AUG 1 5 1980
SR-L0-00467
JSC-16370
81-10033
CR-163556

AgRISTARS

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

Supporting Research

July 1980

MINIMUM VARIANCE GEOGRAPHIC SAMPLING

G. R. Terrell

(E81-10033) MINIMUM VARIANCE GEOGRAPHIC SAMPLING (Lockheed Engineering and Management) 12 p HC A02/MF A01 CSCL 08F
N81-12515
Unclas
G3/43 00033

Lockheed Engineering and Management Services Company, Inc.
Houston, Texas 77058

Lyndon B. Johnson Space Center
Houston Texas 77058

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16370; SR-L0-00467		

4. Title and Subtitle	5. Report Date
Minimum Variance Geographic Sampling	July 1980
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
G. R. Terrell	LEMSCO-15179
9. Performing Organization Name and Address	10. Work Unit No.
Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058	11. Contract or Grant No. NAS 9-15800
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered Technical Report
National Aeronautics and Space Administration Lyndon B. Johnson Space Center Houston, Texas 77058 Tech Monitor: Dr. J. D. Erickson	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Resource inventories require samples with geographical scatter, sometimes not as widely spaced as would be hoped. A simple model of correlation over distances is used to create a minimum variance unbiased estimate of population means.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Correlated samples, minimum variance estimation, resources inventories	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	12	

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

JSC Form 1424 (Rev Nov 75)

NASA — JSC

SR-LO-00467
JSC-16370

MINIMUM VARIANCE GEOGRAPHIC SAMPLING

Job Order 73-345

PREPARED BY

G. R. 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

July 1980

LEMSCO-15179

2. A MODEL FOR DISTANCE DEPENDENCE

Assume the parameter to be estimated is μ, and a single sample point will have variance σ^2 and mean μ. Assume further that samples are nonnegatively correlated with each other as a result of being geographically close. Assuming details of geography are not known, an appropriate model may be derived from the following assumptions:

(1) The correlation ρ_{ij} between the sample values X_i, X_j depends only on the Euclidean distance Δ_{ij} between the locations of the samples.

(2) If the sample points i,j,k are collinear witn j between i and k, then the correlation between X_i and X_k depends only on the correlations between X_i and X_j and between X_j and X_k. Mathematically, this says that the partial correlation $\rho_{ik \cdot j} = 0$.

Assumption two simply means that the effect of one point on its neighbors is via its effect on intermediate points, a sort of "domino" action. From the definition of partial correlation (e.g. Timm (1975)) it is an immediate consequence that $\rho_{ik} = \rho_{ij} \cdot \rho_{jk}$. As a result we have

Theorem: If (1) and (2) hold, then $\rho_{ik} = e^{-K\Delta_{ik}}$ for all pairs of sample locations i and k.

Fitting the model thus requires that we estimate the parameter K.

3. MINIMUM VARIANCE UNBIASED ESTIMATION UNDER DISTANCE DEPENDENCE

The sample $X_1,\dots,X_n$ has covariance matrix $\sigma^2 C$ where $C = (e^{-K\Delta_{ij}})$. We wish to find a vector $\vec{t}$ so that $\vec{t}'\vec{X}$ is the minimum variance linear unbiased estimator for μ. The variance of $\vec{t}'\vec{X}$ is then given by $\vec{t}'C\vec{t}$; the unbiasedness constraint is $\vec{t}'\vec{1} = 1$ where $\vec{1}$ is the n x 1 vector of ones. Minimizing the variance and introducing the constraint via Lagrange multipliers we get

$$\vec{t} = \frac{C^{-1}\vec{1}}{\vec{1}'C^{-1}\vec{1}}$$

The variance of the estimator $\vec{t}'\vec{X}$ is then

$$\frac{\sigma^2}{\vec{1}'C^{-1}\vec{1}}$$

Let us denote $\sum C^{-1} = \vec{1}'C^{-1}\vec{1}$ since it is simply the sum of all terms in the matrix. Then

$$\mathrm{Var}(\vec{t}'\vec{X}) = \frac{\sigma^2}{\sum C^{-1}}$$

By contrast, the variance of the sample mean $\overline{X}$ is

$$\mathrm{Var}(X) = \frac{\sigma^2 \sum C}{n^2}$$

Thus the reduction in variance for our estimator compared to the sample mean is

$$\frac{n^2}{\sum C \sum C^{-1}}$$

This is always less than our estimator, since the sample mean is a linear unbiased estimator, and our procedure has minimum variance in this class.

To illustrate take the simple case where the sample points 1, 2, and 3 are equally spaced on a straight line. Thus $\rho_{12} = \rho_{23} = \rho$ and $\rho_{13} = \rho^2$. We obtain that the estimator is

$$\vec{t}'\vec{X} = \frac{1}{3-\rho} X_1 + \frac{1-\rho}{3-\rho} X_2 + \frac{1}{3-\rho} X_3$$

which has variance $\frac{1+\rho}{3-\rho}$ and reduction in variance $\dfrac{9(1+\rho)}{(3-\rho)(3+4\rho+2\rho^2)}$

If ρ =.5 we obtain a reduction factor of roughly .98. Significant reductions are obtained most readily when spacings are very unequal and correlations are high.

4. FITTING THE MODEL

Ideally one would have sound theoretical reasons for choosing k in the correlation model. Lacking that, it is possible to estimate k from large samples from populations that are believed to have k similar to the target population.

If we know the distance Δ_{ij} between the sites of two sample points i and j, then by standard sampling considerations

$$E\left[\frac{(X_i-X_j)^2}{2}\right] = \sigma^2_{error} ,$$

the residual variance associated with one of the sample points knowing the other. But

$$\rho^2_{ij} = 1 - \frac{\sigma^2_{error}}{\sigma^2}$$

so,

$$E\left[(X_i-X_j)^2\right] = 2\sigma^2\left(1-e^{-2K\Delta_{ij}}\right)$$

From the triples (X_i, X_j, Δ_{ij}) we can estimate K (and σ^2) by nonlinear least-squares techniques.

An example (illustrating more than anything else the pitfalls of fitting the model) is shown in Table 1. The corn production in 33 approximately one-square-mile sections in Missouri were obtained and the distances calculated between each pair. The distances are shown plotted against the absolute value of the difference in the percentage of acreage planted in corn in each section. The hoped-for trend is not apparent to the eye; one would expect an increase in mean difference from lower left to upper right. The equation on page 5 was fitted to this data using the Statistical Analysis System (see SAS (1979)) program procedure NLIN. The best fit was k = .17, which would suggest that the correlation drops to 1/2 at a distance

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M 14:15 TUESDAY, JUNE 3, 1980 1

PLOT OF DIFCO*DIST   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

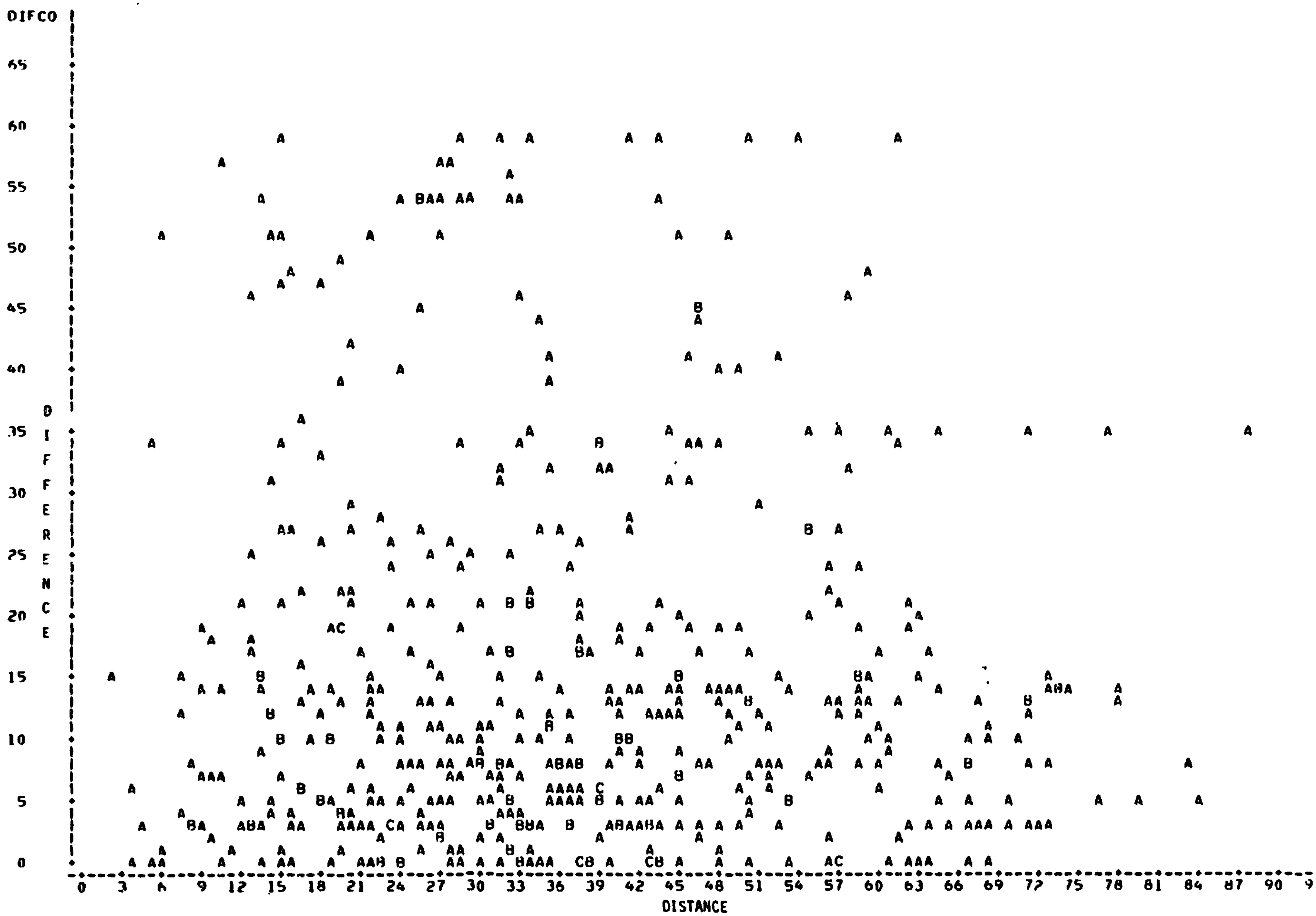

Table 1. Differences in Corn Production vs. Geographical Distance

of four miles. Since this is a very short distance in this population, the sample is close enough to an independent random sample for all practical purposes, and the sample mean is quite adequate. Furthermore, the model did not fit well enough to reassure us even of the positivity of K which is necessary in our model. Thus, the method fails to be helpful for this small data set.

5. CONCLUSION

Further work needs to be done to test the efficacy of this approach when high geographical correlations are present and K can be estimated from either theoretical considerations or extensive previous experience.

6. BIBLIOGRAPHY

SAS Institute, (1979), SAS User's Guide, pp. 317-330. Raleigh, North Carolina.

Timm, Neil H. (1975), Multivariate Analysis with Applications in Education and Psychology, Brooks/Cole Publishing Company.