to indicate stratified systematic unaligned sampling is the most appropriate design to use for land use change study (Smart and Grainger, 1974; Zeimetz, 1978; Fitzpatrick-Lins, 1981; Rosenfield, 1982). This sampling scheme is area weighted. Nevertheless, it consistently provides better results than other sample designs.

At present, there is an active interest in designing an effective program for updating land use information at national and regional levels (Vesterby, 1988; Zeimetz et al, 1976; Frazier and Shovic, 1980). Land use change statistics are now used in the broader context of understanding land use dynamics. They are used to verify statistical models correlating changes with microeconomic, macroeconomic, and social variables in an attempt to explain and predict the occurrence of urbanization (Heimlich and Anderson, 1987).

III. Description of Test Area

Washington County, Oregon, has an area of approximately 731 square miles. It was selected as the target area because of tremendous population growth between 1970-1980. In the span of 10 years, the county gained approximately 90,000 people (U.S. Bureau of Census Data), which qualified it as a fast-growth county (growth of 25000 people and 25% population change). The test area in the southeastern part of Washington county has an area of about 163 square miles (Figure 1). Delineation of the test area boundary took into account practical constraints, excluding Federal lands, Indian reservations, and large bodies of water. A further constraint was imposed by the availability of control data. A complete census of the test area is only available for the years 1981-1982.

Through the 1950s the economy of this county was based on agricultural and lumber products with the population being farm oriented. After the war several industries located in the county and roads were improved, providing better access to the more industrialized Portland area. This development gradually changed the character of eastern Washington county from an agricultural area to a suburban, bedroom area for Portland. Furuseth (1978) stated that over 50% of this county was classified as farmland in the 1950s. However, in the past decades agricultural land use has declined. This decrease may be attributed to two factors; consolidation of smaller farms into larger, more efficient enterprises and conversion to urban uses.

Future impacts on this area include possible annexation of unincorporated territory by nearby cities and expansion of Interstate 5 through the area (The Oregonian, 1988). If these plans are implemented the effects on land use will result in a shift to higher-intensity uses and a decrease in prime agricultural land.

IV. Considerations Prior to Sampling

Determining an effective sampling design to accomplish the goals for this study of land use required careful consideration of the design options and technological and operational alternatives. An efficient design would integrate these three key elements into an overall plan that accomplished the survey objectives.

IV.1 Design Options

In an effort to obtain more desirable sampling distributions, estimates of their characteristics, and particularly a reduction of sample variance, the following design options were manipulated: sampling frames, sample size, form of estimator, and method of sampling. In addition, the final selection for the best option(s) depends on the evaluation of the cost of the different alternatives. The following discussion is focused on the analysis of prior considerations involved in planning the sample design.

IV.1.1. Defining Sampling Objectives

First, the need to identify the survey objectives is important because it is one of the principal criteria by which the utility of a sampling design is judged. In this study, there are two primary objectives; (a) to estimate the composition of the study area in 1982 using the land use classes in Table 1; (b) to examine the effects of varying sample point density, allocation, and stratification on the sample estimates.

IV.1.2. Sampling Frames

The general sampling design utilized in this study can be described as an areal-point sampling methodology. It is a two-level sampling frame stratification of the spatial area of interest. The first frame organizes the population of land use classes into equal-sized, contiguous squares referred to here as the primary sampling units (PSUs). The second frame is the set of intersections of grid lines within each PSU. Each intersection is referred to as a secondary sampling unit (SSU). An observational unit is an area mask or "decision cell" of specified size around each of the sampled SSUs.

IV.1.2.1. Grid Size

One of the basic problems of data acquisition using grid frameworks for a regional study is the selection of minimum cell size to overlay on the study area. Coarser resolution cells result in "loss" of information while finer resolution cells usually lead to unnecessary redundancy of information that is more costly to collect and analyze. Nonetheless, there are criteria for an ideal or optimal cell size to guide the final selection. These include a grid size that reduces error estimates to an acceptable level, minimizes the locational error of data, and does not require too many data cells.

An extensive study of the effects of varying grid resolutions based on measures of the percent deviation of an estimated value from the actual value was conducted with grid cells varying from 2x2 km to 16x16 km in size (Bircham, 1979). Assessment for optimal size was made on the basis of land area to be sampled which is indicative of potential costs. A 2x2 km grid turned out to be the most cost effective size.

In this study, a one square mile grid-size was chosen to maintain consistency with the pilot study. In addition, this cell size is nearly equivalent to the 2x2 km cell that Bircham (1979) determined to be most cost-effective. A total 163 cells were identified from the area's map (1:250,000) to constitute the set of PSUs.

IV.1.2.2. Dot Density Within Each Cell

Another sampling design consideration was determining the density of secondary sampling units within each grid cell. There are varying effects of different alignments and spacings of SSUs in capturing the land use patterns within a square mile. High point density will ultimately increase cost and redundancy of information obtained. Low density may result in less reliable land use estimates depending on type of data and complexity of the cell.

In the pilot study, each cell was further divided by grid lines with equal spacing of 1/10th of a mile. This resulted in 121 grid intersections, including the points along boundary of the grid cell. To maintain consistency with the pilot study the same grid framework was used in this study.

IV.1.3. Estimation of Sample Size

The next step in the planning of the sample survey was choosing the appropriate sample size. There were sample sizes in two sampling frames to be considered: sample size from set of PSUs (areal units) and sample size from set of SSUs (point units). The choice of these two sample sizes is a function of the precision required.

In the pilot study, the authors prescribed the degree of precision for each land use class derived from a prior five-percent survey of the study area. Each observation unit was treated as a random experiment of Bernoulli trials. The probability of success, say p, for each land use class remains the same from trial to trial. That is, the outcome can be classified as a success or a failure. Mean percent of each land use type is estimated, with a separate confidence statement for each land use class. This procedure limits inference between land use classes. Based on the assumption that land use change average percent estimates are independent events, the binomial approach is used to determine sample size.

The calculations of sample size led to a series of conflicting values of n, one for each land use type. The authors reconciled these values by selecting the largest n from the most prevalent land use category, cropland and pastures. This sample size was further revised if $n(i) > .1N$. The final selection of 33 PSUs was chosen as an appropriate representation of the population. The sample size for the SSUs was set at 20 points per PSU. This sample density followed Berry's (1962) methodology of aerial-point sampling.

Another approach to estimating sample size is based on the multinomial sample estimation. In this case, however, the model provides estimate of proportion total for each land use class. This approach is more precise in estimating land use distribution. It does not limit inference between the land use classes. Paired differences of observed PSUs or absolute difference of the sample estimates provide land use change statistics.

In the current project, neither approach was used. The stratified systematic unaligned methods examined in this research paper preclude the approximation of sample size for the primary sampling units. The spatial choice of PSUs is confined by the grid structure. Generation of selected PSUs to be sampled is area weighted and predetermined by the choice of grid overlay. The selection process is discussed below.

The choice of the secondary sampling units to be sampled constitutes the major focus of this research. Theoretically, the optimal size can be statistically determined. The type of data required for this study did not mandate intensive sampling at this stage; nonetheless, the effect of varying the density of SSUs on precision of estimates was of major interest. Seven variations of stratified systematically unaligned selection of SSUs together with allocation of PSUs were examined.

IV.1.4. Sampling Methods

Once the sampling frames have been defined, structured, and the appropriate sample sizes estimated; then the question of sample selection can be addressed. The selection process of PSUs and SSUs make up an important component of the sampling designs. The guidelines for choosing the best design are the desired coverage of the study area, the minimum number of PSUs and SSUs to be sampled, and bias-free selection.

IV.1.4.1. Two-stage Random

This design was used in the 1985 pilot study. In order to ensure adequate representation due to the irregular boundary of the study area, grid cells which are 50% or more in the delineated polygon were included in the selection. Thirty-three PSUs were selected randomly from the area frame of 163 possible grid cells. Twenty SSUs were also selected randomly from the grid intersections sampling frame within each PSU. A total of 660 points were interpreted for each date. Matched comparison was used to improve precision in detecting land use changes.

The forms of estimator to determine mean statistics of land use distribution for a specific land use class using the two-stage random are as follows:

let $p(i)$ = 1982 percent occurrence of a specific land use type in the ith PSU

$n = 33$

land use distribution statistics:

$$\text{Mean}(p) = \Sigma\, p(i)/n$$

$$\text{var}(p) = \Sigma\, [p(i) - \text{mean}(p)]^2/n-1$$

The pilot study revealed two disadvantages of this method. First, there was inadequate coverage of the study area due to clustering of PSUs. Secondly, the model is not adequate to approximate the population parameters. The method is based on the assumption that populations are independent and normally distributed. Sample error estimates of the least prevalent classes indicated that populations of these classes were highly skewed. Improved estimates of these classes might be achieved by sampling more in these land use types.

For this sampling design, there is an improvement of land use change estimates using an estimator based not on paired difference of PSUs but on differences of means. Change estimates based on sample means may be improved by an improvement of sample data for the land use distributions.

IV.1.4.2. Two Stage Systematic

Tests of several representational schemes of this design showed poor coverage of the study area. The organization of population elements into a grid structure results in linear patterns of PSU distribution. Consequently, the sample estimates from this scheme are less accurate due to lack of coverage. The distinctive linear patterns can be reduced by changing the grid resolution to smaller than one square mile. This would, however, result in another deviation of the sample design. Because of incompatibility of this method with the PSU sampling frame, it was excluded from further analysis.

IV.1.4.3. Systematic Unaligned

This sampling design is the major focus of the paper. Seven variations of PSU allocation and SSU density were examined to determine which combinations provide the better estimates.

IV.2. Technological Options

There are three possible sources for obtaining land use information: use of existing records (e.g., USGS land use/land cover maps), field surveys, and remote sensor images (Loveland, 1976).

Among these available options, aerial photography constitutes a valuable source for observable land use data. In aerial-point sampling, the technique of using this remote sensing tool to provide detail on the dynamics of land use change is usually relatively inexpensive. Aerial photography of the study area was obtained at the scale 1:58000. The normal 60% overlap provided stereoscopic coverage of the whole area. The use of color infrared images further enhanced detection capability.

IV.2.1. Spatial Resolution

Sample survey using aerial photos requires the appropriate scale of images to fit the type of data needed. In this regional study, the resolution is small, which results in high levels of generalization. Air photos at the optimal scale should adequately describe the traits of land use class variation in a single square mile. The choice of an appropriate scale of air photos will influence the accuracy and precision of sample data. Color infrared National High Altitude Program contact prints were used. The minimum interpretation unit (area mask) as well as the level of detail is also influenced by scale of the air photos.

IV.3. Operational Alternatives

There are several systematic methods of obtaining land use information from aerial photos. In this project, the data production operational strategy follows the procedure adopted by the previous pilot study. Consistency with the operational method in the pilot study allowed for direct comparison of sample estimates. The goal was to obtain variations in sample estimates that were the results solely of sample design selection. The details of data collection are as follows:

1. A 1:24,000 USGS map (taped together) is used as the reference base map for obtaining locational accuracy in sample selections
2. Construct an overlay grid to cover the entire study area with one grid cell per square mile - another grid overlay was constructed on drafting film at 1:24,000
3. Initial plotting of sample points - Insert square grid 1:24,000 under each USGS map quad on a light table and align with designated sample area

4. Transfer of designated sample points from base maps to aerial photos

....use zoom transferscope

....place map on transfer scope' stage, examine photo on table using 4X lens with clear acetate taped as flat as possible to the photo frame (Use cotton gloves to prevent smudge)

....transfer points with 000 tip ink pen

5. Point interpretation

....use area masks of 1.7 mm2 and 3.5 mm2 at 1:58,000 scale cut under magnification from opaque tape mounted on acetate

....use zoom stereoscope with .5X monocular lens to interpret point

....each point is assigned a land use category based on dominant use or higher intensity use when no single use was dominant

6. Record keeping: a page-sized grid was labeled for each square mile

- land use classification number code is entered into the circle with each point

V. <u>Systematic Unaligned Sampling Methods</u>

The selection procedure for choosing points that are systematically unaligned is as follows. The first point is chosen at random (Figure 2). Then, points for the first column of grid cells are determined by random selection of x-coordinates and y-coordinate of the starting point. Conversely, points for the first row are determined by random selection of y-coordinates and x-coordinate of the starting point. The remaining points are determined by y-coordinate and x-coordinate of the corresponding column and row (Smart and Grainger 1974; Cochran, 1977; Stoddard, 1982; Barber, 1988).

In this study, the fundamental basis for the selection of systematically unaligned points is adapted to the two sampling frames used. This modification allows selection of PSUs (areal squares) that are also systematically unaligned. A prerequisite for this procedure is the construction of grid with equal-size stratum. In Figure 2, for a 2x2 stratum, the first square in stratum (1,1) is determined randomly and denoted by A1. The second square,(A2), in stratum (1,2) is chosen based on the x-coordinate of A1 and a random y-coordinate. The third square, A3, in stratum (2,1) follows the same y-coordinate of A1 and random x-coordinate. The fourth cell in stratum (2,2) is determined by x-coordinate of A3 and y-coordinate of A2. A large scale representational scheme of the systematically unaligned areal squares is illustrated in Figure 3.

Application of this selection procedure to two-stage grid-structured sampling frames enabled formulation of a general probability model for the selection of systematic unaligned units. Expanding the selection concept further to include both point and area sampling units, probability models for the selection process can be derived. If n is the number of possible units of selection in a stratum then:

(a) for points selection

let A be the point selected

..if A does not have adjacent squares, $P(A) = 1/n$

..if A shares two adjacent squares, $P(A) = 2/n$

..if A has four adjacent squares, $P(A) = (4/n) - (2/n^2)$

(b) for areal unit selection

each areal unit has 1/n probability because there are no overlapping domains

Probability of selections computed for the sample units used in this study are as follows:

(a) sample point (SSU)

if point does not have adjacent squares,	p = 1/121	= .00826
if point shares two adjacent squares,	p = 2/121	= .01653
if point has four adjacent squares,	p = 482/121	= .03292

(b) areal square (PSU), p = 1/4 = .25000

These probabilities for the sample units show that this method of selection closely resembles simple random sampling, where sample units have equal probability of being chosen.

It is important to note that sampling frames used in this study impose two constraints on the flexibility of sample unit selection. A two-stage design of this method would set the possible grid intersections (SSUs) to be sampled to certain specific values determined by the grid structure. With exceptions to these numbers, to increase or decrease the points density would require changing the SSU sampling frame.

Maintaining consistency and simplicity of the model were the primary priorities. A compromise alternative is to generate SSUs for all the PSUs in the study area. Manipulating the density of SSUs needed only affect the number of times to repeat the selection process.

Another problem associated with the sampling frame structure is the selection of sample size for PSU's. The selection method adopted in this study presets the size of PSUs to be sampled, also the number does not agree with the statistically-determined estimation. Nevertheless, the sample size employed is greater than the estimated size based on a maximum allowable error of 5%. Thus, it is likely to produce better results because, in estimating proportions, it is the absolute size of the sample which improves the outcome. This problem does not apply when all the PSUs in the study area are included in the sample.

Different variations of the systematic unaligned selection process were employed for the selection of primary sampling units and secondary sampling units to determine an efficient design that fulfilled sampling objectives.

V.1. <u>Sampling All PSUs With One Point/PSU</u>

In this sampling scheme, all the possible primary sampling units are included in the sample. The observational units at the secondary stage is restricted to only one point for each PSU. The procedure for generating systematic unaligned sample points using this variation of the method, with specific application to the defined study area is as follows:

(a) a square grid overlay is placed on entire study area

(b) reduce matrix of cells from (20,20) to (13,17)
(to offset the effect of irregular boundaries of study area)

(c) generate potential points for each cell in the 13x17 matrix

(d) select sample points that fall into the defined boundaries

In this study, the resulting sample size is n=163 points. The estimator for calculating the variance is similar to estimate of proportions (Cochran, 1977:51); y=1 if sampling unit is in class of interest and y=0 if not.

N = population size = 163

n = sample size = 163

A = no. in class in population

a = no. in class in sample

sample proportion: $p = a/n$

$$s^2 = pq$$

$$\text{estimated variance: } v(p) = s^2/n$$

The advantage of this sampling plan lies in the reduction of the number of points to be interpreted (i.e., 163 instead of 660 sampled in the pilot study). Presumably, the same kind of accuracy can be achieved with complete coverage at the primary level of the study area. Nonetheless, this method ignores the heterogeneity of land use class within each primary sampling unit. Sample point replication within each PSU captures the heterogeneity of the primary sampling units influencing the overall estimate of population means. If, however, the primary sampling units are homogeneous then replication is not essential. Another weakness of this sampling scheme is the requirement of photo coverage for the entire study area. In a larger regional study this can be too costly to implement.

V.2. Sampling All PSUs With Four Points/PSU

Four replications of observational units for each primary sampling unit were tested, at the suggestion of Dr. Helen Berg, Oregon State University Survey Department. The rationale is to account for heterogeneity of PSUs. In this case, each observation point is given .25 weight and the number of points of observation is increased to 652. The selection procedure involved four replications of the same selection process mentioned above with different starting points. Due to the higher probability of points along edges being selected, samples with no overlapping points were chosen.

The appropriate estimators for this sample design follow the formulas for single-stage cluster sampling (Cochran, 1977:279):

N = total number of primary units = 163

M = total number of secondary units = 121

n = sample size of PSUs = 163

a(i) = no. of elements in ith secondary unit in class

m = no. of elements (SSUs) in each PSU = 4

$f(1) = n/N$; $f(2) = m/M$

$p(i) = a(i)/m$ proportion for ith PSU in class

$p = \Sigma p(i)/n$ proportion of class

sample variance between PSUs: $s(b)^2 = \Sigma\, [p(i) - p]^2/n-1$

sample variance between SSUs: $s(w)^2 = \dfrac{m}{n(m-1)} \Sigma\, [p(i)\, q(i)]$

Unbiased estimate of population variance is:

$$v(p) = \frac{1 - f(1)}{n} s(b)^2 + \frac{1/m - f(2)}{n} s(w)^2$$

This sampling scheme allocates more observation units throughout the entire study area. The effect of sample point density on estimates of land use classes can be evaluated by comparing estimates obtained from the two sample methods (1/PSU and 4/PSU). The requirement of total photo coverage is a weakness of this design.

V.3. Two-stage Selection With Equal Density

Another variation of the systematic unaligned sampling method was derived by integration of a two-stage selection process into the procedure. The selection units are aerial squares at the primary level and grid intersections (points) at the secondary level. This method samples a fraction of N areal squares (PSUs) and also subsamples m out of M possible SSUs (points) in each PSU. The procedure to select systematic unaligned PSUs, given a specific placement, is described below:

(a) a square grid overlay is placed on entire study area

(b) divide the new delineated boundary into equal-size strata of 2x2 each by aggregating sets of four contiguous cells

...grid placement provides matrix of (14,18)

(c) reduce matrix from (14,18) to (7,9)

(d) generate potential PSUs for the matrix (7,9)

...this will offset the effect of irregular boundaries of studied area

...maximum number of PSUs, n = 63

(e) select PSUs which fall within the original study area's boundary

...possible PSUs for sampling, n = 40

This arrangement of PSUs decreased the original grid resolution from one square mile to four square miles per stratum. One out of the four aggregated squares is chosen as PSU to be sampled. The benefit of this selection procedure is that the same sampling frame was employed as in the pilot study. Direct comparison with two-stage random method to determine effect of PSUs distribution pattern is possible because both samples are derived from the same defined population elements. However, the determination of PSU sample size is dependent on the grid resolution used per stratum. This constraint can be ameliorated by first deriving the appropriate sample size required to achieve a specified precision. Then, the statistically determined sample size will be the threshold value in guiding the decision about grid resolution. For example, the pilot study's desired precision for land use estimates was set at the 95% confidence level. The sample size estimated was 33 PSUs. In this study, the four-square mile grid resolution yielded a sample size of 40 PSUs which results in a greater coverage of the study area.

Based on previous studies, the SSUs sample size per grid cell with this variation of the method is set at 20 observational units (Zeimetz, 1976; Frazier and Shovic, 1980; Behm and Pease, 1985). This SSU sample size was considered an efficient sample point density given the PSU allocation in the areal-point sample survey. However, three different densities were also employed, which allows a good comparison of the effect of sample point density on estimates within PSU's. Limiting the pattern of distribution to systematically unaligned, one of the easiest ways to obtain units of observation is through replication of the selection process employed when generating sample units for all PSUs. Following this procedure, sets of 4, 10, 16, and 20 iterations of the systematic unaligned procedure were generated. The total number of SSUs to be sampled were 160, 400, 640, and 800 points respectively. The appropriate estimators for this two-stage sampling are as follow (Cochran, 1977:279):

let $a(i)$ = no. elements in ith SSU in class

n = sample size = 40

m = no. elements in ith SSU in class (4, 10, 16, and 20)

$p(i) = a(i)/m$ value obtained for ith PSU

$p = \Sigma p(i)/n$ overall sample mean %

variation between PSUs:

$$s(b)^2 = \frac{\Sigma\,[p(i) - p]^2}{n - 1} = \frac{\Sigma\, p(i)^2 - np^2}{n - 1}$$

variation between PSUs:

$$s(b)^2 = \frac{\Sigma\, [p(i) - p]^2}{n - 1} = \frac{\Sigma\, p(i)^2 - np^2}{n - 1}$$

variation within PSUs:

$$s(w)^2 = \frac{m}{n(m-1)}\ \Sigma\, [p(i)\ q(i)]$$

estimate of population variance:

$$v(p) = \frac{1 - f(1)}{n(n-1)}\ \Sigma\, [p(i) - p]^2 + \frac{f(1)\ (1-f(2))}{n^2\ (m-1)}\ \Sigma[p(i)q(i)]$$

V.4. Two-stage Selection With Unequal Density

In this design, the same 40 PSUs generated for the two-stage selection with equal density were used. Varying the density of observational units in each PSU is another variation of the systematic unaligned procedure. A Land Use District map produced by Washington County's Transportation Department provided the basis for stratification (Table 2). A complexity index map with numbers representing amount of different types of land uses within the PSU's was created. From this map, PSUs with just one land use type are designated as low density of sample points, PSUs with two land use types as medium, and PSUs with more than two land use classes as high. Strata are constructed such that the proportion in a defined class varies as much as possible from stratum to stratum (Cochran, 1977:107). This criterion assigned 17 PSUs to the low

Table 2. General Description of Land Use Districts

1. Exclusive Forest and Conservation (EFC)
   - Intended to provide forest uses and the continued use of lands for renewable forest resource production, retention of water resources, recreation and other related or compatible uses.

2. Exclusive Farm Use (EFU)
   - Intended to preserve and maintain commercial agricultural land for farm use consistent with existing and future needs for agricultural products, forests, and open spaces.

3. Agriculture and Forest - 20 (AF- 20)
   - Intended to preserve and maintain agricultural land in uses consistent with those included in an exclusive farm use district ... a 20 acre minimum lot size.

4. Agriculture and Forest - 10 (AF-10)
   - Retain the area's rural character and conserves natural resources while providing for rural residential uses. ... a 10 acre minimum lot size.

5. Agriculture and Forest - 5 (AF-5)
   - Retains the area's rural character and conserves natural resources .... a 5 acre minimum lot size.

6. Rural Residential - 5 (RR-5)
   - Recognizes rural areas which have been committed or developed for suburban residential uses with minimum farm and forest use .... minimum 5 acre lot size.

7. Rural Commercial (R-COM)
   - Provides for commercial activities which serve the convenience goods and service needs of rural residents.

8. Land Extensive Industrial (MA-E)
   - Provides land for farm and forest-related industrial uses.

9. Rural Industrial (R-IND)
   - Provides for industrial uses in the County.

(Source: Washington County Transportation Department) density class, 11 PSUs to the medium class, and 12 PSUs to the high category. Observational units are allocated as follows:

Low density - 4 points/PSU

Medium - 10 points/PSU

High - 16 points/PSU

The total number of observational units selected given these criteria is 370 points. This sample scheme attempts to allocate proportional sampling according to the complexity of each PSU. Decrease in observational units is expected not to affect the precision because the intensity of sampling is reduced only in the more homogeneous PSUs. The estimators used are adapted from the formula for stratified sampling (Cochran, 1977:90):

N(h) = total number of PSUs in stratum h

n(h) = number of PSUs in sample

p(hi) = value obtained for ith PSU

W(h) = N(h)/N stratum weight

f(h) = n(h)/N(h) sampling fraction in the stratum

p(h) = Σ [p(hi)]/n(h) sample mean

p(st)= Σ [W(h) p(h)] stratified sample mean

Assuming each PSU in stratum is approximately homogeneous, the variation within PSUs is ignored and the estimator for the sample variance is

$$s(h)^2 = \frac{1}{n(h)-1} \Sigma [p(hi) - p(h)]^2$$

The population variance estimator is:

$$v[p(st)] = \sum \frac{W(h)^2 \, s(h)^2}{n(h)} - \sum \frac{W(h) s(h)^2}{N}$$

VI. Data Evaluation

The following methods of analysis are employed to gain understanding about the effect of stratification on sample estimates and to determine which systematically unaligned sample design best describes the population. Sample design characteristics, including area weighting of each point, provide basic guidelines for comparison (Table 3). Density of secondary sampling units and allocation of primary sampling units are two major components of variation among the designs. In the detection of the least prevalent classes, two variables (area weighting and number of PSUs) most influence sample approximation of the actual value. Design B has complete coverage of the area but the third lowest area weight for each point. Design C has the least point weighting, but covers only one-quarter of the total area. Summary statistics of estimates obtained from the samples are illustrated (Table 4).

VI.1. Comparison of Variance and Means

Means and variances obtained from the samples are shown in Table 5. Assuming that these data are normally distributed, the following conclusions are possible. Coverage of all PSUs with four points/cell provides the closest approximation to the control data. Correlation analysis (Table 6) performed on sample means to determine which design has the most positive association with the inventory data showed that design B has the highest correlation coefficient (Freeman et al, 1978:127).

Table 3. Sample designs characteristics and area weighting of each point

Design				Grids(PSUs)	Pts(SSUs)	Total Pts.	Acres/point
One-stage	Sys.	Unalign.	(A)	163	1	163	640
"	"	"	(B)	163	4	652	160
Two-stage	"	"	(C)	40	20	800	130
"	"	"	(D)	40	16	640	163
"	"	"	(E)	40	10	400	261
"	"	"	(F)	40	4	160	652
"	"	"	(G)	40	(4,10,16)	370	(652,261,163)
Two-stage	Random		(H)	*33	20	660	158

* PSUs are randomly distributed, others are systematically unaligned.
(equal size sampling, except G)

Table 4. Summary statistics of sums and sums of squares for sample data

$\Sigma p(i)^2$	Land Use Class									
	P1	P2	P3	P4	P5	P6	P7	P8	P9/P10*	P11
163 - 1(A)	**27	11	1	-	91	6	-	26	-	1
163 - 4(B)	20.125	4.375	0.438	0.625	76.813	1.438	0.375	8.688	0.063	0.125
40 - 4(C)	5.063	1.250	0.063	-	22.813	0.188	0.063	0.688	-	-
40 -10(D)	4.720	0.530	0.060	0.030	19.320	0.130	0.010	0.820	-	0.010
40 -16(E)	5.051	0.449	0.023	0.027	18.688	0.176	0.004	0.930	0.004	0.016
40 -20(F)	4.985	0.508	0.015	0.028	18.203	0.258	0.003	0.875	0.008	0.010
40(G)- 4	4.188	0.438	-	-	7.500	-	-	0.250	-	-
-10	0.460	0.180	-	0.010	6.610	0.080	-	0.190	-	0.010
-16	0.440	0.016	0.020	0.004	5.230	0.070	0.004	0.480	-	-

* P9: 40-16(E), 40-20(F) ; P10: 163-4(B)
** Frequency of occurrence

$\Sigma p(i)$	Land Use CLass									
	P1	P2	P3	P4	P5	P6	P7	P8	P9/P10*	P11
163 - 4(B)	30.750	10	1.250	1.500	94.250	4.750	1.00	18.750	0.250	0.500
40 - 4(C)	7.750	2.500	0.250	-	26.250	0.750	0.25	2.250	-	-
40 -10(D)	8.500	2.000	0.400	0.300	24.400	1.100	0.10	3.100	-	0.100
40 -16(E)	8.688	1.813	0.250	0.313	23.750	1.448	0.06	3.500	0.063	0.125
40- 20(F)	8.600	1.950	0.200	0.350	23.550	1.550	0.050	3.500	0.150	0.100
40(G)- 4	5.750	1.250	-	-	9.000	-	-	1.000	-	-
-10	1.000	0.600	-	0.100	7.900	0.600	-	0.700	-	0.100
-16	1.813	0.125	0.188	0.063	7.563	0.625	0.063	1.563	-	-

* P9: 40-16(E), 40-20(F) ; P10: 163-4(B)

Key to land use symbols (refer Anderson's classification - Table 1)

Table 5. Comparison of means, variances and frequency of occurrence in cells (Confidence level: 95%)

Land-Use Class		Design A	B	F	E	D	C	G	H	Actual Mean
	n=	163	163	40	40	40	40	40	33	
	m=	1	4	4	10	16	20	4,10,16	20	
P1 - Mean%		16.56	18.86	19.38	21.25	21.72	21.50	21.41	21.82	16.8
Variance%		8.48	8.86	18.67	14.78	15.73	15.52	11.91	20.89	
No.of cells		27	59	15	21	21	22	21	N/A	
P2 - Mean%		6.75	6.13	6.25	5.00	4.53	4.88	4.94	5.45	5.4
Variance%		3.86	2.33	5.93	2.32	1.91	2.10	2.40	2.62	
No.of cells		11	28	6	10	11	12	6	N/A	
P3 - Mean%		0.61	0.77	0.63	1.00	0.63	0.50	0.47	0.30	1.0
Variance%		0.37	0.27	0.39	0.33	0.13	0.08	0.03	0.04	
No.of cells		1	4	1	3	3	3	2	N/A	
P4 - Mean%		-	0.92	-	0.75	0.78	0.88	0.41	2.12	2.0
Variance%		-	0.38	-	0.18	0.15	0.14	*0.02	0.62	
No.of cells		-	4	-	3	4	5	2	N/A	
P2 + P3 + P4=		7.36	7.82	6.88	6.75	5.94	6.26	5.82	7.87	8.4
P5 - Mean%		55.83	57.82	65.63	61.00	59.38	58.00	61.16	54.24	60.0
Variance%		15.13	13.80	26.12	22.40	22.78	21.48	13.42	31.36	
No.of cells		91	131	35	36	36	36	35	N/A	
P6 - Mean%		3.68	2.91	1.88	2.75	3.59	3.88	3.06	4.55	3.3
Variance%		2.18	0.81	1.12	0.64	0.72	1.05	*0.08	1.77	
No.of cells		6	17	3	10	15	15	12	N/A	
P7 - Mean%		-	0.61	0.63	0.25	0.16	0.13	0.16	0.30	0.3
Variance%		-	0.23	0.39	0.06	0.005	0.02	0.007	0.09	
No.of cells		-	3	1	1	1	1	1	N/A	
P8 - Mean%		15.95	11.50	5.63	7.75	8.75	8.75	8.16	11.06	10.7
Variance%		8.22	4.05	3.50	3.18	3.27	2.95	1.12	7.24	
No.of cells		26	56	8	17	21	23	15	N/A	
P9 - Mean%		-	-	-	-	0.16	0.38	-	0.15	0.1
Variance%		-	-	-	-	0.005	0.04	-	0.02	
No.of cells		-	-	-	-	1	3	-	N/A	
P10- Mean%		-	0.15	-	-	-	-	-	-	0.2
Variance%		-	0.04	-	-	-	-	-	-	
No.of cells		-	2	-	-	-	-	-	-	
P11- Mean%		0.61	0.31	-	0.25	0.31	0.25	0.25	-	0.3
Variance%		0.37	0.08	-	0.06	0.08	0.05	*0.03	-	
No.of cells		1	1	-	1	1	1	1	-	
Total %		99.99	99.98	100.03	100.00	100.01	100.03	100.02	99.9	100.1

*error - negative value variance
Key to land use symbols refer Table 1

Table 6. Correlation Coefficient (CC) and Cumulative Difference (CD)

Design	CC	CD	CD*
163 - 1(A)	0.97318	14.69	11.99
163 - 4(B)	0.98742	7.94	6.48
40 - 20(C)	0.98929	18.81	16.45
40 - 16(D)	0.98505	11.00	11.00
40 - 10(E)	0.98950	10.65	10.65
40 - 4(F)	0.98410	12.07	12.07
40 - **(G)	0.98514	11.62	11.62
33 - 20(H)	0.98183	13.81	13.47

* combined % estimates for urban classes

** stratified sample

test: correlation coefficient (r)

$$r = \frac{\text{Cov }(x\ ,\ y)}{(\text{Sd of } x)\ (\text{Sd of } Y)}$$

Cov (x,y) = (avg of products xy) - (avg of x)(avg of y)

The absolute number of sample points for design B is less than designs C and H, but it has higher detection capability for the less prevalent land use classes due to complete coverage of the study area. Comparison of cumulative difference (CD) for all designs also supports the contention that this method best describes the population parameters (Table 6). Cumulative difference measures the overall absolute deviation of estimated means from each design to the actual values from the inventory.

test: cumulative difference (CD)

CD = sum of [x (i) - y(i)]

Cumulative difference (CD*) for most of the designs tends to decrease when urban classes are combined. This suggests improved sample estimates may be achieved by generalizing data classification.

The distribution pattern of primary sampling units for large samples is not as important in improving estimates. Systematic unaligned sample means with 40 PSUs (C, D, E, F, G) did not differ much with the results from the two-stage random with 33 PSUs. Sample variance differences may be largely attributed to the estimators used. Direct comparison between systematically unaligned sample designs and the two-stage random sampling is not possible because the different estimators used contribute to the variation. However, this group of designs seems to provide an improvement on reducing sampling error.

The effect of sample point density and PSUs allocation among designs C, D, E, and F shows that increasing point density does not necessarily provide closer approximation of the actual population parameters. Frequency of occurrence in cells for land use types among designs C-G indicates (Table 5) a stabilizing detection rate for the primary class (i.e., cropland and pasture). This trend is also observed in the residential class, where occurrences vary between 15 and 21. As the sample point density increases the observed frequency stabilizes around a particular value. A sample point density of 10 observational units for each PSU seems adequate to estimate the major land use classes. These four designs are not sensitive to the less prevalent land use classes. Underestimation of these classes suggests the scale used in this sampling study was inadequate to detect small land uses. Improved estimates of these classes may require a full inventory of the study area. These could be separated out for ground survey.

The stratified sampling design (G) yields three negative sample variance estimates (Table 7). These errors occur in the smaller land use classes and are likely due to unequal proportions of the different complexity index classes over the whole study area, or to unequal representation of these classes in the sampled PSU's. The errors in the variance estimates are attributed to the two factors, particularly in smaller land use classes.

Estimating precision of the sample means shows designs A and B with smaller 95% confidence intervals (Table 8). The group of designs with 40 PSUs have consistently higher values, suggesting the conservative estimate due to sampling from a fraction of the population. Estimators for computing

Table 7a. Summary statistics on stratified sampling (design G)

	Stratum Means			Stratum variances			
Land use	p(1)	p(2)	p(3)	$s(1)^2$	$s(2)^2$	$s(3)^2$	v(p)
P1	0.338	0.091	0.151	0.14033	0.03689	0.01594	0.001191
P2	0.074	0.055	0.010	0.02153	0.01467	0.00134	0.002412
P3	-	-	0.016	-	-	0.00155	3.136E-6
P4	-	0.009	0.005	-	0.00091	0.00034	-2.276E-6
P5	0.529	0.718	0.630	0.17142	0.09392	0.04247	0.001342
P6	-	0.055	0.052	-	0.00467	0.00341	-8.320E-6
P7	-	-	0.005	-	-	0.00034	6.880E-7
P8	0.059	0.064	0.130	0.01193	0.01449	0.02520	0.000112
P11	-	0.009	-	-	0.00091	-	-2.964E-6

* P9,P10 land use classes are not detected

Table 7b. Complexity Index Classes

Stratum(h)	N(h)	n(h)	W(h)	n(h)/n	n(h)/N(h)
I	83	17	50.9	42.5	0.2048
II	29	11	17.8	27.5	0.3790
III	51	12	31.3	30.0	0.2350
Total	163	40	100.0	100.0	

Table 8. Estimating precision of sample means at 95% confidence level (p ± . 1.96 s/√n)

Land-Use Class		A	B	F	E	D	C	G	H	Actual Mean
		Design								
	n=	163	163	40	40	40	40	40	33	
	m=	1	4	4	10	16	20	4,10,16	20	
P1 - Mean%		16.56	18.86	19.38	21.25	21.72	21.50	21.41	21.82	16.8
Interval		4.47	4.57	13.39	11.91	12.29	12.21	10.70	15.57	
Covers		+	+	+	+	+	+	+	+	
P2 - Mean%		6.75	6.13	6.25	5.00	4.53	4.88	4.94	5.45	5.4
Interval		3.02	2.34	7.55	4.72	4.28	4.49	4.80	5.52	
Covers		+	+	+	+	+	+	+	+	
P3 - Mean%		0.61	0.77	0.63	1.00	0.63	0.50	0.47	0.30	1.0
Interval		0.93	0.80	1.94	1.78	1.12	0.88	0.54	0.68	
Covers		+	+	+	+	+	+	+	+	
P4 - Mean%		-	0.92	-	0.75	0.78	0.88	0.41	2.12	2.0
Interval		-	0.95	-	1.31	1.20	1.16	0.44	2.69	
Covers		-	no	-	+	no	+	no	+	
P5 - Mean%		55.83	57.82	65.63	61.00	59.38	58.00	61.16	54.24	60.0
Interval		5.97	5.70	15.84	14.67	14.79	14.36	11.35	19.11	
Covers		+	+	+	+	+	+	+	+	
P6 - Mean%		3.68	2.91	1.88	2.75	3.59	3.88	3.06	4.55	3.3
Interval		2.27	2.38	3.28	2.48	2.63	3.18	0.88	4.54	
Covers		+	+	+	+	+	+	+	+	
P7 - Mean%		-	0.61	0.63	0.25	0.16	0.13	0.16	0.30	0.3
Interval		-	0.74	1.94	0.76	0.22	0.44	0.26	1.02	
Covers		-	+	+	+	+	+	+	+	
P8 - Mean%		15.95	11.50	5.63	7.75	8.75	8.75	8.16	11.06	10.7
Interval		4.40	3.09	5.80	5.53	5.60	5.32	3.28	9.18	
Covers		no	+	+	+	+	+	+	+	
P9 - Mean%		-	-	-	-	0.16	0.38	-	0.15	0.1
Interval		-	-	-	-	0.22	0.62	-	0.48	
Covers		-	-	-	-	+	+	-	+	
P10- Mean%		-	0.15	-	-	-	-	-	-	0.2
Interval		-	0.31	-	-	-	-	-	-	
Covers		-	+	-	-	-	-	-	-	
P11- Mean%		0.61	0.31	-	0.25	0.31	0.25	0.25	-	0.3
Interval		0.93	0.43	-	0.76	0.88	0.69	0.54	-	
Covers		+	+	-	+	+	+	+	-	
Total (Mean) %		99.99	99.98	100.03	100.00	100.01	100.03	100.02	99.9	100.1

*error - negative value variance
Key to land use symbols refer Table 1
Interval = ± 1.96 s/√n

confidence intervals weighted the standard deviation with the square root value of sample size (n). Consequently, designs A and B, with four times as large a sample size, will be weighted much less. This confidence statement may not be appropriate to use for analysis, except for design A, because the designs are not of random samples (Freeman et al, 1978).

VI.2. Approximation of Population Distribution

The distribution of the population elements in the study area is highly linearalized and concentrated. A distinct corridor pattern of urban type land use classes occurs along the highways between Forest Grove, Hillsboro, and West Slope. It runs through the middle of the study area in an East-West direction, dividing it into two approximately equal segments. The residential class is also congregated along this major artery which connects the test area to the city of Portland. The most prevalent class, cropland and pasture, comprises most of the northwest and southern part of the area. Forest area and other land use types are widely scattered throughout the area.

Designs A and B provide the closest approximation of the spatial distribution of population parameters (Figure 4). Design B with greater point density per grid cell captures the major land use patterns of the area. The 40-PSU sample designs (C, D, E, F, and G) only sample a fraction of the study area; thus, they do not provide adequate information to approximate spatial population distributions because of non-random spatial distribution of land use classes.

V.3. Relative Precision of Sampling Designs

Direct comparisons among the eight sampling methods considered in this study is not possible because of varying sample sizes. But relative precision between sample designs B, D, and H can be determined as these designs have approximately the same number of sample units (652, 640, and 660 respectively). The measure of efficiency of these three different designs can be appraised in terms of deff (design effect). It describes the ratio variance of the estimate from the more complex sample to the variance estimate from a simple random sample of the same number of units (Cochran, 1978). Analysis of results obtained from this computation shows that design B (163-4) has more gain in efficiency in reducing the sampling error than design D (40-16).

VII. Utility of Sampling Information

One of the sampling objectives is to estimate means of specified land uses. Mean proportions of each land use class provide details of the study area composition. Initial data classification precludes certain uses of the sample estimates. In this study, the primary application of the sample data is to provide reliable estimates of major land use classes for a given area and time. This land use information is then used as the basis for comparison between two dates to obtain details of land use change. Quantifiable data on land use change are needed for a synoptic analysis at the national scale and for answering the following questions (Vesterby, 1988):

1. What shifts are occurring among major land use categories?
2. Are shifts to and from certain uses more prevalent than shifts between other uses, such as from cropland to urban?
3. What land uses are changing to urban and at what rates?
4. What are the impacts of population growth within urbanizing areas?

Knowledge of existing details of land resources and the dynamics of land use improve the effectiveness of policies in inventorying, monitoring, planning, and management of land uses. Additionally, at the local level, the sample data can serve to describe land use distributions.

VII.1. Ground Truth

In aerial survey, the criteria of 85% interpretation accuracy is required to support remote sensing investigations. In this project, ground-truthing was conducted before sample data were interpreted and served to increase the investigator's familiarity with the study area. A total of 160 points were included in this analysis. These points comprised a random selection of four observational units in each of the 40 PSUs sampled for designs C-G. The frequency distribution of the observations is indicated in Table 9 based upon the categories of land use. There were 31 interpretation errors made. Besides genuine interpretation error this relatively high percentage error is largely due to conversions of land use to higher intensity uses subsequent to the date of the aerial survey.

Table 9. Error Matrix for Ground Truth Results

Photointerpreted Land use	Observed land Use 11	12	14	17	21	22	23	3	4	5	6	7	Interpreted Total
P1	33	-	-	-	-	-	-	-	1	-	-	-	34
P2	2	7	-	-	-	-	-	-	-	-	-	-	9
P3	-	-	2	-	-	-	-	-	-	-	-	-	2
P4	-	-	-	-	1	-	-	-	-	-	-	-	1
P5	7	3	-	1	76	7	2	-	4	-	-	-	100
P6	-	-	-	-	-	5	-	-	-	-	-	-	5
P7	-	-	-	-	-	-	-	-	-	-	-	-	-
P8	2	-	-	-	1	-	-	-	6	-	-	-	9
P9	-	-	-	-	-	-	-	-	-	-	-	-	-
P10	-	-	-	-	-	-	-	-	-	-	-	-	-
P11	-	-	-	-	-	-	-	-	-	-	-	-	-
Observed Total	45	10	2	1	78	12	2	-	11	-	-	-	160

VII.2. Limitations

Systematic sampling is area weighted (Rosenfield et al, 1982), that is, most sample points chosen are those in categories that cover most of the map area. This led to situations where some small polygons in sparse categories might not be sampled at all. The deficiency of this model can be adjusted by sampling more in an underrepresented areas. Assignment of additional points in these sparse categories will ensure the specified minimum number of points in each category. A problem which might arise concerns how many points to allocate for each land use class. Prior knowledge of the land use distribution is then necessary to make an appropriate estimate.

Another limitation of systematic design is the estimators used for estimate calculation. A common practice is to use a random sample formula for data taken systematically because of simplicity and an assumption of normality. Biased estimates of the sampling error of totals or means may result (Osborne, 1942; Cochran, 1978).

This sampling study also did not measure nonsampling errors. Selection bias and operational errors contributed to this problem. The sampling frame for the secondary units resulted in an unequal probability of selection for points along the edges. These boundary points have a higher chance of being included in the sample. However, since bias is consistent throughout the selection procedure, the effect is assumed to be minimal. Manual operations of grid scribing and transferring designated sample points from base maps to aerial photos constituted the operational errors.

Furthermore, cost and efficiency limit the optimal design(s) for consideration. The choice of systematically unaligned sampling designs is not exhaustive. Other possible variations were not included in the analysis.

Evaluation of the effect of density of SSUs and allocation of PSUs on estimates of population parameters are limited to the designs selected for the project.

Finally, data classification also influenced efficiency of the sampling method. In this study, indistinguishable land use classes on the aerial photos affect the accuracy of sample estimates. Particularly for urban classes, it is difficult to differentiate commercial complexes from urban residential. This problem can be remedied by collapsing these relatively indistinguishable classes into a larger or more general category.

VIII. Summary and Conclusions

The data obtained illustrate the inefficiency of all variations of the systematically unaligned method to estimate less prevalent land use classes. This deficiency is expected for most sample designs. Accurate information on these classes requires census of the study area. The nonrandom nature of land use also contributes to relatively large margins of error in the other classes. The random sampling scheme is also not sensitive to sporadic or concentrated distributions of certain land use classes. Systematically unaligned provides a more representative coverage of the test area, but sample estimates did not indicate much improvement due to certain linear trends of land uses. In this study, sampling all possible PSUs with four points/grid seems to be most efficient because of better representation in the less prevalent classes. Overall, the systematically unaligned sample designs have smaller variances compared to the two-stage random.

Stratification based on the Land Use District Map yields sample estimates which include negative variances for the smaller land use proportions. In the larger classes, the sample estimates have relatively lower variance than estimates obtained from the other designs. Improvement on stratum construction is also needed to ensure representative coverage of the population.

Finally, the choice between a two-stage random or systematically unaligned sampling methods depends on the data required. In this study, data classification needs to be adjusted because certain land use types are consistently under-represented. This affects the overall evaluation of the sample design. If estimates are required only for the major land use types, design E (40 -10), which has relatively minimal sample point density and PSUs allocation, would be an appropriate choice.

Bibliography

Arkin, H. and Colton, R.R. 1970. Statistical Methods, 5th edition. Barnes and Noble Inc.

Barber, M.G. 1988. Elementary Statistics for Geographers, The Guilford Press.

Behm, K. and Pease, J.R. 1985. "A pilot study for estimating land use change in fast growth counties." Unpublished paper, Agreement No. 58-3198-2-00276, Department of Geography, Oregon State University.

Berry, J.L.B. 1962. Sampling coding and storing: Flood plain data. Agriculture Handbook No. 237, U.S. Department of Agriculture, Farm Economics Division, Washington, D.C.

Bircham, D.P. 1979. "Sampling methodology study for land use monitoring in Canada: A detailed methodology and data report." Unpublished paper, Land Use Monitoring Division, Lands Directorate, Environmenta Canada.

Bonnor, M.G. 1975. " The error area estimates from dot grids," Canadian Journal of Forest Research 5(10).

Clawson, M. and Stewart, L.C. 1965. Land Use Information: A Critical Survey of U.S. Statistics Including Possibilities for Greater Uniformity, Resources For the Future Inc., The John Hopkins Press, Baltimore.

Clawson, M., Burnell, Held R. and Stoddard, H.C. 1960. Land for the Future, The John Hopkins, Baltimore.

Frazier, B. and Shovic, F.H. 1980. "Statistical methods for determining land-use change with aerial photographs," Photogrammetric Engineering and Remote Sensing 46(8):1067-1077.

Finney, D.J. 1948. "An example of variation in forest sampling," Forestry (23).

Freeman, D., Pisani, R. and Purves, R. 1978. Statistics, W.W. Norton and Co.

Fitzpatrick-Lins, K. 1981. "Comparison of sampling procedures and data analysis for a land-use and land-cover map," Photo. Eng. and Rem. Sen. 47(3):343-351.

Kalton, G. 1983. Introduction to Survey Sampling, Beverley Hills:Sage Publications.

Greig-Smith, P. 1983. Quantitative Plant Ecology, 3rd ed. Berkeley, California University.

Heimlich, E.R. and Reining, C.R. 1988. "Land use change in fast-growth areas: ERS's role in urbanization research," Unpublished paper, Workshop sponsored by ERS, USDA, The Farm Foundation.

Heimlich, E.R. and Anderson D.W. 1987. "Dynamics of land use change in urbanizing areas: Experience in the Economic Research Service," Sustaining Agriculture Near Cities, edited by William Locheretz, Soil and Water Conservation Society, USA.

Jolliffe, R.F. 1986. Survey Design and Analysis. Department of Mathematics and Statistics, Brunel University, John Wiley and Sons.

Lounsbury, F.J., Sommers, M.L., and Fernald A.E. (Editors). 1981. Land Use: A Spatial Approach, Kendall/Hunt Publishing Co.

Loveland, R.T. 1976. "Urban land use inventorying with small scale aerial photography," Thesis (MS in Sc.), South Dakota State University.

Madow, G.W. and Madow H.L. 1944. "On the theory of systematic sampling, I," Annals of Mathematical Statistics 15:1-24.

Madow, H.L. 1946. "Systematic sampling and its relation to other sampling designs," American Statistical Association Journal 41:204-217.

Maxim, D.L. and Harrington, L. 1983. "Aerial survey design: A systems-analytic perspective," Photogrammetric Eng. and Rem. Sen. 49(10)1425-1435.

Maxim, D.L. and Harrington, L. 1984. "On optimal two-stage cluster sampling for aerial surveys with detection errors," Photo. Eng. and Rem. Sen. 50(11): 1613-1627.

Norton-Griffiths, M. 1988. "Aerial point sampling for land use surveys," Journal of Biogeography 1(15):149-156.

Odeniyi, O.P. 1980 "Land use change analysis using sequential aerial photography and computer techniques," Photo.Eng. and Rem. Sen. 46(11):1447-1464.

Osborne, G.J. 1942. "Sampling error of systematic and random surveys of cover-type areas," American Statistics Association Journal 37:256.

Platt, H.R. 1985. "The farmland conversion debate: NALS and beyond," Professional Geographer 37(4):433-442.

Rosenfield, H.G., Fitzpatrick-Linsk, K. and Ling, S.H. 1987. "Sampling for thematic map accuracy testing," Photo. Eng. and Rem. Sen. 48(1):131-137.

Scheaffer, L.R., Medenhall, W. and Ott, L. 1979. Elementary Survey Sampling, 2nd edition. Duxbury Press.

Senter, R.J. 1969. Analysis of Data: Introductory Statistics for the Behavioral Sciences, University of Cincinnati, Scott, Foreman and Co.

Smartt and Grainger. 1974. "Sampling for vegetation survey: Some aspects of unrestricted, restricted, and stratified techniques," Journal of Biogeography 1:193-206.

USDA. 1986. "Land use change analysis," C-1684, Photointerpretation Training Manual, Earth Sat.

Williams, B. 1978. A Sampler on Sampling, John Wiley and Sons, New York.

Wood, F.W. 1955. "Use of stratified random samples in a land use study", Annals AAG 45:350-367.

Vesterby, M. 1987. Land use change in fast growth counties: An analysis of Study Methodology. Resources and Technology Division, Economic Research Service, USDA, No. AGES880510.

Vesterby, M. 1988. Agricultural Land Use Change in Fast-Growth Areas, Paper presented at the National Governor's Conference.

Vesterby, M. and Brooks H.D. 1988. "Land Use Change in Fast Growth Areas,1950-1980," Paper presented at Land Use Transition In Urbanizing Areas: Research and Information Needs, June 6-7, 1988, Washington, D.C.

Zeimetz et al. 1976. "Dynamics of land use in fast growth areas," Agricultural Economic Report 325, Economic Resource Service, USDA.

Zeimetz et al. 1976. "Using area point samples and airphotos to estimate land use change," Agricultural Economics Research 28(2):1.

Figure 1. Regional map with location of study area
(adapted from Furuseth, 1978)

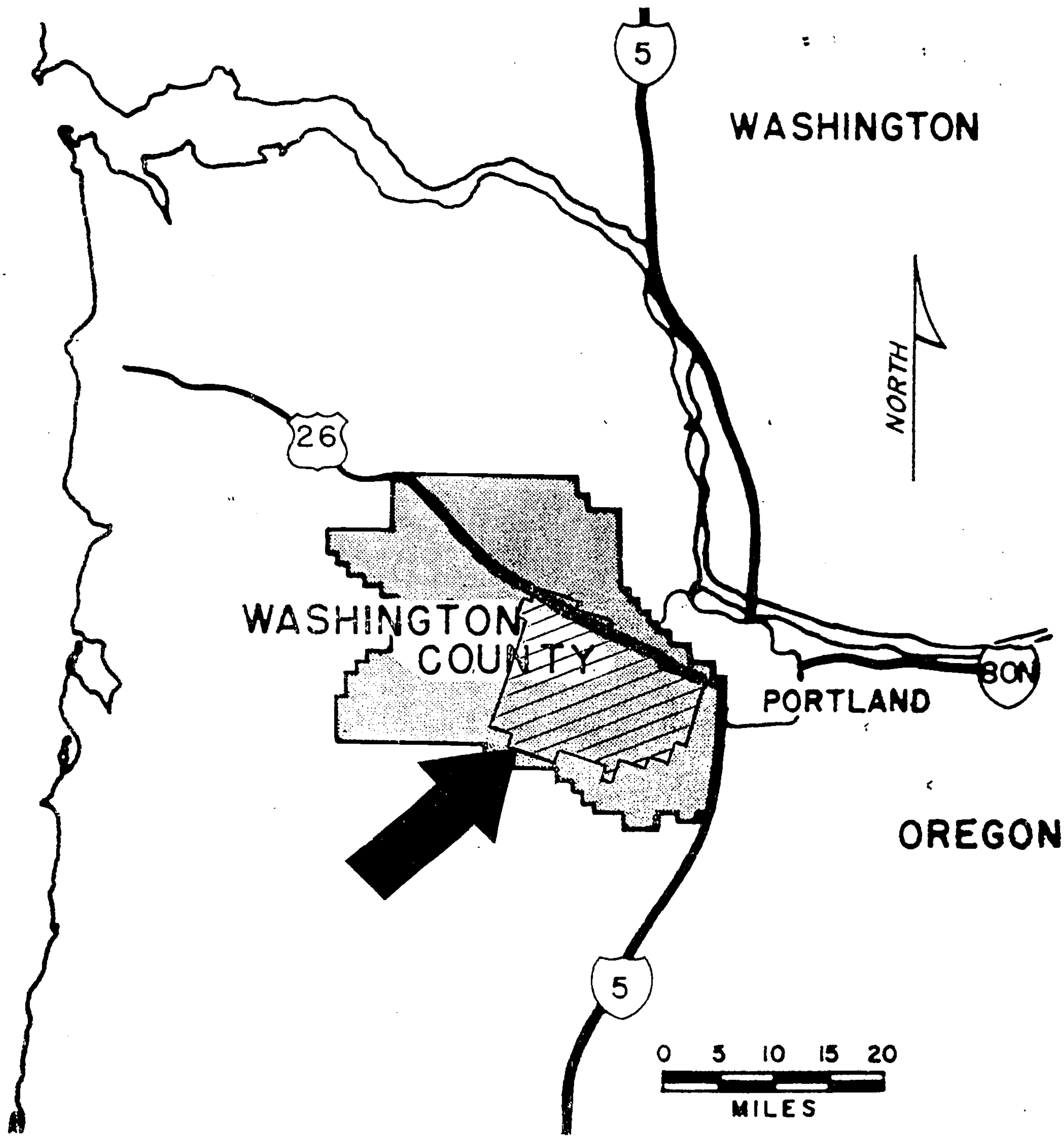


Figure 2. Systematically unaligned sample points
(adapted frpm Stoddard,1982)

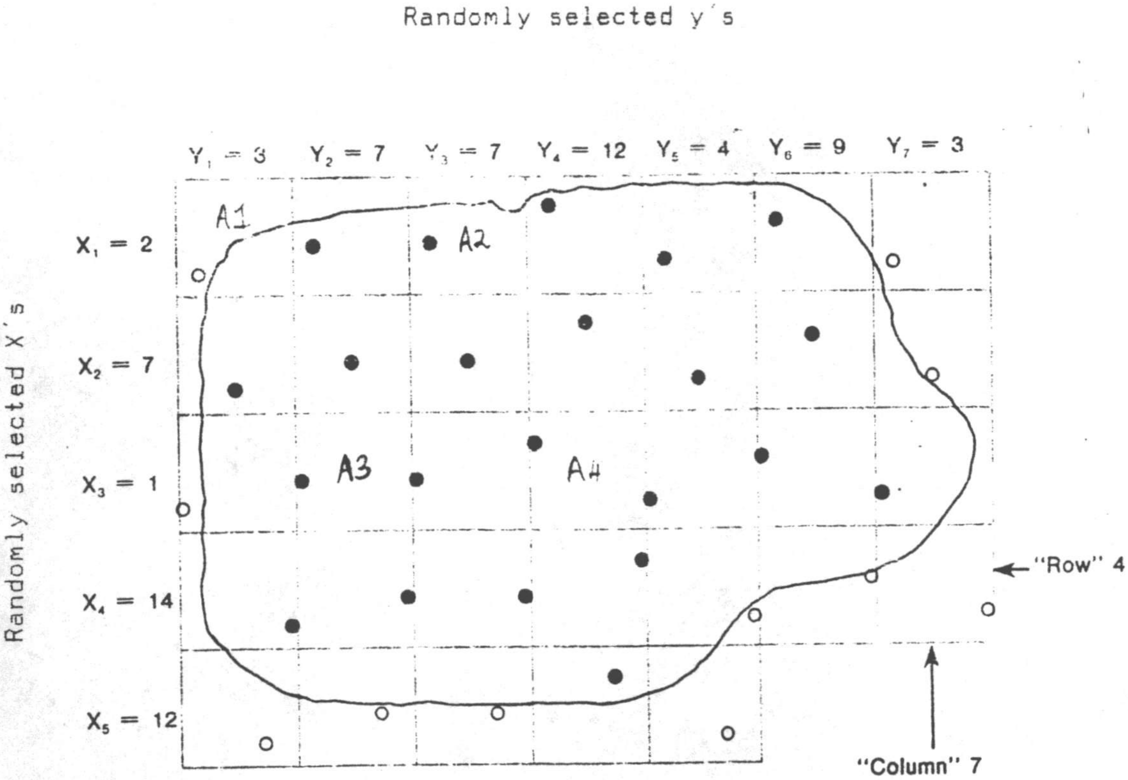


Figure 3. Systematically unaligned sample squares

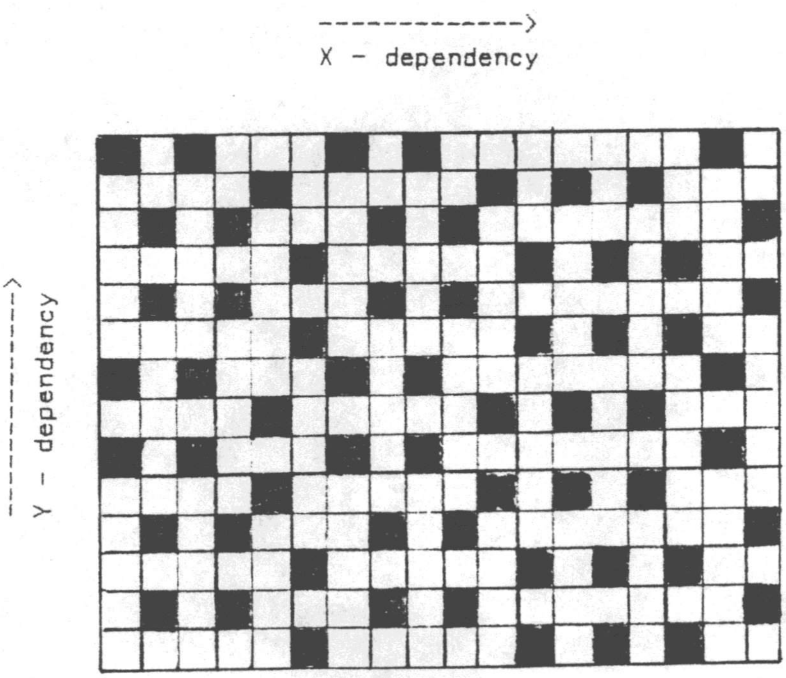
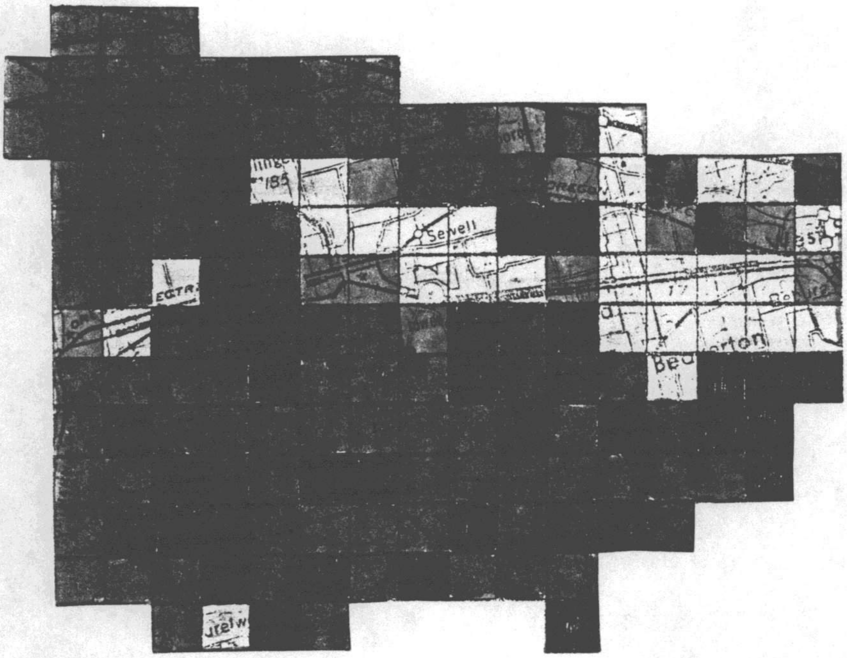
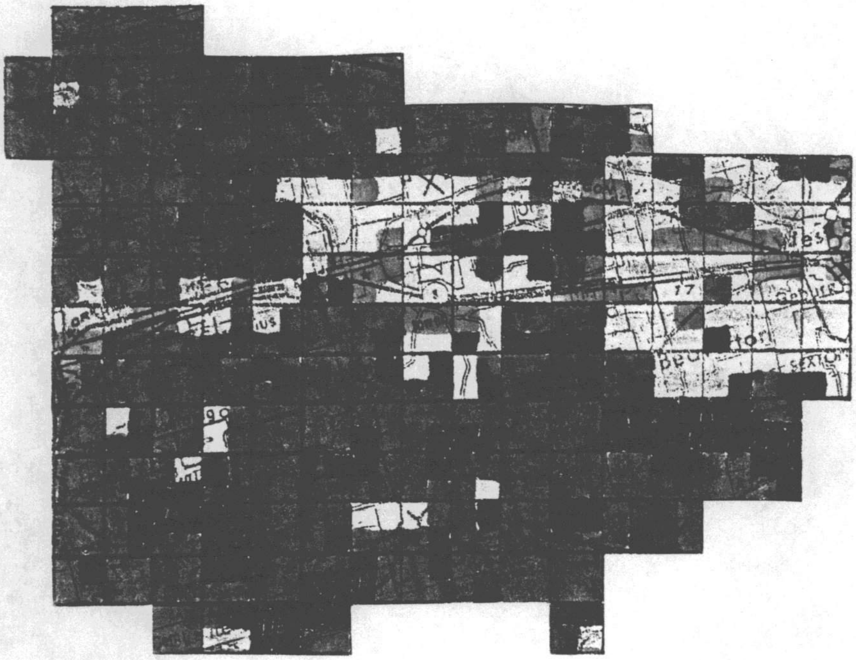


Figure 4. Approximation of population distribution using 2 sample designs

Design A : 163 - 1



Design B : 163 - 4



LEGEND

P1 Residential

P2 Commercial & Services, Industrial, Industrial Complexes, Mixed Urban

P3 Transportation, Communication, &

P4 Urban Built-up Land

P5 Cropland & Pasture

P6 Orchards, Groves, Vineyards, Nurseries & Ornamental Horticultural Areas

P7 Other Agric. Land & Confined Feeding Operation

P8 Forest Land

P10 Wetland

P11 Barrenland