UNITED STATES

DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

INTERAGENCY REPORT USGS-213

AREAL MEASUREMENT ERROR WITH A DOT PLANIMETER: SOME EXPERIMENTAL ESTIMATES

by

Robert S. Yuill
Department of Geography
Dartmouth College
Hanover, New Hampshire

August 1971

Prepared by the U.S. Geological Survey (USGS) for the National Aeronautics and Space Administration (NASA) under NASA contract No. W-13318, Task No. 160-75-71-03. Work performed by Dartmouth College for the USGS Geographic Applications Program under USGS Contract No. 14-08-0001-19958.

This report will not be accessioned by NASA.

TABLE OF CONTENTS

I. INTRODUCTION

II. MEASUREMENT OF A BOUNDED SHAPE

III. MEASUREMENT OF PARCELS WITHIN AN AREA

FIGURES

Figure 1. Test Shapes for Measuring Area with a Dot Planimeter

Figure 2 and 3. Number of Points vs Errors

ABSTRACT

A common geographic and geologic requirement is the measurement, on a map or photo, of areas which vary tremendously in size and shape, such as the different categories of land use within census tracts in and around a city. For such tasks the dot planimeter has obvious advantages.

The literature indicates that in the straightforward measurement of areas the accuracy of the dot planimeter is improved if the problem area is large in size and regular in shape. The present study analyzes shapes more complex than previously studied, and utilizes a computer to simulate a multiplicity of dot grids mathematically. Results indicate that the number of dots placed over an area to be measured provides virtually the entire correlation with accuracy of measurement, the indices of shape being of little significance.

The present study provides equations and graphs from which the average expected error, and the maximum expected range of error, for various numbers of dot points can be read off. Figures 2 & 3 (page 5) make it apparent that point counts of less than 100 are inappropriate for precise measurement, but that 200 dots per area tighten down the error to an average of approximately 1/2 of 1 percent, and the maximum expected error to approximately 3%.

In many practical situations the number of parcels is simply too great (150,000 in the case of the 1970 High Altitude Land Use Map at the Boston Area) and the size of them too small (down to 1 millimeter in size) to permit measurement of each one individually with a mathematical device. Areal sampling is required. Sampling with a dot planimeter can under most circumstances provide an appropriate number of points and permit the measurement in a single operation of more than one heirarchy-level of size (land use categories within census tracts for example). Tests were made with 20 different combinations of parcel patterns, using 20 different randomly oriented sets of dot planimeter measurements on each. The multiple regression equation formulated from these tests provides the basis for accurately quantifying the areas of small-percentage land uses, at the same time that the measurement of the larger, more discrete areas is carried out. Again, number of dots is the critical factor.

AREAL MEASUREMENT ERROR WITH A DOT

PLANIMETER: SOME EXPERIMENTAL ESTIMATES

I. INTRODUCTION

The dot planimeter has occupied a rather equivocal place in areal measurement since its first use by Abel.[1] It has, as Wood pointed out, the advantages of speed and simplicity of operation [2] especially when measuring irregular shapes. Expenditure for equipment is also minimal. On the other hand, it has also been generally regarded as the least accurate of the various area measuring devices.[3]

The number of studies involving areal measurement has greatly increased since Wood in 1954 noted a trend toward quantification.[4] Since many investigators may not have mechanical devices such as the polar planimeter or access to automatic coordinate recording equipment, the dot planimeter remains useful for areas measurement.

Many studies such as the present Dartmouth College Project in Remote Sensing, which involves measurement of land use areas within census tracts, have a twofold requirement for areal measurement. The areas of a number of bounded shapes must be determined and a quantitative measure of the various categories of subdivision within those shapes obtained. Very often the subdivisions are both too numerous to measure individually[5] and too small to be measured accurately by a mechanical device. An areal sampling is thus required, and may be readily obtained by a dot planimeter.[6] Further, in some instances both the area and the subdivisions can be measured simultaneously by a dot grid. It is therefore useful to evaluate the measurement accuracy of the dot planimeter for both of the above uses.

---

[1] C.A. Abel, "A Method of Estimating Area in Irregularly Shaped and Broken Figures," Journal of Forestry, 37 (1939), 344-5.

[2] Walter F. Wood, "The Dot Planimeter, A New Way To Measure Map Area," The Professional Geographer, 6 (1954), 12-14.

[3] John W. Gierhart, "Evaluation of Methods of Area Measurement," Surveying and Mapping, 14 (1954), 463.

[4] Wood, op. cit., 12.

[5] A specific case of this is found in A.R. Stobbs, "Some Problems of Measuring Land Use in Underdeveloped Countries: The Land Use Surveys of Malawi," The Cartographic Journal, 5 (1968), 107.

[6] This assumes of course that there is no problem with spatial periodicity. For further comments on this matter see Brian J.L. Berry and Alan M. Baker, "Geographic Sampling", in Spatial Analysis ed. by Berry & Marble (Englewood Cliffs: Prentice Hall, 1968).

II. MEASUREMENT OF A BOUNDED SHAPE

It has been indicated by a number of authors that the measurement accuracy of a dot planimeter is related to two parameters: size and shape. 1. Size. In itself the size of an area is not particularly important. The critical aspect is the relation of the size to the dot grid spacing. Frolov and Maling, for example, have shown that the order of measurement accuracy is directly related to the number of points counted.[7] Gierhart has shown that measurement error decreases with decreasing grid spacing,[8] which is essentially the same thing. 2. Shape. It has been indicated both by Frolov and Maling,[9] and by Gierhart,[10] that the more serpentine or less compact an area the greater the relative inaccuracy of measurement by a dot planimeter.

Theoretical accuracies for dot counts have been calculated by Frolov and Maling for certain simple geometric shapes. The goal of the present study is to extend their investigation on an empirical basis, using more irregular shapes. Sixteen shapes (Figure 1) were used as the basis for the investigation. The variety of shapes was designed to test the effect of area shape on measurement. Five different-sized, square, dot grids were used to test the effect of the grid size or number of sample points.

A computer sampling process was used to minimize error. The coordinates of the perimeter of each area were read into an especially written computer program. When the program was given a specific grid spacing, it superimposed a mathematical point-grid having that spacing over the figure, with a random orientation. (The orientation of the grid was determined by the computer's pseudo-random number generator, thus eliminating possible human bias.) Using an algorithm based on the Jordan curve theorem,[11] the imaginary grid points within the boundary of the figure were counted. An arrangement was made for the possibility of points falling on the perimeter,[12] but with infinitely small (mathematical) points as a grid, this did not occur in the sample. The familiar problem of whether the cartographic representation of a point was on, within, or outside a drawn line was virtually eliminated.

---

[7] J.S. Frolov and D.H. Maling, "The Accuracy of Area Measurement by Point Counting Techniques," The Cartographic Journal, 6 (1969), 34.

[8] Gierhart, op. cit., 461.

[9] Frolov and Maling, op. cit., 31.

[10] Gierhart, op. cit., 464.

[11] Paul S. Aleksandrov, Elementary Concepts of Topology, trans. by Alan Farley, (New York: Dover, 1961) 16.

[12] Wood's (op. cit.) procedure of counting half the points falling on the perimeter was followed.

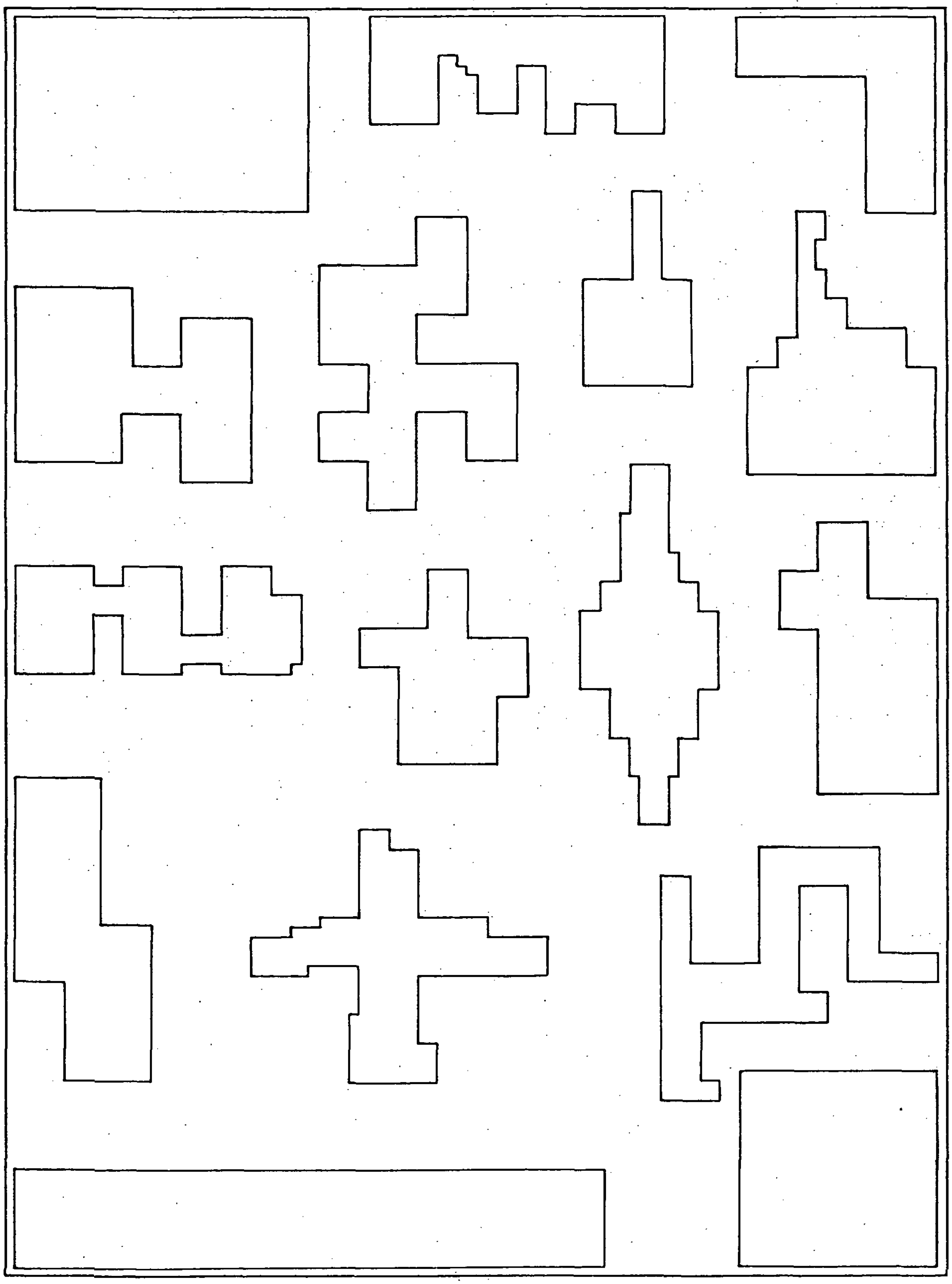


Figure I. Test Shapes for Measuring Area with a Dot Planimeter.

Since a dot grid is a sampling device, a series of measurements under constant conditions must be made to assess its accuracy. For this, the area of each shape was measured 30 times with each of two different-sized grids (one was measured with three) giving a total of 33 series of point-grid counts. From these was calculated the average number of sample points for each series, and the error.

The assumption was made (and experimentally verified on several of the data series) that the members of a series of measurements under constant conditions are normally distributed about the mean. The differences or errors of the measurements from the mean will also be normally distributed. The standard deviation may then be used to describe the dispersion of the measurement errors for each series.

It is assumed further that the mean (μ) of an infinite sample equal the true area of the shape ($\mu = A$). Since the sample means measured here vary so little from the true mean (average difference = .33%), for the purpose of this study they are assumed to equal the true mean. Hence $x = \mu = A$ and the standard deviation measures the dispersion from the shape's true area.

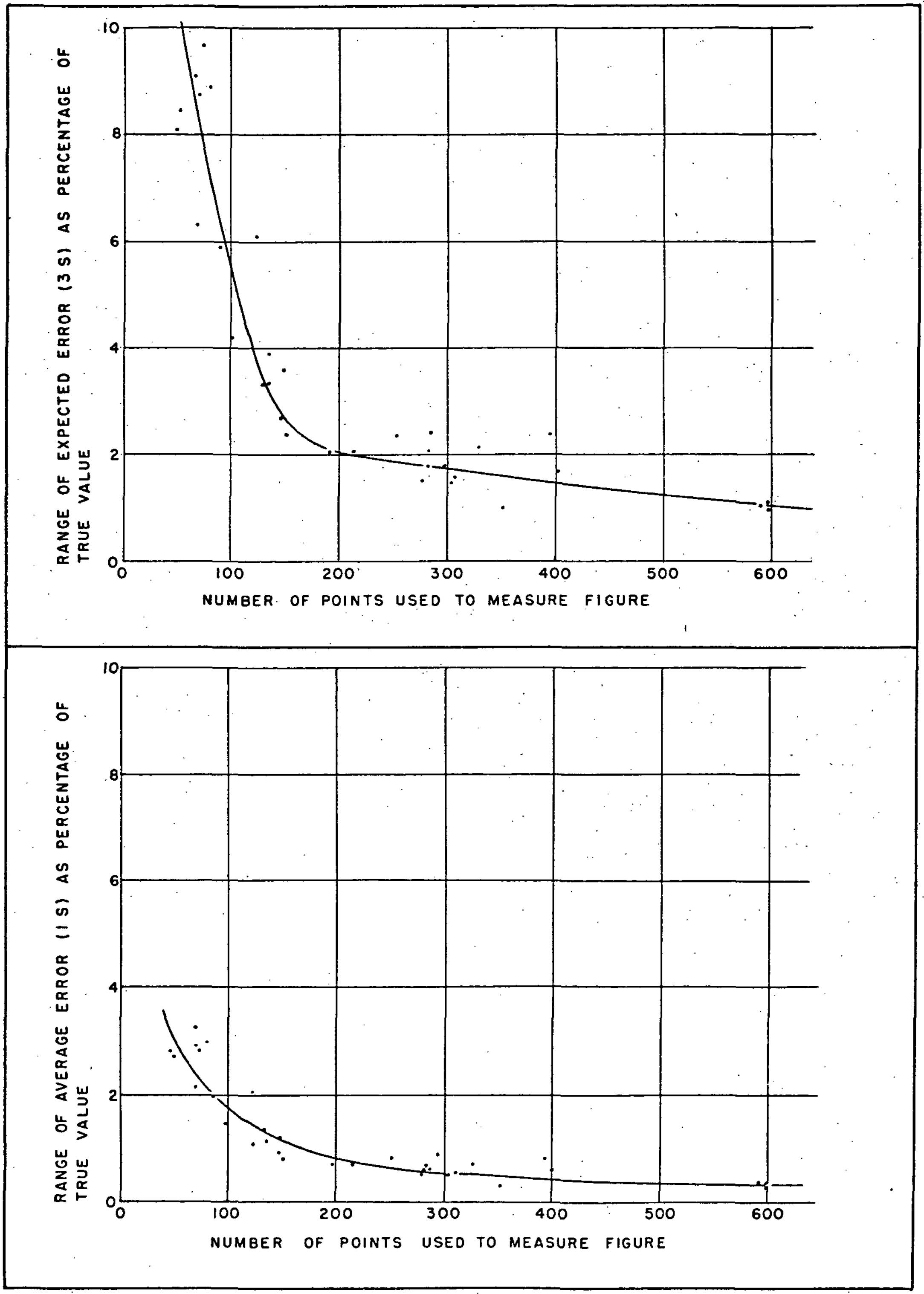
For each series of measurements, the mean and standard deviation define a probability density function. From this, the statistical expectation is that 99% of the individual area measurements of a shape will fall within $\pm$ 3S of the true area of that shape. Concomitantly, an individual measurement has a .99 probability of being within $\pm$ 3S of the true value of the area. It is this error range (3S) which is compared with the size and shape of the measured figures.

The initial comparisons of error to shape and size were made in a multiple regression format. Size was represented by the average number of points used to measure the figure for that series. A variety of indices for shape was used, ranging from the radials of Boyce and Clark[13] to an index developed by the author. None of the shape indices was found to be significantly related at any level to the error range (correlation co-efficients ranged from 0.09 to 0.13). Although this is contrary to expectations, it may be that when the size of the dot sample exceeds a certain threshold, shape becomes relatively unimportant as a factor influencing accuracy.

Since the number of points provided virtually all the explanation in the multiple regression equation, this factor may be treated individually. Figure 2 shows a good curvilinear relationship between the mean number of points and the error range defined above. This relationship may be measured by transforming the variables into a linear equation. To obtain linearity it was necessary to transform both variables (square root on the error range and logarithmic on the number of points). The regression equation is:

---

[13] Ronald R. Boyce, and W.A.V. Clark, "The Concept of Shape in Geography," The Geographical Review, 54 (1964).



Figures 2 and 3. Number of Points vs Errors.

$$\sqrt{Y} = 6.278 - 0.8534 \ln X \qquad \text{Eq. 1}$$

or

$$Y = (6.278 - 0.8534 \ln X)^2 \qquad \text{Eq. 2}$$

WHERE

Y = The expected range of error (3 standard deviations from the mean expressed as a percentage of the Mean)

X = Number of points used to measure the area.

The linear correlation is -0.922, significant at the .001 level. The independent variable (number of points) explains 85% of the dependent variable (error range), a relationship sufficient to serve as a guide for expected errors in dot planimeter measurements.

Equation 2 may be regarded as a calibration curve for estimating the error range of a given size of point sample. For example, with 250 points Equation 2 shows that there is .99 probability that the area calculated by the dot planimeter is within 2.45% of the true value. For 100 points this value would be 5.51%. Although this appears to be a fairly sizable error, this figure represents a <u>maximum</u> expected range of error. The <u>average</u> expected error would be shown by one standard deviation from the mean. At 100 points the average expected error would be 1.89% which agrees with the findings of Rolov and Maling.[14] The relationship of the average expected error to the number of sample points is shown in Figure 3. The corresponding linear regression equation is:

$$Y = \frac{(6.278 - 0.8534 \ln X)^2}{3} \qquad \text{Eq. 3}$$

WHERE

Y = The average expected error (1 standard deviation from the mean)

X = Number of points

One immediate conclusion that may be drawn from the figures and equations is that for precise measurements, point counts under 100 are obviously inappropriate. A minimum of 200-250 points should be employed so that the expected measurement errors fall in the flat part of the curve. However, it is also obvious that error decreases slowly in the flat portion of the curve, so large point counts may be inefficient except when greater precision is desired.

---

[14] Frolov and Maling, <u>op. cit</u>., 34.

A word of caution: the above conclusions have been derived from mathematical grids and lines. Unless dot grids are constructed and used with extreme care, the measurement error ranges will be greater than those shown in Figures 2 and 3.

III. MEASUREMENT OF PARCELS WITHIN AN AREA

Many geographic studies are faced with the problem of measuring the areas or proportions of the subdivisions within some study region. (For convenience, an areal subdivision will be called a land use.) The most accurate mensuration of the land use proportions would entail the measurement of each discrete parcel of each type of land use. However, these parcels are commonly highly irregular and often extremely small, so this method of measurement is beyond the resources of most projects. For example, the Dartmouth College Remote Sensing Project's land use map of the Boston area [15] contains on the order of a 150,000 discrete parcels of land use. The only practical method of area measurement is sampling.

Two major factors were found in a preliminary investigation[16] by the present author to affect accuracy in determining land use proportions by sampling: the number of points used to sample the area, and the percentage of that land use in the study region. Shape was not taken into consideration because it was shown not to be significant in the first part of this paper.

The effects of the two factors were tested by superimposing a square grid of dots over various sample regions and calculating the percentage of the dots falling within specified areas or land uses.

The sample regions were rectangular, constructed on .1 in. engineering grid paper. The .01 in.^{2} cells within these regions, and delimited by the grids, were defined as the land use parcels of the regions. The regions contained from 400 to 1000 parcels. For testing, a given percentage (ranging from 1 to 100%) of the total number of parcels was marked as a particular land use. The locations of these marked cells were randomly assigned within the region.

The dot grid used for sampling had a grid interval of .2 in. This interval was purposely larger than the unit cell size to more closely simulate actual conditions of measurement. Under these conditions, a grid point

---

[15] Robert B. Simpson, Production of a High Altitude Land Use Map and Data Base for Boston, (Hanover, New Hampshire: Dartmouth College Project in Remote Sensing, 1970).

[16] Ibid., Annex C.

falling upon a cell over-represents the area covered by that cell. This over-representation is compensated by the fact that the superimposed grid will not fall upon all the cells of the given land use.

Following the procedure described in Section II, each of 20 different combinations of regions and percentages of land use was tested by a series of 20 measurements, with the dot grid randomly oriented for each measurement. For each measurement the total number of points and the number of points falling in the cells of the given land use were recorded. From these and the percentage of the land use in the area were obtained three variables: the total number of points used to measure the region, the actual percentage of the region given over to this particular land use, and the error or deviation of the measured percentage from the true percentage. (This latter variable is predicated upon the same assumptions about the universal and sample means as was used in Section II, and is measured in the same manner: the standard deviation.)

The relationships of the variables are expressed in a multiple regression equation. For linearity, the variables were transformed as in Section II. (The new variable, the area percentage, also showed a curvilinear relationship with the error range, so had to be transformed.) The equation is:

$$\sqrt{Y} = 29.64 - 3.203 \ln X_1 - 2.317 \ln X_2 \qquad \text{Eq. 4}$$

or

$$Y = (29.64 - 3.203 \ln X_1 - 2.317 \ln X_2)^2 \qquad \text{Eq. 5}$$

WHERE:

Y = The expected range of error (3 standard deviations) from the true proportion of land use expressed as a percentage of the true proportion.

X_1= Number of points required to measure entire region

X_2= True areal percentage of the given land use (Range = 0. to 100).

The multiple correlation is .932 and the r^2 is .868, significant at the .01 level of confidence (F - ratio). Given the strength of the relationship, an expected range of error can be predicted. If a region is measured with 100 points and a given land use within the region occupies approximately 20% of the area, there is a .99 probability that the measurement of that land use is within ± 63% of the true value. If a land use covers 70% of the same region, the expected error range would be ± 49.7%. Increasing the number of sample points has considerably more effect: 250 points on a land use covering 20% of the area yield an error range of 25%.

It is readily apparent from these examples that for accurate measurement of low percentages of a given land use (e.g. 5 - 10%) a large number of sample points is required. This has important implications in planning the land use measurements of any given region. The trade-offs between accuracy and resources can be calculated in advance of a measuring operation to estimate the optimum allocation of resources.

The above results are experimental values. More detailed investigations may yield greater degrees of explanation or additional variables. In the interim, the calibrating equation derived here may be of utility for both area and sub-area measurements.

###

N72-12530

STANDARD TITLE PAGE FOR TECHNICAL REPORTS	1. Report No. USGS-DO-71-Ø14	2. Govt. Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Aerial Measurement Error with a Dot Planimeter: Some Experimental Estimates	5. Report Date August 1971 6. Performing Organization Code
7. Author(s) Robert S. Yuill, Dartmouth College	8. Performing Organization Rept. No. IR-USGS-213
9. Performing Organization Name and Address U.S. Geological Survey Geographic Applications Program Washington, D.C. 2Ø242	10. Project/Task/Work Unit No. 16Ø-75-71-Ø3 11. Contract/Grant No. NASA W-13318 USGS 14-Ø8-ØØØ1-19958
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 2Ø546	13. Type of Report & Period Covered Final report 14. Sponsoring Agency Code

15. Supplementary Notes Prepared by Dartmouth College for the USGS Geographic Applications Program under USGS contract no. 14-Ø8-ØØØ1-19958.

16. Abstracts The present study analyzes shapes more complex than previously studied, and utilizes a computer to simulate a multiplicity of dot grids mathematically. Results indicate that the number of dots placed over an area to be measured provides virtually the entire correlation with accuracy of measurement, the indices of shape being of little significance. Equations and graphs are provided from which the average expected error, and the maximum range of error, for various numbers of dot points can be read.

17. Key Words and Document Analysis. (a). Descriptors

Photointerpretation, Measuring instruments, Aerial photography, Measurement accuracy

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group 8E 8F

18. Distribution Statement	19. Security Class (This Report)	21. No. of Pages
Unclassified - Unlimited	UNCLASSIFIED	13
	20. Security Class. (This Page)	22. Price
	UNCLASSIFIED	$3.00

FORM NBS-667(1-70)