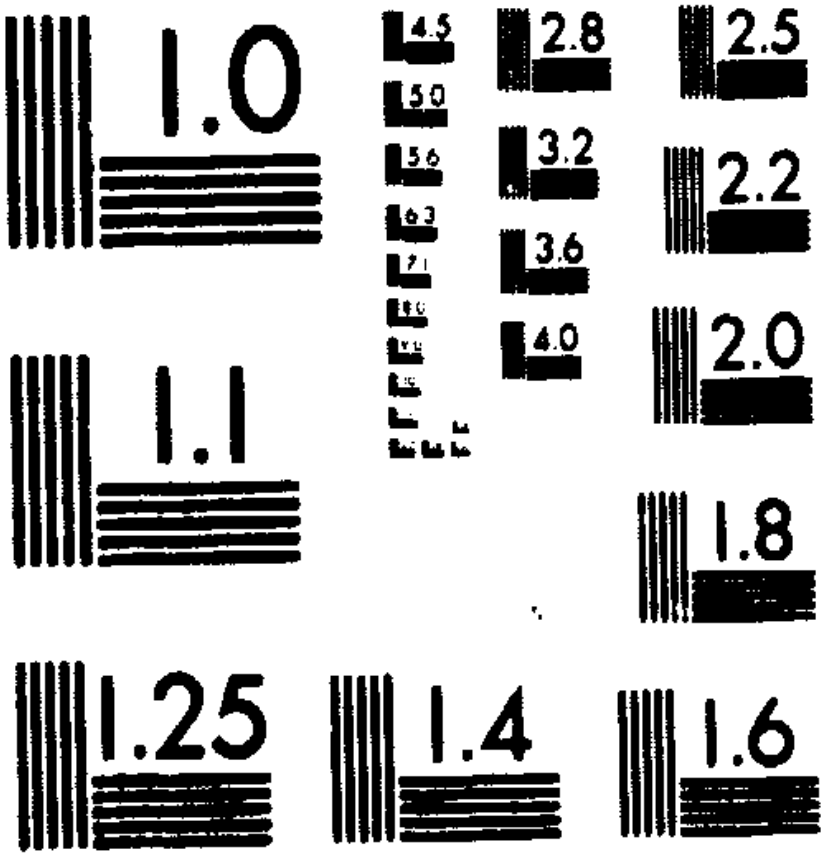

AgRISTARS

SR-L0-00460 80-10272
JSC-16362
JUN 26 1980 NASA CR-160730

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

June 1980

Supporting Research

(E80-10272) SAMPLING OF RECTANGULAR REGIONS (Lockheed Engineering and Management) 16 p HC A02/MF A01 CSCL 08B

N80-29819

Unclas
G3/43 00272

SAM

George Terrell

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
1830 NASA Road 1, Houston, Texas 77058

SR-LO-00460
JSC-16362

TECHNICAL REPORT

SAMPLING OF RECTANGULAR REGIONS

Job Order 73-306

This report describes classification activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

George Terrell

APPROVED BY

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright, Manager
Development and Evaluation Department

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.

Under Contract NAS 9-15800

For

Earth Observations Division
Space and Life Sciences Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

June 1980

LEMSCO-14806

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SR-LO-00460, JSC-16362		

4. Title and Subtitle	5. Report Date
Sampling of Rectangular Regions	June 1980
	6. Performing Organization Code LEMSCO-14806
7. Author(s) George Terrell	8. Performing Organization Report No.
9. Performing Organization Name and Address Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058	10. Work Unit No. 63-2307-3306
	11. Contract or Grant No. NAS 9-15800
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058; Technical Monitor: J. D. Erickson	13. Type of Report and Period Covered Technical Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

A method based on numerical quadrature is proposed to estimate parameters of a Landsat segment. The method is used to estimate proportion of small grains in ten Transition Year segments.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
systematic sample small random samples quadrature approach	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	15	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75) NASA — JSC

CONTENTS

TABLES

1. INTRODUCTION

Problems in remote sensing generate large picture element (pixel) data sets from which it is often necessary to extract a subset for more intensive analysis. There is a tendency, because of classical statistical thinking, to treat these subsets as independent random samples, whereas in practice, there are usually important geographical patterns to be allowed for. In this report, a simple method of geographical sampling is proposed which accounts for a wide variety of existing patterns in the pixel data.

2. SAMPLING OF RECTANGULAR REGIONS

To estimate a parameter of a population which can be thought of as having a uniform spatial distribution, it may be necessary to modify the usual considerations of sampling theory. In particular, there may be correlations between spatial neighbors, thus violating the assumptions of independence.

There are several possible approaches to dealing with these assumptions. If it is evident that or one is willing to assume that correlations between neighbors in the variables of interest decrease rapidly with distance, then small random samples can be taken in the hope that, on the average, the sites sampled will be sufficiently separated. Alternatively, a systematic sample may be taken at sites that are intentionally spaced as far apart as practical. For example, if the region is rectangular, it may be subidvided into isomorphic subrectangles and a point sampled as near as possible to the center of each subrectangle. If the correlations over substantial distances fail to be negligible, other strategies are in order.

If estimation is to be linear, the x and y coordinates of each sample or power thereof may be used as covariates. Alternatively, explicit assumptions about the correlations between points may be made, and the estimation adjusted accordingly.

Both of these approaches depend on fairly specific assumptions and can be quite complicated to carry out in practice. Accordingly, an alternative method is proposed herein allowing for spatial correlations. The method is very simple to use and depends on fairly weak assumptions.

3. ALTERNATIVE METHOD

Assume that there is a need to estimate parameters of a large population distributed uniformly over a rectangular region R. Let the parameter

$$a = \frac{1}{\mu(R)} \int \int_R f_a(x,y)dxdy$$

where f_a is square-integrable on R. If (x_i, y_i) are the coordinates of any sample value, then that value is an estimate of $f_a(x_i, y_i)$. Thus, f_a expresses the spatial dependence of whatever is being estimated.

The problem of estimating parameter a from a sample becomes (in this formulation) a problem in quadrature (i.e., numerical integration). A random sample then gives a Monte Carlo estimate of the integral

$$\hat{a} = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_a(x_i, y_i)$$

It is well known that certain systematic choices of points (x_i, y_i) give better approximations to such an integral than nonsystematic choices. The simplest such set of choices is derived by the midpoint rule. For $n \times m$ points (x_i, y_j), $i = 1, \cdots, n$; $j = 1, \cdots, m$ in the rectangle $(a,b) \times (c,d)$. Choose

$$\left[X_i = \frac{2i - 1}{2n} (b - a) + a\right] \times \left[Y_i = \frac{2j - 1}{2m} (d - c) + c\right]$$

The points chosen are the midpoints of a set of isomorphic rectangles subdividing the large rectangle. Then

$$\hat{a} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} f_a(x_i, y_j)$$

The error in this approximation goes down as $1/n^2$, $1/m^2$ and depends on the second derivative of f. Because the weights given to sampled points are all equal, the variance in $\hat{a}$ that does not depend on x and y is the same as for a random sample of size $n \times m$.

It is known from numerical analysis that if we waive the requirement that the systematic sample be evenly spaced and evenly weighted, it is possible to make very large improvements in the accuracy of a numerical integration. Gauss-Jacobi (ref. 1) integration specifies a set

$$0 < Z_{i,n} < 1 \quad ; \quad i = 1, \cdots, n$$

and a set

$$0 < W_{i,n}$$

such that if

$$x_i = Z_{i,n}(b - a) + b \quad ; \quad y_j = Z_{j,m}(d - c) + c$$

then

$$\hat{a} = \sum_{i=1}^{n} \sum_{j=1}^{m} W_{i,n} W_{j,m} f(x_i, y_j)$$

This set is much more economical of sample points than is the midpoint rule. Because the $W_{i,n}$ sample points are not all equal, the variance, due to factors other than x and y values, will be larger than that for the midpoint rule. Properties of the Gauss-Jacobi sampling scheme will be derived in the following section.

4. GAUSS-JACOBI QUADRATURE

The Gauss-Jacobi quadrature is developed in most standard works on numerical analysis or orthogonal polynomials [e.g., Erdelyi (ref. 2)]. However, since this method is apparently not well known among workers in statistical theory, some of the more elementary properties are derived below.

Gauss-Jacobi quadrature on a rectangle is simply the Cartesian product of quadrature on intervals. Therefore, consider the problem of integration on an interval [0,1]. It is well known that square-integrable functions can be approximated arbitrarily well in the L^2-norm by polynomials. Let $f \in L^2$ be a polynomial of degree n and p, such that

$$\int_0^1 (f - p_n)^2 dx < \epsilon \text{ small}$$

Then from Schwartz's inequality (ref. 1)

$$\left[\int_0^1 f - \int_0^1 p_n\right]^2 < \epsilon$$

Thus, approximating f by polynomials allows one to approximate the integral of f by the integrals of polynomials.

An orthonormal basis for L^2 functions is given by the Legendre polynomials on [0,1]. This is a sequence of polynomials $L_0, L_1, L_2, \cdots, L_n$ such that L_i is a polynomial of degree i; $\int_0^1 L_i^2 = 1$; $\int_0^1 L_i L_j = 0$, $i \neq j$; and $L_i(1) > 0$ for all i. This set of properties uniquely characterizes the set L_i. Furthermore, it is easy to see that $L_0, \cdots, L_n$ form a basis for the polynomials of degree $\leq n$. A further fact: all roots of L_i are real, separated, and lie in (0,1). Observe that if L_i had $j < i$ roots in (0,1), a polynomial of degree j, p_j with those roots could be specified; and, away from the roots, it must have either always the same or always the opposite sign to L_i. Thus, $\int_0^1 p_j L_i \neq 0$. But $p_j = a_0 L_0 + a_1 L_1 + \cdots + a_j L_j$ for some constants a_k, so $\int_0^1 p_j L_i = 0$, contradiction.

Supose it is desired to use n points of quadrature to estimate f. Let p_{2n-1} by a polynomial of degree 2n - 1 approximating f pointwise and in L^2. Using formal division of polynomials, one can decompose

$$p_{2n-1} = L_n q_{n-1} + p_{n-1}$$

where

q_{n-1} and p_{n-1} are polynomials of degree n - 1

Then

$$\int_0^1 p_{2n-1} = \int_0^1 L_n q_{n-1} + \int_0^1 p_{n-1}$$

But the first term on the right is zero, since q can be written as a combination of L_i's where $i < n$ and L_n is orthogonal to both of these. Thus, $\int_0^1 p_{2n-1} = \int_0^1 p_{n-1}$ and the problem is reduced to integrating the lower degree polynomial. In order to accomplish this, note that at the points x_i, $i = 1, n$ in (0,1) such that $L_n(x_i) = 0$, $p_{2n-1}(x_i) = p_{n-1}(x_i)$. By standard considerations, p_{n-1} is uniquely determined by its values at the n points x_i; therefore, so is its integral. The process of fitting a polynomial to the pairs $[x_i, p_{2n-1}(x_i)]$ and integrating the polynomial is readily seen to determine a set of weights $W_1, \cdots, W_n$, depending only on n such that

$$\int_0^1 p_{2n-1}(x)dx = \sum_{i=1}^{n} W_i p_{2n-1}(x_i)$$

The sets $\{x_i\}$, $\{W_i\}$ are tabulated and specify the Gauss-Jacobi quadrature of order n (ref. 1). It is as accurate as interpolating a polynomial of order 2n - 1 in f(x) and integrating the result. By contrast, the midpoint rule is equivalent to interpolating a piecewise linear function and is, therefore, prone to much higher error for function f exhibiting global regularity.

5. TEST OF THE QUADRATURE APPROACH

To test the relevance of the quadrature approach, 10 Transition Year (TY) segments were studied. These segments are rectangles, 117 pixels by 196 pixels in dimension. It was hoped to estimate the area proportion of small grains in each segment. The function f is then the density of small-grain cultivation near a point. This was approximated by setting f equal to one for an entire pixel if the pixel were primarily small grains and to zero if otherwise. Ground-truth information was used to obtain this value. For comparison, a 6 by 10 grid was used. The grid averaged values over the 60 pixels which were numbered an odd number of tens (10, 30, 50, 70) in each dimension. This regular grid was the one used in Procedure 1 as the primary type-2 grid. The corresponding 6 by 10 Gauss-Jacobi grid was that set of pixels nearest the n = 6 Gauss points in the first dimension and nearest the m = 10 Gauss points in the second dimension. These pixels and weights are given in table 5-1. The estimated proportions of small grains were then compared to the true proportions obtained from the ground-truth studies. These results are given in table 5-2. The mean squared error for the Gauss-Jacobi procedure was substantially lower than that for the type-2 grid, even though the difference was not significant according to an F-test. The Gauss-Jacobi procedure was essentially unbiased. It is interesting to note that treating each of the two estimates as a random sample (ignoring geographical correlations) would predict that the Gauss-Jacobi estimate would average ~4/3 as much variance as the regular grid because of unequal weights.

TABLE 5-1.- THE PIXEL NUMBER AND WEIGHTS OF THE GAUSS-JACOBI GRID[a]

[Modified from Abramowitz and Stegun (ref. 1)]

First coordinate: 117 pixels (n = 6) to be sampled

Points (pixel number)	Weights
4	0.085662246
20	.1803807865
45	.233956967
73	.233956967
98	.1803807865
114	.085662246

Second coordinate: 196 pixels (m = 10) to be sampled

Points (pixel number)	Weights
3	0.03333567215
14	.0747256746
32	.1095431813
56	.1346333597
84	.1477621124
113	.1477621124
141	.1346333597
165	.1095431813
183	.0747256746
194	.03333567215

[a] $\text{Estimate} = \sum_{i=1}^{10} \sum_{j=1}^{6} f(p_j^1, p_i^2)\, w_j^1\, w_i^2$

TABLE 5-2.- RESULTS OF A COMPARISON OF ESTIMATED PROPORTIONS OF SMALL GRAINS VERSUS TRUE PROPORTIONS OF SMALL GRAINS OBTAINED BY GROUND TRUTH

Segment number	Ground truth, percent	Gauss-Jacobi dots (small grains)			Type 2 dots (small grains)		
		Percent	Δ	$\tilde{\Delta}$	Percent	Δ	$\tilde{\Delta}$
1394	35.45	37.96	2.51	6.3	32.14	-3.31	10.96
1457	47.72	53.79	6.07	36.85	28.81	-18.91	357.59
1472	37.95	25.84	-12.11	146.65	36.21	-1.74	3.03
1518	34.16	40.82	6.66	44.36	33.93	-0.23	0.05
1584	51.58	49.03	-2.55	6.5	48.33	-3.25	10.56
1602	30.42	34.38	3.96	15.68	23.33	-7.09	50.27
1619	47.91	53.00	5.09	25.91	38.33	-9.58	91.78
1668	9.49	3.88	-5.61	31.47	8.33	-1.16	1.35
1909	22.35	22.10	-0.25	0.06	33.33	10.98	120.56
1918	15.02	12.67	-2.35	5.52	12.28	-2.74	7.51
Mean bias, percent			Mean squared error		Mean bias, percent	Mean squared error	
-.14			0.00319		-3.7	0.00654	

6. CONCLUSION

This experiment should be replicated with more segments in order to establish whether or not the observed improvement is due to chance. Further consideration should be given to what other properties of the segment can be estimated more efficiently when using Gauss-Jacobi sampling rather than using regular sampling. Some possibilities are yield (i.e., small-grains production on a more refined scale than 0 and 1) and various parameters of the spectral distribution. Since these two possibilities are more nearly continuous than the mere presence or absence of grain, Gauss-Jacobi sampling may achieve relatively better results in these cases.

7. REFERENCES

1. Abramowitz, A.; and Stegun, I. A.: Handbook of Mathematical Functions. Dover Publishing, 1972, pp. 887-888 and 916-919.

2. Erdelyi, A., et al.: Higher Transcendental Functions. McGraw-Hill, vol. II, Chapter 10, 1953, pp. 160-161.