Some Aspects of Spatial Analysis of Uniformity Data

Janet Shy Modjeska

Inst. of Statistics Mimeo Series No. 1345

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

INTRODUCTION

Soil heterogeneity is one of the largest contributing causes of error in field experiments. Because its effect can be so large as to overshadow those effects being investigated, researchers have been searching for a quantitative expression of soil variability and ways to decrease the effect of soil variability. Quantitative expressions of, or corrections for, soil heterogeneity have been difficult to obtain but several factors have been found useful in diminishing the effect of soil heterogeneity. These factors, dependent on the presence of fertility gradients, on equipment and techniques used in the care and harvest of crops and on cost, are plot size, plot shape and numbers of plots, and attempt to minimize both variance among plots and cost per plot.

Mercer and Hall (1911) estimated variations in yield of various sized plots for both wheat and mangolds, plotted yields across the field in both directions and plotted the coefficient of variation against plot size. The coefficient of variation (c.v.) decreased as plot size increased. They called the point of sharpest change the optimum plot size. Optimum plot sizes for mangolds and wheat were found to be 1/40th and 1/50th of an acre, respectively. Mercer and Hall did not see any superiority of long and narrow plots to square plots for either crop. They concluded that experimental error could be further reduced by "increasing the number of plots similarly treated and by scattering them about the area under experiment, but that there was not much to be gained by increasing the number of scattered plots above five." For practical purposes the authors recommended that "in any field experiment each unit of comparison should be given five plots of 1/40th of an acre each, systematically distributed within the experimental area."

Harris (1915) noted that without special precautions irregularities in the substratum could have a greater influence on numerical results of an experiment than the factors the investigator was seeking to compare. To obtain some measure of soil heterogeneity Harris calculated the correlation among adjacent plots of a given size. Plots were formed using different numbers (m) of unit plots and correlations were computed for each size. As soil heterogeneity increased, the correlation also increased; and the more nearly the correlation approached zero, the more homogeneous the soil. Harris concluded that if an experimental field exhibited irregularities in conditions which influenced in a measurable degree the yield of neighboring small plots, this heterogeneity, measured in terms of the correlation between plots, should become apparently less as the size of plot combined increased.

Batchelor and Reed (1918) investigated the variation in yields of fruit trees taken singly and in groups to determine the effect on variability of various combinations of unit plots. Yields of plots of any size and shape were compared with one another and their variability determined. The c.v., the basis of comparison, decreased as the number of adjacent trees per plot increased. Batchelor and Reed concluded, based on the change in c.v., that little was gained by including more than eight adjacent trees in a plot for all fruit trees studied (oranges, lemons, apples and walnuts). Increasing the number of trees to the plot in systematically distributed units of either four or eight trees gave a more typical sample of the productivity of the total planting than the same number of adjacent trees, with the four-tree unit apparently giving a greater degree of accuracy. They also found that the difference in variability of plots of different shapes was insignificant.

Christidis (1931, 1939) investigated the effect of plot shape and orientation on minimizing the effect of soil heterogeneity. He concluded that in fields with random patchy fertility, plot shape was theoretically immaterial as long as all plots were the same size. With fields containing a gradual change in fertility the longer the plot, the greater the probability of eliminating sectional variations. The c.v. depended primarily on the width to length ratio of the plot and on the plot's angle with respect to the direction of the gradient. Thus, as plot width decreased and its length increased, the c.v. decreased. And as the angle to the gradient increased from 0 to 90 degrees, the c.v. increased. Christidis therefore concluded that in no case could square plots be more uniform than long plots; and the smaller the $w:\ell$ ratio or the more complex the change in soil fertility, the greater the advantage of the long plot.

In 1938 H. F. Smith pointed out that the region of maximum curvature in a plot of c.v. versus plot size depends on the scale chosen and therefore considered it an unsatisfactory method for determining optimum plot size. After combining data from adjacent plots to get yields of different sizes and shapes of plots and determining their variability, Smith regressed log of variance on log of plot size and found that the relative decrease in variance for a relative increase in plot size was approximately constant throughout the range observed. His equation describing soil and plant heterogeneity, adjusted to consider a field of infinite size, is

$$(V_x)_\infty = \frac{(V_1)_\infty}{x^b}$$

where

V_1 = variance among unit plots,

V_x = variance among plots of x unit plots,

x = number of unit plots per plot, and

b = index of soil heterogeneity such that $0 \leq b \leq 1$.

A b of 1 implies independence among the units making up the plot and (V_x) is the variance of a mean of x independent units; a b approaching 0 implies perfect correlation and therefore extreme uniformity within plots up to the size considered. Adding a cost function, the optimum plot size was found to be:

$$x = \frac{bK_1}{(1-b)K_2}$$

where

K_1 = cost of plot regardless of size and

K_2 = cost per unit plot so that total cost per plot is $(K_1 + K_2x)$.

Koch and Rigney (1951) demonstrated that Smith's b could be estimated from experimental data in which treatment effects were present as well as from uniformity trial data. They proposed techniques that simulated uniformity data from analyses of incomplete block and split-plot designs and estimated b using unweighted least squares fit.

Li and Keller (1951) tested the randomness of a set of uniformity data by using a circular definition of serial correlations to obtain information relative to a trend in soil fertility and to compare the relative efficiencies of different sizes and shapes of plots. The yield data (Y_{ij}) were arranged into a vector using a serpentine pattern with respect to either rows or columns of the field and the lag one serial correlation ($\hat{\rho}$) was computed for each sequence. For example, a 3x3 matrix would be considered as a row vector in the sequence 11, 12, 13, 23, 22, 21, 31, 32, 33 and as a column vector in the sequence 11, 21, 31, 32, 22, 12, 13, 23, 33. A pronounced difference in the correlations of the horizontal and vertical sequences, $\hat{\rho}_h$ and $\hat{\rho}_v$, indicated whether rows or columns should be used as replicates. If $\hat{\rho}_h > \hat{\rho}_v$, columns should be used as replicates, and if $\hat{\rho}_v > \hat{\rho}_h$, rows should be used. Their equation for $\hat{\rho}$ was

$$\hat{\rho}_j = \frac{\sum_i X_i \; X_{i+1} - \frac{(\sum_i X_i)^2}{N}}{\sum_i X_i^{\,2} - \frac{(\sum_i X_i)^2}{N}}$$

where

i = element index in data vector for sequence j and

j = h or v.

It was pointed out that the primary problem with this method was that the experimental field is two-dimensional while the $\hat{\rho}$ measures the randomness only of a one-dimensional sequence.

Whittle (1954) computed spatial correlations of various lags in two dimensions based on total variation and covariation using the data of Mercer and Hall (1911) and of

Batchelor and Reed (1918). For a series of mn observations his empirical covariance of lag j,k was

$$C_{j,k} = \frac{1}{mn} \sum_{s} \sum_{t} \xi_{st}\, \xi_{s+j,t+k}$$

where

s is summed from 1 to m-j and

t is summed from 1 to n-k neglecting end effects.

He used those correlations in a model fitting and testing exercise to obtain information on the nature of the two-dimensional autoregression models operating in the field, but not to infer an optimum plot size and shape.

Pearce (1955) added another term to Smith's empirical relationship to account for the variation between trees that was independent of the characteristics of the field. He said the two sources of variation, due to position and due to the unique characteristics of the individual tree, were independent of each other and accounted for them in the equation

$$V_x = \frac{V}{x^b} + \frac{V'}{x}$$

where

$\frac{V}{x^b}$ represents soil heterogeneity and

$\frac{V'}{x}$ represents variation between trees independent of field characteristics.

Pearce concluded that in situations where soil heterogeneity was dominant, small plots gave an economy of land and material relative to the precision being sought and large plots economized labor. Where genetic variation was dominant, V'/x became the important term and plot size had little effect on the accuracy of the experiment and he recommended using the largest plot possible to still have enough degrees of freedom.

Hatheway and Williams (1958) noted that the difficulty in estimating Smith's b which was pointed out by Koch and Rigney (1951) with respect to experimental data (that of highly correlated variance estimates for different-sized plots built up from common components) applied equally well to uniformity data. They also agreed that the simple weighted least squares estimate, weighting by d.f. of each estimate, did not take those correlations into account.

Hatheway and Williams proposed using a nested plot structure from which the orthogonal sums of squares and mean squares could be obtained for computing variances of different-sized plots. When put on a sub-plot basis, the variances of plots of x units (V_i') were shown to be linear functions of the mean squares (V_i) from the nested analysis of variance. From assuming normality of the observations and knowing the linear relationship, the variance-covariance matrix of the V_i' was estimated and generalized least squares was used to obtain an unbiased estimate of Smith's b with asymptotically minimum variance.

Pearce (1976) evaluated Smith's equation and found that the relationship between $\log(V_x)$ and $\log(x)$ was not necessarily linear when calculated from a finite area, that it could be concave up or down at times. When plots were very small, each point was virtually completely correlated with every other point ($b \to 0$). As long as that is true, any change in plot size will have no effect on V_x, which is an unrealistic model. But when plots are very large, each point will be appreciably correlated with only a few other points, giving a b nearer 1. Though Smith's b cannot be constant over x as Smith asserted, Pearce concluded that Smith's equation is almost inevitably a good approximation over small changes in plot size.

As Smith himself pointed out, his equation implies that adjacent areas are equally correlated irrespective of their size and therefore this relationship must break down at some point. Pearce (1976) showed that the relationship held only for small changes in x and not over the range of very small to very large x. Viewing the problem as a spatial correlation problem, it is clear that while Smith's model may be a reasonable approximation, Smith's model imposes fairly rigorous constraints on the behavior of the spatial correlations.

This paper investigates the usefulness of the spatial correlation approach (Whittle, 1954) as an alternative computational procedure for determining optimum plot size. The spatial correlation model assumes only that plots having the same spatial relationship have the same correlation and imposes no constraints on the relationship between correlations of different lags. They can, in fact, go negative or increase with increasing lag, neither of which is possible with Smith's model. By not obscuring the nature of the correlations, it appears that more information may be obtained relative to the real character of the soil heterogeneity.

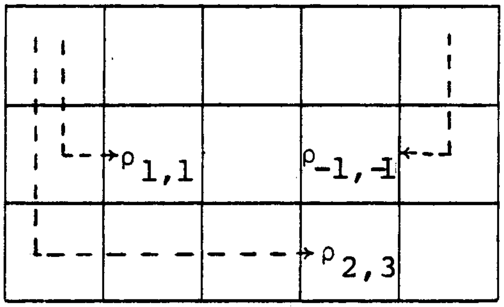
This paper presents analyses of two simulated data sets and five uniformity trials. The analysis consists of computing spatial correlations, covariances and mean square errors for plots of all possible sizes up through one half the dimensions of the field in both directions. Regression coefficients (from plotting the ℓn of the variance among plots of size bxa within blocks the size of the field on a per unit basis divided by the number of unit plots (x), i.e., $\ell n(\mathrm{StdMSE}/x)$, versus $\ell n(x)$) were calculated for all five uniformity trials and compared to those Smith calculated for those same data sets. A method of adjusting MSE's such that they are comparable for blocks less than the size of the field is presented. A singular value decomposition was done on all spatial correlations and is presented where it helps clarify results. Fertility contour maps of uniformity trials and plots of $\ell n(\text{StdMSE}/x)$ versus $\ell n(x)$ are also presented.

MATERIALS AND METHODS

The use of the spatial correlation approach to the analysis of uniformity data was studied in three stages. The first two stages studied the nature of the spatial correlations obtained from a regular response surface with no error and from a uniform response surface with random fluctuations only. The last stage studied results from actual uniformity trial data from which inferences are made as to optimum plot size and shape and block shape and the results compared to those obtained from the conventional Smith's analysis.

Spatial Correlation Model - Theory and Computations

Spatial correlations were calculated from Y_{ij}, the yield of unit plot ij in a field of size BxA, $i = 1,2,\ldots B$ and $j = 1,2,\ldots A$. If ℓ and k define the spatial relationship of any two unit plots in the field, where $\ell = 0,1,2,\ldots B-1$ and $k = 0,1,2,\ldots A-1$, then $p(\ell,k)$ is the correlation between pairs of unit plots a distance of ℓ units apart vertically and k unit plots apart horizontally, ℓ and k defined as before. Because of symmetry of correlations, it is obvious that $p(\ell,k) = p(-\ell,-k)$. For example:



P+ was then defined as the B/2 x A/2 matrix of spatial correlations, $P+ = \{p(\ell,k)\}$, $0 \leq \ell \leq \frac{B}{2}-1$, $0 \leq k \leq \frac{A}{2}-1$ where

$$\hat{\rho}_{\ell,k} = \frac{\sum_{ij} Y_{ij}Y_{i+\ell,j+k} - \frac{(\sum_{ij} Y_{ij})(\sum_{ij} Y_{i+\ell,j+k})}{(B-\ell)(A-k)}}{\left[\left(\sum_{ij} Y_{ij}^2 - \frac{(\sum_{ij} Y_{ij})^2}{(B-\ell)(A-k)}\right)\left(\sum_{ij} Y_{i+\ell,j+k}^2 - \frac{(\sum_{ij} Y_{i+\ell,j+k})^2}{(B-\ell)(A-k)}\right)\right]^{\frac{1}{2}}}$$

and P- was similarly defined as $P- = \{p(\ell,-k)\}$, $0 \leq \ell \leq \frac{B}{2}-1$ and $0 \leq k \leq \frac{A}{2}-k$ where

$$\hat{\rho}_{\ell,-k} = \frac{\sum_{ij} Y_{i,j-k}Y_{i+\ell,j} - \frac{(\sum_{ij} Y_{i,j-k})(\sum_{ij} Y_{i+\ell,j})}{(B-\ell)(A-k)}}{\left[\left(\sum_{ij} Y_{i,j-k}^2 - \frac{(\sum_{ij} Y_{i,j-k})^2}{(B-\ell)(A-k)}\right)\left(\sum_{ij} Y_{i+\ell,j}^2 - \frac{(\sum_{ij} Y_{i+\ell,j})^2}{(B-\ell)(A-k)}\right)\right]^{\frac{1}{2}}},$$

Y_{ij}, B, A, ℓ and k are as defined previously and all summations are over $1 \leq i \leq B-\ell$ for P+, $1 \leq j \leq A-k$, and for P-, $k+1 \leq j \leq A$.

The variance among field plots made of several unit plots is the sum of the variances of the individual unit plots plus the sum of all possible covariances among the unit plots. Assuming similar covariances between pairs of unit plots having the same spatial relation, the variances can be expressed in terms of spatial covariances defined as:

$$\mathrm{cov}_{\ell,k} = \frac{\sum_{ij} Y_{ij}Y_{i+\ell,j+k} - \frac{(\sum_{ij} Y_{ij})(\sum_{ij} Y_{i+\ell,j+k})}{(B-\ell)(A-k)}}{[(B-\ell)(A-k) - 1]}$$

and

$$\mathrm{cov}_{\ell,-k} = \frac{\sum_{ij} Y_{i,j-k}Y_{i+\ell,j} - \frac{(\sum_{ij} Y_{i,j-k})(\sum_{ij} Y_{i+\ell,j})}{(B-\ell)(A-k)}}{[(B-\ell)(A-k) - 1]}$$

where Y_{ij}, B, A, ℓ, k and the limits of summation are as defined previously. Cov(ℓ,k) was defined as the covariance of all pairs of Y_{ij} ℓ units apart vertically and k units apart horizontally. Again because of symmetry, cov(ℓ,k) = cov($-\ell$,$-$k). Because of the equal weighting of cov(ℓ,k) and cov(ℓ,$-$k) in computing the variance among plots of size b x a, their average ($\overline{\mathrm{cov}}(\ell,k)$) was used. The variance among plots of size b x a on a per unit basis is defined as MSE(ba|BA) where

$$MSE(ba|BA) = \frac{1}{ba} [(ba)\overline{cov}_{00} + 2b \sum_k (a-k)\overline{cov}_{0k}$$

$$+ 2a \sum_\ell (b-\ell)\overline{cov}_{\ell,0} + 4 \sum_{k\ell} (b-\ell)(a-k)\overline{cov}_{\ell,k}]$$

and

$$\overline{cov}_{\ell,k} = \frac{cov_{\ell,k} + cov_{\ell,-k}}{2} \quad .$$

The matrix MSE is defined as MSE = {MSE(ba|BA)}, b = 1,2,3,...B/2 and a = 1,2,3,...A/2. These MSE's estimate the variance among plots of size b x a units, on a unit basis, within a block of size B x A, and as such can be compared directly for the relative efficiencies of plots of different sizes and shapes. $[MSE(ba|BA)]/ba = (V_{ba})_{BA}$ in Smith's notation.

The estimates of the spatial covariances as given above can be shown to be biased even under the null model because of common $Y_{ij.}$ in the correction for the two means. Thus, under the null model

$$\mathcal{E}(cov_{\ell,k}) = \begin{cases} \dfrac{-(B-2\ell)(A-2k)}{(B-\ell)(A-k)[(B-\ell)(A-k)-1]}\,\sigma^2 & \text{if } B-2\ell \geq 0 \text{ and } A-2k \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Since σ^2 is also estimated, as $\text{cov}_{0,0}$, each of the biased covariances can be corrected for the null model bias, if necessary, by

$$\text{cov}_{\ell,k} + \frac{(B-2\ell)(A-2k)}{(B-\ell)(A-k)[(B-\ell)(A-k)-1]} \cdot \text{cov}_{0,0} \; .$$

To compare efficiencies of different-sized plots one can either assume a block the size of the uniformity trial and use the MSE's as computed or adjust the MSE's for smaller block sizes according to the procedure that follows.

Let MSE(ba|BA) represent the variance of plots of $b \times a$ units within blocks of size $B \times A$ and suppose a uniformity trial of size $B \times A = N$ is divided in a nested manner into r blocks of size $t \times p$ and each block is subdivided into t plots of p unit plots. The analysis of variance for this nested structure gives the following:

Source	df	Mean Square
total	$(N-1)$	$\dfrac{\sum_{ijk} Y_{ijk}^2 - \dfrac{(\sum_{ijk} Y_{ijk})^2}{N}}{N-1} = MSE(1\mid N)$
among blocks of size tp	$(r-1)$	$\dfrac{\dfrac{\sum_i Y_{i..}^2}{tp} - \dfrac{(\sum_{ijk} Y_{ijk})^2}{N}}{(r-1)} = MSE(tp\mid N)$
among plots within blocks	$r(t-1)$	$\dfrac{\dfrac{\sum_{ij} Y_{ij.}^2}{p} - \dfrac{(\sum_i Y_{i..})^2}{tp}}{p(t-1)} = MSE(p\mid tp)$
among unit plots within blocks	$rt(p-1)$	$\dfrac{\sum_{ijk} Y_{ijk}^2 - \dfrac{(\sum_{ij} Y_{ij.})^2}{p}}{rt(p-1)} = MSE(1\mid p)$.

One can see that

$$MSE(p\mid tp) = \frac{[(rt-1)MSE(p\mid N) - (r-1)MSE(tp\mid N)]}{r(t-1)}$$

$$= \frac{[tMSE(p\mid N) - MSE(tp\mid N)]}{t-1}$$

$$+ \frac{\alpha}{t-1}[MSE(tp\mid N) - MSE(p\mid N)]$$

where $MSE(p\mid N)$ and $MSE(tp\mid N)$ are values taken from the MSE matrix computed in the previous analysis and $\alpha = 1/r$. If α

is small so that the second term above may be ignored, one obtains the equivalent of Smith's (1938) result where he adjusted for block size by letting field size go to infinity. It is helpful to note that $\alpha = 1/r = tp/N$ is the fractional part of the original uniformity study within which one wants the variance of plots of size p. Also, it should be noted that if Smith's model holds,

$$\frac{MSE(p|N)}{p} = \frac{(V_1)_N}{p^b} \quad \text{and} \quad \frac{MSE(p|tp)}{p} = \frac{t(1-t^{-b})(V_p)_N}{(t-1)},$$

Smith's result after adjustment for block size.

For purposes of comparing plot shapes within a given size and shape of block, the adjustment for block size is unnecessary since the adjustment by MSE(tp|N) will be the same in all cases. Conversely, MSE(p|tp) will be minimized for a given shape of plot of size p by choosing the shape of block with the largest MSE(tp|N).

It is clear from the above that the formulation of MSE(p|tp) merely confirms the experimental design principle that the plots and blocks in field design should be chosen so as to simultaneously make the block mean square as large as possible and the plot-to-plot variance as small as possible. Thus, from the precision point of view (ignoring costs) minimization of MSE(p|tp) for a given t will be obtained by finding that plot size and plot shape and block shape (of appropriate dimensions to accommodate the t plots) with maximum difference between MSE(p|N) and MSE(tp|N).

The choice of p will almost inevitably end with $p = 1$ unless costs are considered. Taking cost into account would require a table search over p, with t fixed, for a minimum $C(p|tp)$, where

$$C(p|tp) = \frac{(k_1 + k_2 p)MSE(p|tp)}{p} .$$

The total cost for the experiment is $rt(k_1 + k_2 p)$ and is constant with respect to shape of plot and block (k_1 is cost independent of plot size and k_2 is cost dependent on plot size).

In the event analysis of MSE reveals that Smith's model adequately represents the data,

$$\frac{MSE(p|N)}{p} = (V_p)_\infty = \frac{(V_1)_N}{p^b} .$$

Smith's b can be estimated from the MSE matrix by any one of several methods: nonlinear least squares using the model,

$$MSE(p|N) = V_1 p^{1-b};$$

linear least squares using

$$\ell n(MSE(p|N)) = \ell n(V_1) + (1-b)\ell n(p) ,$$

or a visual fit of the ℓn-ℓn plot, because of the density of the points and the inherent "smoothing" involved in the computations of MSE may be sufficient. Then Smith's formula for optimum plot size can be applied.

In cases where Smith's original model does not adequately represent the field, a two-dimensional version can be proposed as

$$V_{xy} = \frac{(V_1)_\infty}{x^{b_1} y^{b_2}} ,$$

the variance of plots of size x by y where b_1 is the regression coefficient in the x direction and b_2, the y direction. The mean square adjusted for plots of size r x c within blocks of size b x a is approximately

$$MSE(rc|ba) \doteq \frac{[tMSE(rc|BA) - MSE(ba|BA)]}{(t-1)}$$

where t, B and A are as defined previously; and the variance of plots of size r x c within blocks of size b x a is

$$\frac{t(V_1)_{BA}}{r^{b_1} c^{b_2}(t-1)} [1-\alpha^{-b_1}\gamma^{-b_2}]$$

where

$ba = rct,$

$a = \alpha r,$

$b = \gamma c,$ and

$\alpha\gamma = t$.

Substituting into Smith's cost formulation, one sees that the cost per unit of information is

$$\frac{t(V_1)_{BA}}{(t-1)} \cdot \frac{[1-\alpha^{b_2-b_1}t^{-b_2}][k_1 + k_2rc]}{r^{b_1}c^{b_2}} \quad .$$

If $b_1 = b_2 = b$, cost will be minimized with a plot size determined from Smith's one-dimensional model, plot shape and block shape being irrelevant,

$$x = rc = \frac{bk_1}{(1-b)k_2} \quad .$$

For $b_1 \neq b_2$, and assuming without loss of generality that $b_1 > b_2$, cost is minimized by choosing c as small as possible and r as large as possible; i.e., c = 1 and r = optimum plot size, and by choosing α as small as possible for a given t, i.e., $\alpha = 1$. Cost per unit of information therefore is

$$\frac{t(V_1)_{BA}}{(t-1)} \cdot \frac{[1-t^{-b_2}][k_1 + k_2r]}{r^{b_1}} \quad .$$

Taking the partial derivative with respect to r and setting it equal to 0 gives

$$r_{opt} = \frac{b_1k_1}{(1-b_1)k_2} \quad .$$

Thus, in situations where Smith's two-dimensional model holds, optimum plot size is determined by the larger $|b|$ and optimum plot shape will be $1 \times r$, r = optimum plot size, with plots running in the direction associated with the larger b. Optimum block shape will be t such plots wide.

Spatial Analysis of a Strictly Patterned Surface

The relationship between a given response surface and its spatial correlations was investigated by generating a two-dimensional data set of size $B \times A$ and by computing P+ and P-. Two response surfaces were simulated. The first response surface simulated was a regular, cyclical cosine wave. All variation in the field was strictly patterned according to the equation:

$$Y_{ij} = \alpha_1 \cos(\alpha_2 2\pi x_i) + \beta_1 \cos(\beta_2 2\pi y_j)$$

where

$i = 1,2,\ldots B$,

$j = 1,2,\ldots A$,

$x = \frac{i-1}{B-1}$, $0 \leq x \leq 1$,

$y = \frac{j-1}{A-1}$, $0 \leq y \leq 1$,

α_1 and β_1 control amplitude, and

α_2 and β_2 control frequency.

Letting Y_{ij} represent the yield of unit plot ij, all spatial correlations were computed according to the previously defined equations. Inspection of P+ and P- revealed not only the nature of the spatial correlation and the primary fertility pattern, but also the orientation of the major fertility patterns with respect to the axes of the field. Both Y_{ij} and the spatial correlations were plotted for comparison.

Spatial Relationship of a Strictly Random Surface

The relationship between a response surface where all variation is strictly random and its spatial correlations was also investigated. Because the field fluctuated randomly from uniformity, the spatial correlations should estimate zero. This configuration was termed the null case.

Five thousand 8x8 data sets of that type were generated. A standard normal random number, generated by McGill's random number package, was added to a constant to give Y_{ijm}, the data used in all computations:

$$Y_{ijm} = Z + RNOR(0)_{ijm}$$

where

$i = 1,2,3 \ldots B$,

$j = 1,2,3 \ldots A$,

$m = 1,2,3 \ldots 5000$,

Z = a constant, and

RNOR is a $N(0,1)$ random variable.

Spatial correlations were computed from each of the 5000 null data sets. Because of the negative bias in the correlations, another method of computing the variance of different-sized plots was used, that of computing covariances directly from the data. Subsequently, 2000 8x8 null data sets were generated and all possible covariances through lag 4x4 were computed. Though these too were biased, the analysis was continued with this method. Covariances of a given lag were averaged, i.e., $1/2[\mathrm{cov}(l,k)m + \mathrm{cov}(l,-k)m] = \overline{\mathrm{cov}}(l,k)m$. They were also averaged over the 2000 data sets to give a matrix of the following form:

$$\overline{\mathrm{Cov}} = \begin{bmatrix} \overline{\overline{\mathrm{cov}}}_{0,0} & \overline{\overline{\mathrm{cov}}}_{0,1} & \cdots & \overline{\overline{\mathrm{cov}}}_{0,\frac{A}{2}-1} \\ \vdots & & & \vdots \\ \overline{\overline{\mathrm{cov}}}_{\frac{B}{2}-1,0} & & & \overline{\overline{\mathrm{cov}}}_{\frac{B}{2}-1,\frac{A}{2}-1} \end{bmatrix} .$$

The variance among the 2000 averaged covariances of a given lag was computed for all lags (VCOV).

The variance of plots of all possible sizes up through lag 4x4 was computed and put on a per unit basis (MSE(ba|BA)) using the following matrix procedure:

$$\mathrm{MSE} = \frac{(A@B)*(D@E)*}{(F@G)}\ \mathrm{VEC}\overline{\mathrm{COV}}$$

where

$$A = {}_{\frac{B}{2}}\begin{bmatrix} 1 & 0 & . & . & 0 \\ 2 & 1 & 0 & . & 0 \\ 3 & 2 & 1 & 0 & 0 \\ \vdots & & & & \vdots \\ \frac{B}{2} & \frac{B}{2}-1 & \frac{B}{2}-2 & . & 1 \end{bmatrix}_{\frac{B}{2}},$$

$$B = {}_{\frac{A}{2}}\begin{bmatrix} 1 & 0 & . & . & 0 \\ 2 & 1 & 0 & . & 0 \\ 3 & 2 & 1 & . & 0 \\ \vdots & & & \vdots & \vdots \\ \frac{A}{2} & \frac{A}{2}-1 & \frac{A}{2}-2 & . & 1 \end{bmatrix}_{\frac{A}{2}},$$

$$D_{B/2xB/2} = \mathrm{diag}(1 \quad 2 \quad 2 \quad . \; . \; . \quad 2) \;,$$

$$E_{A/2xA/2} = \mathrm{diag}(1 \quad 2 \quad 2 \quad . \; . \; . \quad 2) \;,$$

$$F = \underset{\frac{B}{2}}{}\begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ 2 & 2 & 0 & . & . & . & 0 \\ 3 & 3 & 3 & 0 & . & . & 0 \\ \vdots & & & & & & \\ \frac{B}{2} & \frac{B}{2} & \frac{B}{2} & . & . & . & \frac{B}{2} \end{bmatrix}_{\frac{B}{2}} \;,$$

$$G = \underset{\frac{A}{2}}{}\begin{bmatrix} 1 & 0 & . & . & . & . & 0 \\ 2 & 2 & 0 & . & . & . & 0 \\ 3 & 3 & 3 & 0 & . & . & 0 \\ \vdots & & & & & & \\ \frac{A}{2} & \frac{A}{2} & \frac{A}{2} & . & . & . & \frac{A}{2} \end{bmatrix}_{\frac{A}{2}} \;.$$

$$\mathrm{VECCOV} = {}_{\frac{BA}{4}}\begin{bmatrix} cov_{0,0} \\ cov_{0,1} \\ \cdot \\ \cdot \\ \cdot \\ cov_{0,\ \frac{A}{2}-1} \\ cov_{1,0} \\ \cdot \\ \cdot \\ \cdot \\ cov_{\frac{B}{2}-1,\ \frac{A}{2}-1} \end{bmatrix}_{1} .$$

In the above matrices:

@ represents a Kronecker product,

* represents matrix multiplication, and

– represents matrix division such that $\frac{0}{0} = 0$.

The MSE's were then standardized by dividing each element by the 1,1 element of the MSE column matrix:

$$\mathrm{StdMSE} = \frac{\mathrm{MSE}}{\mathrm{MSE}(1,1)} .$$

Since the MSE's are linear functions of the average covariances, their variances (VMSE) can also be computed as linear functions of the variances of the variance-covariance matrix of the estimated covariances:

$$\mathrm{VMSE} = \left\{\frac{(A*D)@(B*E)}{(F@G)}\right\}*\mathrm{VCOV}*\left\{\frac{(A*D)@(B*E)}{(F@G)}\right\}',$$

A, B, D, E, F, and G as defined previously

and

$$\mathrm{VCOV} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}\sigma_{12} & \cdots & \sigma_{11}\sigma_{\frac{BA}{22}} \\ \sigma_{12}\sigma_{11} & \sigma_{12}^2 & \cdots & \cdot \\ \vdots & & & \\ \sigma_{1A}\sigma_{11} & & & \cdot \\ \vdots & & & \\ \sigma_{\frac{BA}{22}}\sigma_{11} & \cdot & \cdots & \sigma_{\frac{BA}{22}}^2 \end{bmatrix}_{\frac{BA}{4}}$$

$\frac{BA}{4}$

Spatial Analysis of Uniformity Data

The final stage of the analysis used the same estimation scheme in computing variances of plots from ten uniformity trials. Four data sets are taken from Mercer and Hall (1911), mangolds (roots and leaves) and wheat (roots and leaves), and six from Batchelor and Reed (1918): aples, lemons, walnuts and three separate orange data sets.

Of the ten data sets analyzed, only five are presented here. The mangold field was rectangular, 20 rows long and 10 columns wide, with 200 unit plots, each 1/200th of an acre. Each plot was approximately 2.20m x 9.22m, the length running with rows. Roots and leaves were weighed separately and yield expressed in pounds per plot. Yield of roots ranged from 267 to 384 pounds per plot, a deviation of 18 percent from the mean. Mercer and Hall also presented straw and grain yield from a wheat uniformity trial. The field was 20 rows by 25 columns with the 500 unit plots each being 1/500th of an acre (Mercer and Hall, 1911).

Several of Batchelor and Reed's data sets were used. Yield was based on individual fruit trees from orchards which had received uniform treatment from the time of planting. The 1915-1916 yields of 1000 24-year-old navel orange trees came from a grove consisting of 20 trees from north to south with 50 trees in a row, planted 22' x 22'. The grove contained distinct high- and low-yielding areas.

The northeast corner and south end contained notably high-yielding trees and the northern two-thirds of the west side contained a large number of low-yielding trees. This entire orchard and each of its four quarters were analyzed but results of the northeast, northwest and southwest quarters are not presented (Batchelor and Reed, 1918).

Lemon yields were obtained in 1915 and 1916 from a grove of 364 23-year-old trees. The grove consisted of 14 rows extending north and south with 26 trees in a row, planted 24' x 24' apart (Batchelor and Reed, 1918).

The walnut tree records were obtained during the seasons of 1915 and 1916 from a 24-year-old grove. The planting was laid out 10 trees wide and 28 trees long. The trees were planted on the square system, 50 feet apart (Batchelor and Reed, 1918).

Data sets analyzed, but not presented, are Mercer and Hall's mangold leaves and their wheat grain and wheat straw data, and Batchelor and Reed's 33x15 navel orange grove, their 20x12 Valencia orange grove and their 28x8 Jonathan apple orchard.

P+, P- and $\overline{COV}$ were computed for each B x A data set for lags up through a distance of half the width and length of the field. The $\overline{COV}$ matrices were not vectorized in these analyses but were left as matrices. This was done to conserve computer space since variances of covariances were not being computed. Thus:

$$\overline{COV} = \begin{array}{c} \\ \frac{B}{2} \end{array} \begin{bmatrix} \overline{cov}_{0,0} & \overline{cov}_{0,1} & \cdots & \overline{cov}_{0,\frac{A}{2}-1} \\ \overline{cov}_{1,0} & \overline{cov}_{1,1} & & \\ \vdots & & & \vdots \\ \overline{cov}_{\frac{B}{2}-1,0} & \cdot & \cdots & \overline{cov}_{\frac{B}{2}-1,\frac{A}{2}-1} \end{bmatrix}_{\frac{A}{2}} .$$

Variances among plots of size bxa were computed and put on a per unit basis using the following matrix operations:

$$MSE = \frac{SL^2 * D * \overline{COV} * E * SU^2}{N} ,$$

where

$$SL = \begin{array}{c} \\ \frac{B}{2} \end{array} \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 1 & 1 & 0 & . & . & 0 \\ \vdots & & & & & \\ 1 & 1 & 1 & . & . & 1 \end{bmatrix}_{\frac{B}{2}} ,$$

$SL^2 = SL*SL = A$ as defined previously,

$$SU = {}_{\frac{A}{2}}\begin{bmatrix} 1 & 1 & 1 & . & . & 1 \\ 0 & 1 & 1 & . & . & 1 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & . & 0 & 1 \end{bmatrix}_{\frac{A}{2}} ,$$

N = SL*J*SU, the number of unit plots per plot

and

$$J = {}_{\frac{B}{2}}\begin{bmatrix} 1 & 1 & . & . & . & 1 \\ 1 & . & . & . & . & 1 \\ \vdots & & & & & \vdots \\ 1 & . & . & . & . & 1 \end{bmatrix}_{\frac{A}{2}} .$$

Again these MSE's, reflecting variances among plots of size b x a within a block of size B x A, were standardized by the procedure

$$\text{StdMSE} = \frac{\text{MSE}}{\text{MSE}(1,1)} .$$

RESULTS

Strictly Patterned Surface

The first stage of the analysis involved computing the spatial correlation matrix for a regular two-dimensional cosine wave simulating a field with very regular spots of high fertility and with no random fluctuation. Peaks with a value of about 2 and valleys of about 0 were 6 units apart horizontally and vertically and peak to peak or valley to valley distance (or cycle length) was 12 units horizontally and vertically (Figure 1). The spatial correlations reproduced this cycle length of 12 units. Those having lags in multiples of 12 in either direction had a maximum value of 1. Similarly, the spatial correlations $\hat{\rho}(6+12k, 6+12k')$ were -1 for all k and k'. Thus in the extreme case of perfectly regular fertility patterns the spatial correlations reproduce the fertility patterns of the original field (Figure 2).

While this is an unrealistic field, it illustrates that spatial correlations can go negative and can increase in value, a result not permitted in Smith's model, and that cycling of the spatial correlations reflects, in an average sense, the distance between fertility peaks and valleys. If such a cycling pattern were strong enough (if the negative correlations were large enough), the possibility arises that greater efficiency could be attained by increasing plot size to coincide with cycle length. In

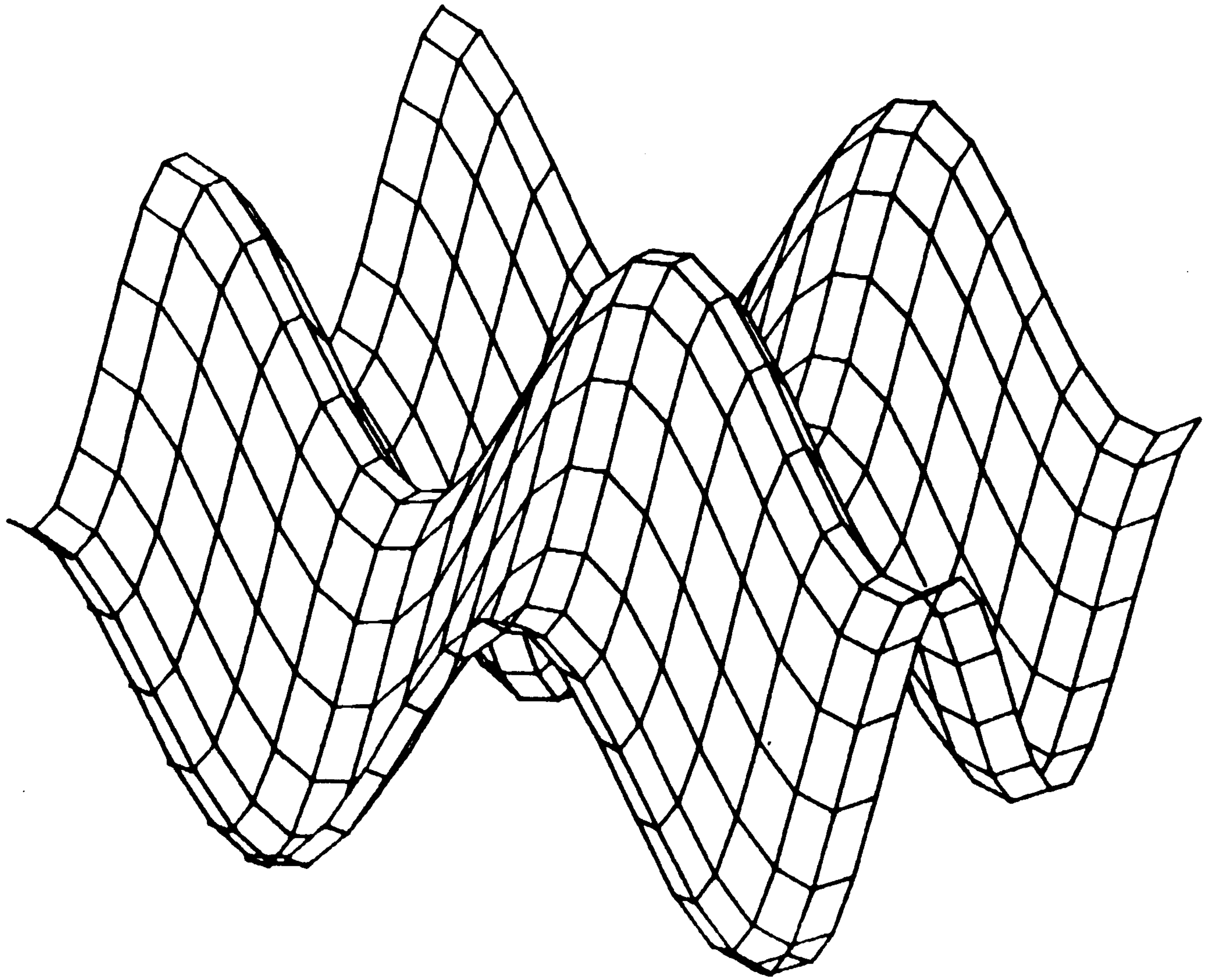


Figure 1. Spatial correlations of a regular two-dimensional cosine wave

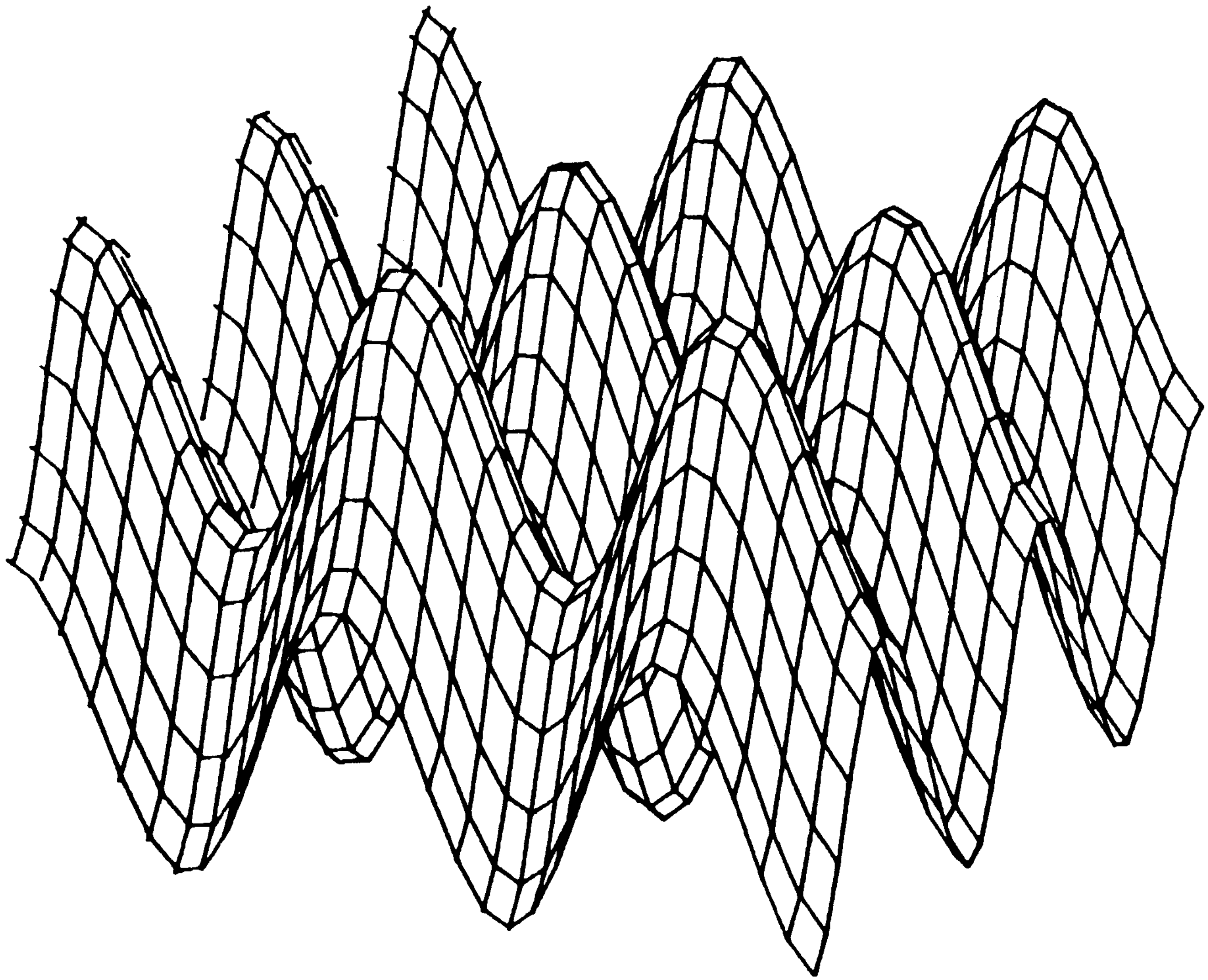


Figure 2. Plot of a regular two-dimensional cosine wave simulating a field with no random fluctuation where amplitude is 1 and frequency is 5

this case because the gradient is at 0 and 90 degrees to the axes of the field and because cycle length is the same along rows and columns, plot shape and plot orientation do not influence optimum plot size. Maximum efficiency would be obtained with plots whose length and width are a multiple of 6.

Strictly Random Surface

The null case investigated a uniform field, i.e., one with no gradients but only random fluctuations. Thus the correlation between any 2 unit plots any distance apart is 0 and the StdMSE's should be 1.

The results of analyzing 5000 such data sets gave evidence of a negative bias among the correlations. The lag (0,0) correlation was 1. Of the remaining 35 correlations calculated from the original 8x8 matrix all were very close to 0, ranging from -0.018 to 0.004, but 25 of the 35 were negative and all correlations in the quadrant of the matrix for $\ell < 4$ and $k < 4$ were negative (Table 1). This proportion is more than expected from a N(0,1) random variable. The matrix of expected biases for the spatial covariances for the null model (and approximately for the spatial correlations since $\sigma^2 = 1.0$) are shown in Table 2. It is clear that correction for this bias removes the observed negative bias when the null model holds. Variances among the 5000 correlations of a given lag were very small, ranging from 0.009 to 0.062.

Table 1. Spatial correlations computed from 5000 data sets for the null case

	k					
ℓ	0	1	2	3	4	5
0	1.000	-0.014	-0.012	-0.011	0.004	0.000
1	-0.016	-0.015	-0.016	-0.010	0.000	0.002
2	-0.018	-0.010	-0.013	-0.009	0.000	0.000
3	-0.007	0.000	0.000	0.002	-0.004	-0.003
4	-0.007	0.000	0.000	0.002	-0.004	-0.003
5	-0.001	-0.003	0.002	0.000	0.008	0.004

Table 2. Expected biases/σ^2 for spatial covariances in an 8x8 field under the null model

		k					
		0	1	2	3	4	5
ℓ	0	u[a]	-0.016	-0.014	-0.010	u	u
	1	-0.016	-0.015	-0.014	-0.010	u	u
	2	-0.014	-0.014	-0.013	-0.009	u	u
	3	-0.010	-0.010	-0.009	-0.007	u	u
	4	u	u	u	u	u	u
	5	u	u	u	u	u	u

[a]Unbiased if $\ell = k = 0$ or if $\ell \geq 4$ and/or $k \geq 4$.

The negative bias in the correlations showed up in the StdMSE's as an apparent gain in efficiency with larger plots. The StdMSE's decreased from 1 (for 1x1 plots) along both rows and columns. The StdMSE for 6x6 plots from the 8x8 field was 0.706, an apparent gain in efficiency of 42 percent over the 1x1 plot (Table 3). Correction for the bias in the spatial covariances gave relative efficiencies very near 1.0 as expected from the null model. The maximum deviation from 1.0 was 0.01. Thus, it is clear that correction for bias of the spatial covariances is necessary if StdMSE's for different sizes of plots are to be compared. For our presentation of the method, however, corrections for bias are not made. Therefore, all MSE's should be viewed with the negative bias in mind. Variance among MSE's of a given lag (VMSE) was also very small, ranging from 0.010 to 0.502.

Covariances were also estimated for the null case from 2000 8x8 trials. The lag (0,0) covariance was 1 and all of the 15 remaining covariances were less than 0, again more than expected from a N(0,1) random variable (Table 4).

The negative bias in the covariances also showed up in the StdMSE's as an apparent gain in efficiency with larger plots. The StdMSE's decreased from 1 to 0.801 for 4x4 plots, an apparent gain in efficiency of 25 percent (Table 5). The StdMSE computed from correlations for 4x4 plots was 0.806 (5000 trials).

Table 3. Standardized mean square errors computed from 5000 data sets

	Number of Unit Plots					
	1	2	3	4	5	6
1	1.000	0.986	0.974	0.962	0.956	0.952
2	0.984	0.955	0.927	0.902	0.889	0.881
3	0.967	0.926	0.884	0.848	0.828	0.816
4	0.955	0.904	0.851	0.806	0.779	0.763
5	0.944	0.888	0.829	0.777	0.747	0.728
6	0.937	0.878	0.812	0.757	0.725	0.706

Table 4. Average covariances computed from 2000 data sets for the null case

	k			
ℓ	0	1	2	3
0	1.000	-0.009	-0.019	-0.011
1	-0.012	-0.016	-0.016	-0.008
2	-0.016	-0.013	-0.013	-0.010
3	-0.018	-0.009	-0.008	-0.006

Table 5. Standardized mean square errors computed from 5000 data sets for the null case

	Number of Unit Plots			
	1	2	3	4
1	1.000	0.988	0.974	0.957
2	0.991	0.962	0.933	0.905
3	0.975	0.931	0.888	0.848
4	0.962	0.906	0.851	0.801

Spatial Analysis of Uniformity Trials

The 26x14 lemon grove appeared to exhibit a fairly regular cycling fertility gradient across the grove with yields ranging from 79 to 510 pounds per individual tree. It was therefore by no means "extremely uniform" as Batchelor and Reed had said. Four vertical strips of alternating high- and low-yielding areas covered the grove (Figure 3). These strips ran the entire length of the field and were approximately three unit plots wide. Yield remained fairly constant within any one of the strips but cycled when going across the four strips. The two gradients are parallel and perpendicular to the axes of the grove.

The fertility gradients were regular enough over the field to be reflected in the correlation matrix (Figure 4). Because yield remained fairly constant in a column, correlations likewise remained fairly constant within a column. Along a row, correlations cycled, decreasing to lag $(\ell,+3)$ or lag $(\ell,+4)$ and increasing again at lag $(\ell,+6)$ for all ℓ; thus the correlation matrix also has strips of high and low correlations two to three units wide running down columns and shows a marked similarity to the fertility contour map of the original grove (Table 6). The spatial correlations show that basic units about three to four units apart and again about seven units apart in a row are essentially independent. Independence was never reached along a

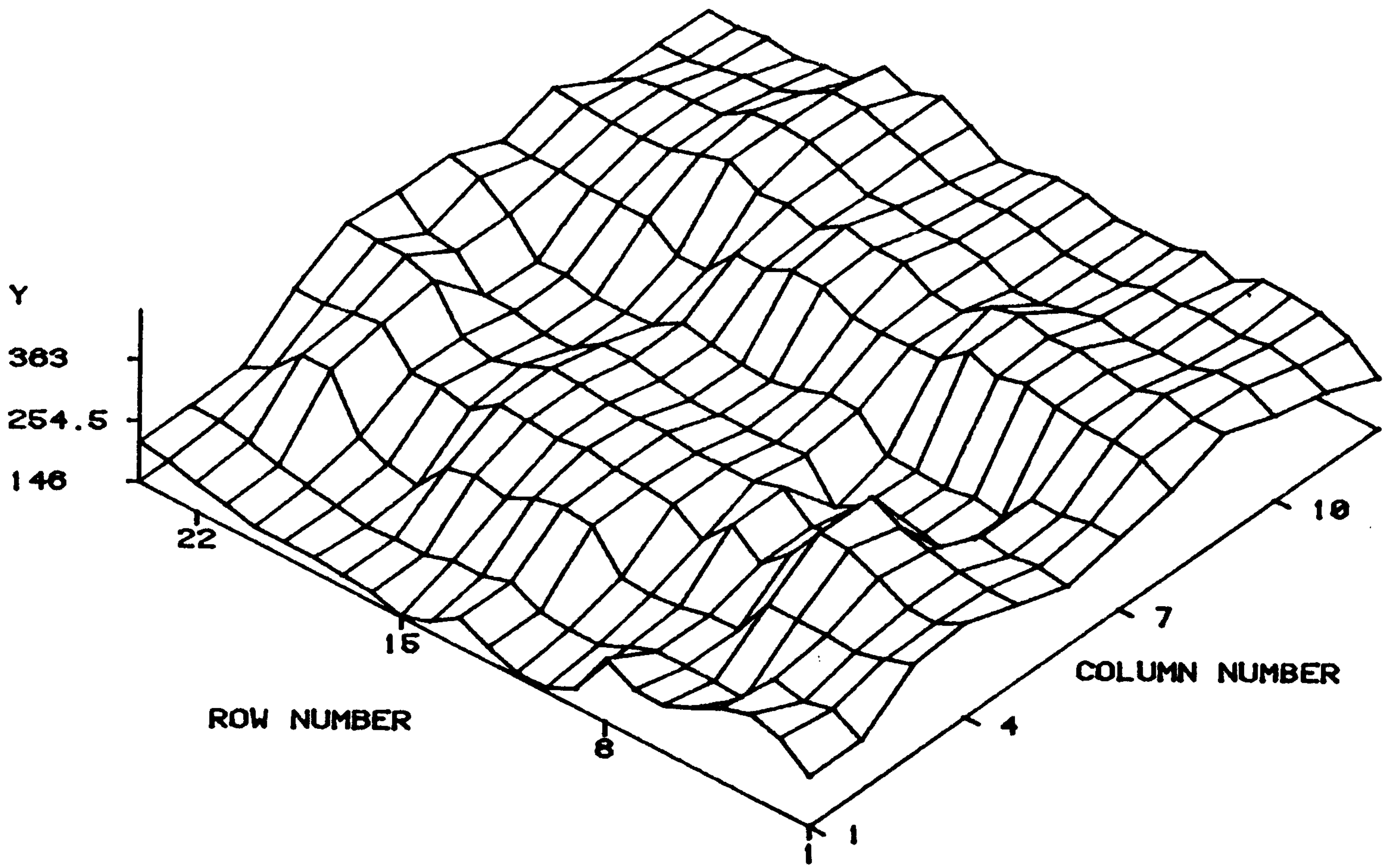


Figure 3. Smoothed response surface using moving medians of all possible 3x3 plots of pounds of lemons per tree from the 26x14 lemon grove (Batchelor and Reed, 1918)

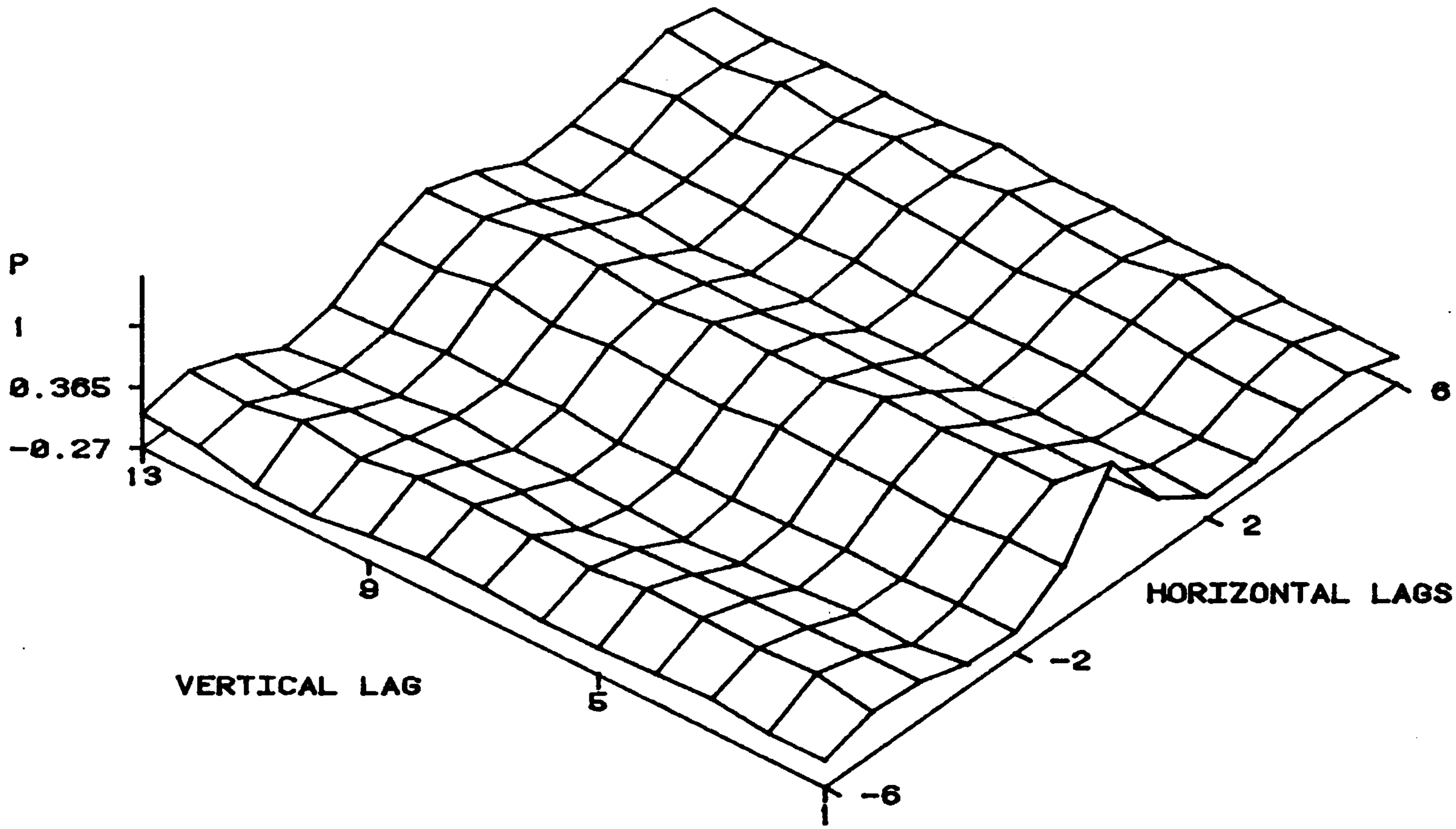


Figure 4. Spatial correlations computed from the lemon data for lags $0 \leq \ell \leq 13$ and $-6 \leq k \leq 6$

Table 6. Spatial correlations of yield in pounds per tree from the 26x14 lemon grove (Batchelor and Reed, 1918)

	k												
ℓ	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
0	0.01	0.17	0.13	-0.03	-0.05	0.28	1.00	0.28	-0.05	-0.03	0.13	0.17	0.01
1	0.02	0.20	0.10	-0.08	-0.02	0.23	0.51	0.28	-0.05	-0.13	0.13	0.27	0.00
2	0.02	0.17	0.18	-0.09	-0.05	0.24	0.50	0.20	-0.09	-0.09	0.11	0.24	0.07
3	0.07	0.31	0.19	0.00	-0.05	0.19	0.48	0.24	-0.08	-0.13	0.10	0.26	0.03
4	0.06	0.20	0.19	-0.01	-0.09	0.22	0.50	0.24	-0.11	-0.17	0.07	0.27	0.01
5	0.10	0.23	0.14	-0.02	-0.05	0.20	0.41	0.26	-0.14	-0.21	0.04	0.19	0.00
6	0.09	0.31	0.16	-0.04	-0.07	0.22	0.39	0.17	-0.14	-0.18	-0.04	0.23	0.04
7	0.15	0.20	0.20	0.01	-0.05	0.18	0.41	0.19	-0.14	-0.16	0.05	0.25	0.05
8	0.10	0.29	0.24	-0.03	-0.06	0.19	0.33	0.23	-0.07	-0.16	0.05	0.25	0.00
9	0.11	0.19	0.16	-0.02	-0.10	0.17	0.41	0.16	-0.10	-0.11	0.03	0.15	-0.09
10	0.12	0.28	0.08	0.03	0.04	0.19	0.43	0.26	0.08	0.18	0.02	0.25	0.08
11	0.09	0.17	0.19	-0.02	-0.10	0.15	0.33	0.14	-0.15	-0.24	-0.09	0.14	0.04
12	0.11	0.24	0.08	-0.04	-0.09	0.16	0.33	0.13	-0.18	-0.27	-0.01	0.19	0.09

column. The two sets of serial correlations (P+ and P-) are very symmetric or similar with respect to the values in each row and column, reflecting the alignment of the fertility gradients with respect to the axes of the field - 0 and 90 degrees (Figure 4, Table 6).

The basic pattern of the spatial correlation matrix was shown by the singular value decomposition on P+, on P- and on their average. The rank 1 least squares fit of the correlation matrices gave a "goodness of fit" of 84, 89 and 90 percent, respectively. The eigenvectors associated with the largest eigenvalues of P+ and P- were plotted against row number and showed the primary row and column effects very well (Figure 5). The column effect was a gradual and approximately linear decrease across rows while the row effect cycled across columns with an approximate cycle length of 6. The approximate coincidence of the P+ and P- plots of row and column effect show the symmetry of the gradient within the grove.

The StdMSE's reflect the relative efficiencies of 1x1 plots to plots of the stated size, or the ratio of two StdMSE's reflect the relative efficiencies of the two plot sizes and shapes within a block the size of the uniformity trial. The high correlations running the length of the grove are reflected in the high bx1 StdMSE's. They are much higher than their 1xa counterparts (compare where b = a) (Table 7). For example, 7x1 plots have a StdMSE of

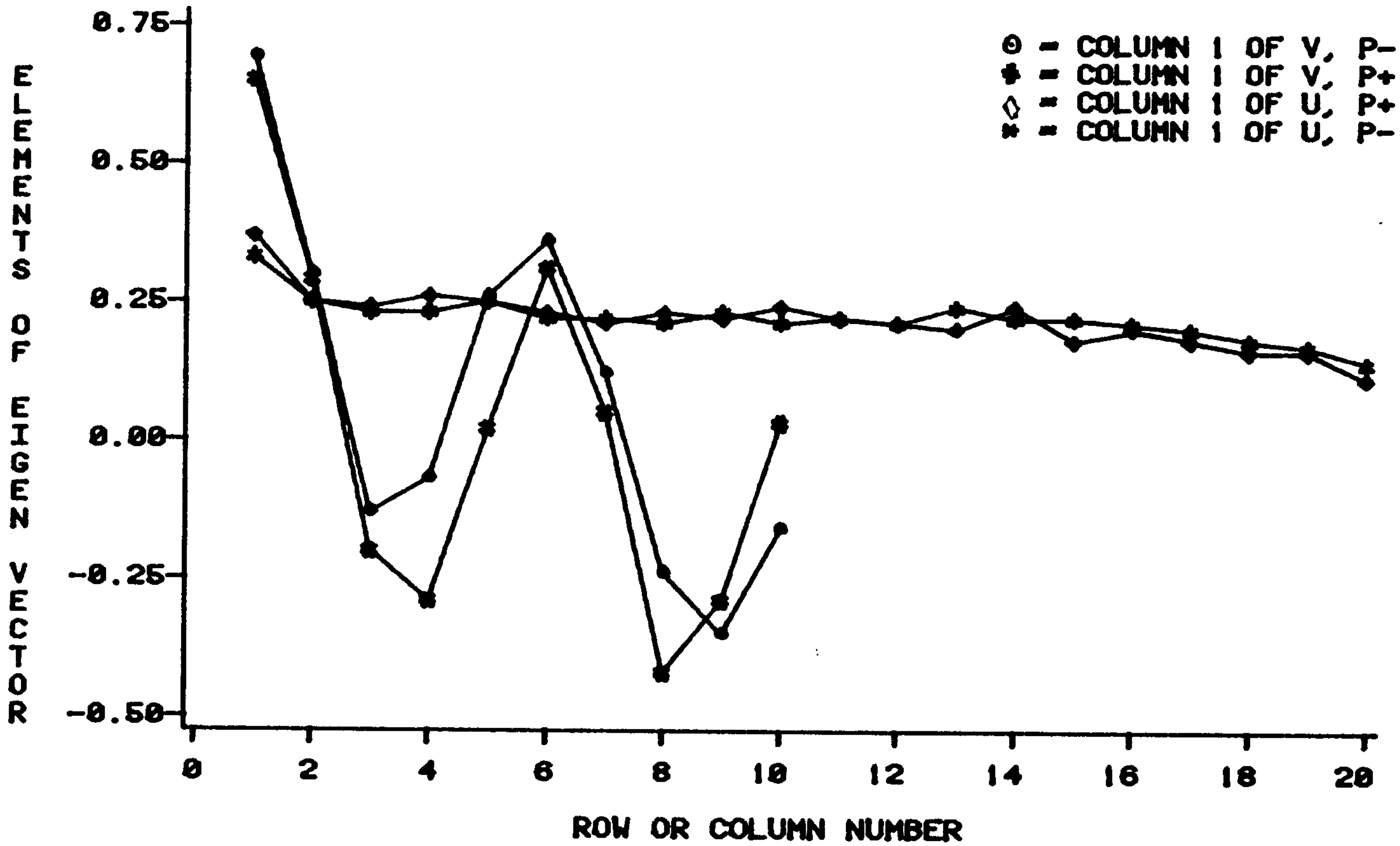


Figure 5. Elements of the eigenvectors corresponding to the largest root of the singular value decomposition of the spatial correlations of the lemon data

Table 7. Standardized mean square errors computed from the lemon data (Batchelor and Reed, 1918)

	Number of Unit Plots						
	1	2	3	4	5	6	7
1	1.00	1.27	1.33	1.34	1.41	1.50	1.56
2	1.51	2.02	2.14	2.13	2.22	2.41	2.55
3	2.02	2.75	2.90	2.87	2.99	3.26	3.47
4	2.52	3.46	3.65	3.59	3.75	4.12	4.42
5	3.02	4.18	4.39	4.30	4.49	4.97	5.35
6	3.49	4.88	5.10	4.98	5.19	5.76	6.21
7	3.94	5.53	5.71	5.61	5.83	6.49	7.01
8	4.38	6.18	6.43	6.22	6.47	7.20	7.80
9	4.79	6.79	7.05	6.81	7.07	7.90	8.58
10	5.20	7.40	7.67	7.38	7.67	8.58	9.33
11	5.81	8.01	8.29	7.97	8.27	9.26	10.08
12	6.81	8.60	8.88	8.52	8.83	9.90	10.78
13	6.40	9.16	9.45	9.04	9.35	10.48	11.43

3.94 and 1x7 plots of 1.56. Without adding a cost function or adjusting for a block less than the size of the grove or adjusting for the bias, approximately the same efficiencies are obtained using a 1x7 or a 2x1 plot (StdMSE's = 1.51). Thus plots in this grove should be rectangular and oriented parallel to rows and blocks should be parallel to columns.

For a given plot size of 2 ($p = 2$) and for blocks with 3 treatments ($t = 3$) the smallest MSE $(2|6)$ is obtained by rectangular blocks parallel to columns. Plots of size 2x1 in 6x1 blocks have an adjusted MSE of 0.52 and 1x2 plots in 3x2 blocks of 0.53. These plots and blocks are much more efficient than 1x2 plots in 1x6 blocks and 2x1 plots in 2x3 blocks (MSE$(2|6)$ is 1.16 and 1.19, respectively).

Regressing ℓn(StdMSE/p) on $\ell n(p)$, and therefore fitting Smith's model, gives very different regression coefficients depending on the orientation of the plots (Figure 6). For example, using 1x1, 1x2, 1x3, . . . plots gives a slope of -0.79 while the set of 1x1, 2x1, 3x1, . . . plots gives a slope of -0.26 (note that the slopes given are the negative of Smith's b). Thus, optimum plot size would vary considerably depending on plot orientation if determined from Smith's equation. The set of 1x7, 2x7, 3x7, . . . plots gives a slope of -0.21 and the set of 7x1, 7x2, 7x3, . . . of -0.76. Log variances within any one row or column of the MSE matrix are approximately linear with $\ell n(x)$ but overall they are not. The four lines

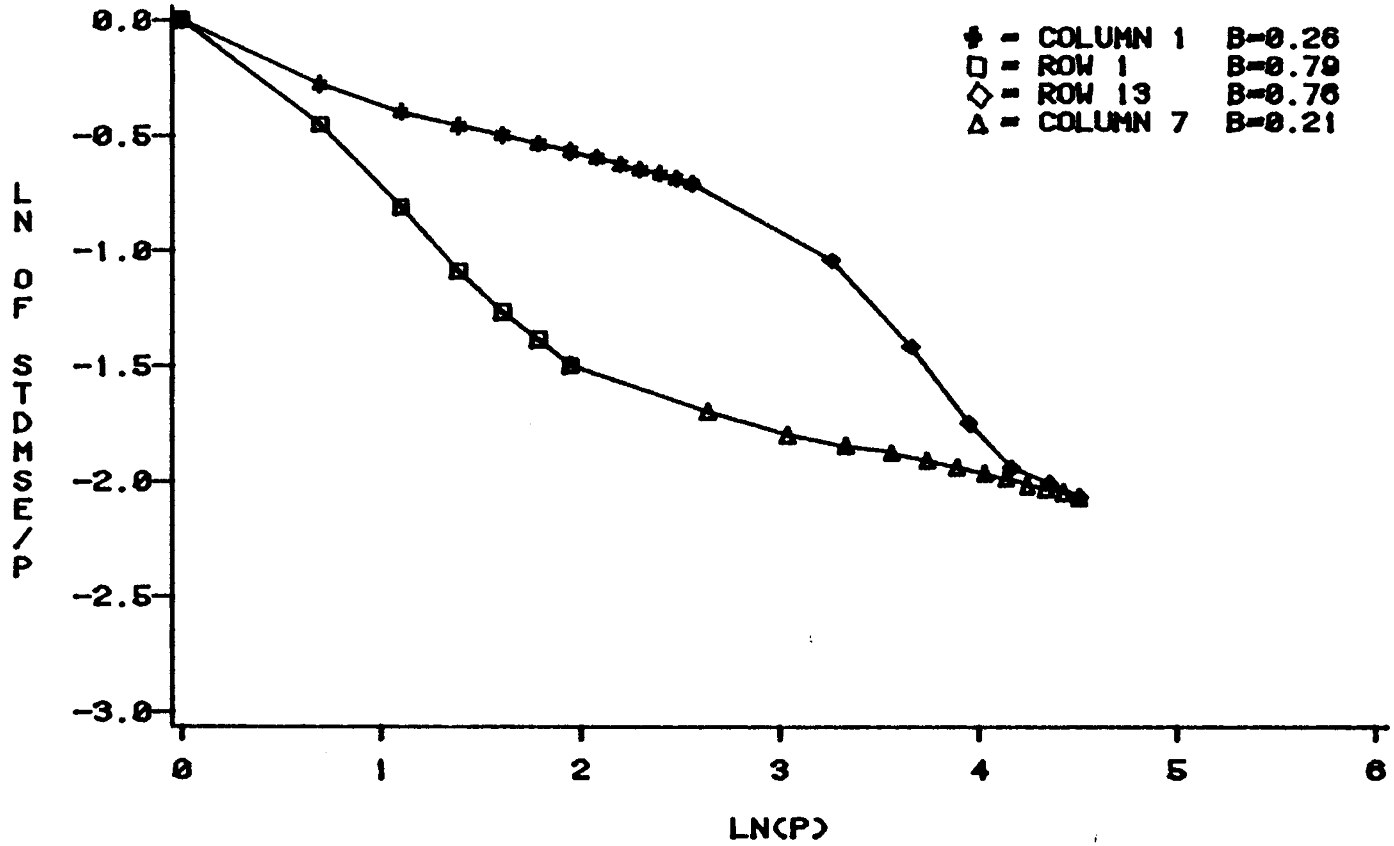


Figure 6. Plot of the ℓn(StdMSE/p) versus ℓn(p) using the first and last rows and columns of the StdMSE matrix from the lemon data showing Smith's B

above form a polygon within which the plot of ℓn(StdMSE/p) versus ℓn(p) will fall for all orientations and nestings of plots considered in this grove. Smith's equation does not differentiate between rectangular plots whose length runs with rows and those whose length runs with columns, or square plots of the same size, but gives a series of plot sizes running more or less through the middle of the polygon with an average b value. Therefore, some of the information available through the StdMSE's, as is seen by this lemon grove analysis, is obscured. In this situation, with Smith's analysis, Smith's equation consistently under-estimates optimum plot size by using an average b instead of the maximum b.

The 28x10 walnut grove had fertility patterns less distinct than those in the lemon grove. Yields ranged from 6 to 215 pounds per tree. A strip of very high fertility ran across the upper left end of the grove, an area of low yield dominated the middle region, and a mixture of high and intermediate yields covered the lower right end (Figure 7). Yield was not constant within any of the three primary regions and appears to change along both row and column. No primary row effect seems immediately evi-dent. The primary column effect appears to be a cycling, but because the pattern lacks regularity, no definite cycle length can be ascertained from Figure 7.

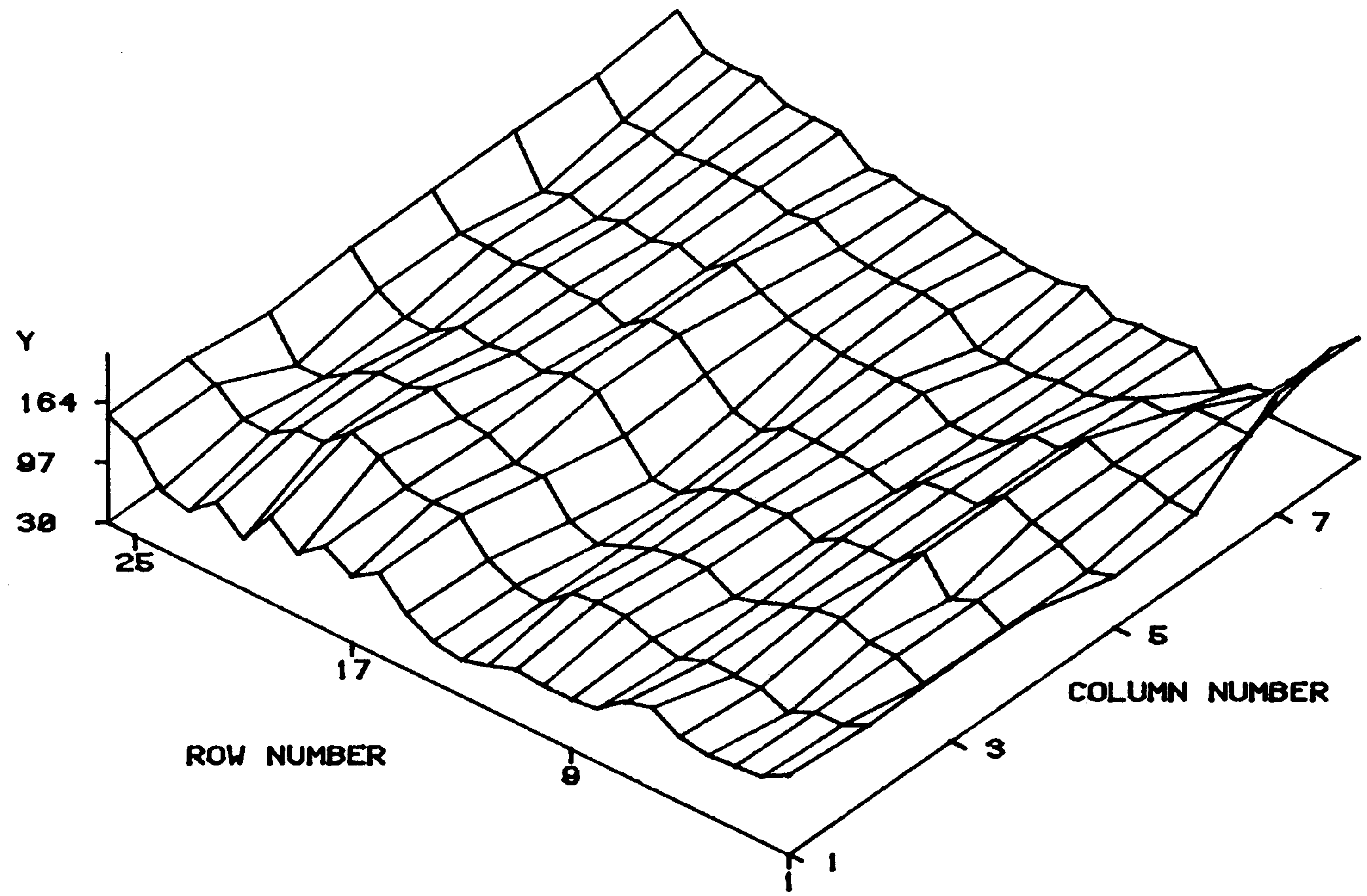


Figure 7. Smoothed response surface using moving medians of all possible 3x3 plots of pounds of walnuts per tree from the walnut grove (Batchelor and Reed, 1918)

Spatial correlations, fairly low at the start, tend to remain fairly constant within rows with a decrease at lag $(\ell,3)$ followed by a slight increase (Table 8). Within columns, correlations alternately decrease and increase, with independence reached at about lag $(5,k)$ for all k. No definite pattern can be detected in the correlation matrix but correlations less than or equal to lag (6,4) are positive. Correlations are negative from lag $(\ell,7)$ through lag $(\ell,13)$ for all ℓ.

The StdMSE's increase a little more rapidly along a row than along a column, reflecting the higher correlation among yields within a row (Table 9). For example, 1x5 plots have a StdMSE of 1.87 and 5x1 plots of 1.72, and therefore suggest the use of rectangular plots oriented parallel to columns. StdMSE's within a column tend to stabilize at about rows 9 to 11. The decrease through row 14 is a result of the negative correlations of the larger lags, due perhaps to some extent to the negative bias observed in the analysis of the strictly random surface. Taken at face value, the StdMSE's indicate that similar efficiencies would be obtained by 7xa and 14xa plots for all a = 1,2,3,4,5. For example, StdMSE's of 1.90 and 1.92 are obtained by 7x1 and 14x1 plots, respectively, and StdMSE's of 5.75 and 5.90 by 7x5 and 14x5 plots. Thus, in this case it would appear that plot size can be doubled from 7 to 14 without loss of information.

Table 8. Spatial correlations of yield in pounds per tree from the 28x10 walnut grove (Batchelor and Reed, 1918)

	k								
ℓ	-4	-3	-2	-1	0	1	2	3	4
0	0.27	0.17	0.24	0.23	1.00	0.23	0.24	0.17	0.27
1	0.11	0.07	0.11	0.25	0.24	0.24	0.19	0.17	0.25
2	0.20	0.14	0.15	0.20	0.18	0.15	0.18	0.17	0.18
3	0.12	0.07	0.11	0.17	0.16	0.00	0.04	0.08	0.15
4	0.08	0.07	0.08	0.16	0.10	0.09	0.14	0.20	0.05
5	0.17	0.02	0.09	0.01	0.04	0.03	0.00	0.09	0.10
6	0.11	0.11	0.06	0.05	0.00	0.13	-0.02	0.17	0.02
7	0.08	-0.10	0.01	-0.07	-0.05	-0.09	-0.03	0.02	-0.01
8	0.04	-0.05	-0.13	-0.09	0.03	-0.05	-0.04	0.08	-0.01
9	0.06	-0.09	-0.10	0.00	-0-05	-0.01	-0.03	-0.04	-0.03
10	-0.04	-0.13	-0.05	-0.03	-0.01	-0.01	-0.06	0.09	0.08
11	-0.19	-0.17	-0.21	-0.02	-0.10	-0.13	-0.02	-0.07	-0.03
12	-0.22	-0.30	-0.15	-0.25	-0.24	-0.18	-0.10	-0.08	-0.24
13	-0.28	-0.14	-0.24	-0.12	-0.21	-0.10	-0.19	-0.13	-0.27

Table 9. Standardized mean square errors computed from walnut data (Batchelor and Reed, 1918)

	Number of Unit Plots				
	1	2	3	4	5
1	1.00	1.23	1.45	1.65	1.87
2	1.23	1.69	2.09	2.43	2.81
3	1.41	2.07	2.63	3.11	3.64
4	1.58	2.36	3.04	3.63	4.28
5	1.72	2.62	3.41	4.11	4.87
6	1.82	2.82	3.68	4.47	5.34
7	1.89	2.98	3.92	4.80	5.75
8	1.94	3.07	4.06	4.99	6.01
9	1.98	3.14	4.14	5.11	6.18
10	2.00	3.17	4.18	5.18	6.28
11	2.02	3.20	4.21	5.21	6.34
12	2.02	3.19	4.19	5.18	6.32
13	1.98	3.12	4.08	5.04	6.16
14	1.92	3.02	3.92	4.83	5.90

Adjusting MSE's for plots of size 4 in blocks containing 2 treatments, as an example, gives the following MSE(4|8)'s:

Plot Shape	Block Shape	Mse(4|8)
1x4	1x8	1/
	2x4	0.87
2x2	2x4	0.96
	4x2	1.03
4x1	4x2	0.81
	8x1	1.23

One can see that in this situation the smallest adjusted MSE is obtained by placing rectangular 4x1 plots side by side to form 4x2 blocks.

The plot of ℓn(StdMSE/p) vs ℓn(p) gives a set of symmetric lines concave downward at the extreme end (Figure 8). The series of 1x1, 2x1, 3x1, . . . plots has a slope of -0.73 and the series of 1x5, 2x5, 3x5, . . . plots has a slope of -0.54. The slope increases slightly as each of the 5 columns of the StdMSE matrix is plotted, but the shape of the curve stays about the same. The series of 1x1, 1x2, 1x3 plots has a slope of -0.61 and the series of 14x1, 14x2, 14x3, . . . plots of -0.31. Any one set of nested rectangular plots doesn't deviate too badly from linearity if only small plots are considered. Points representing the larger plots tend to curve downward,

---

[1]Could not calculate.

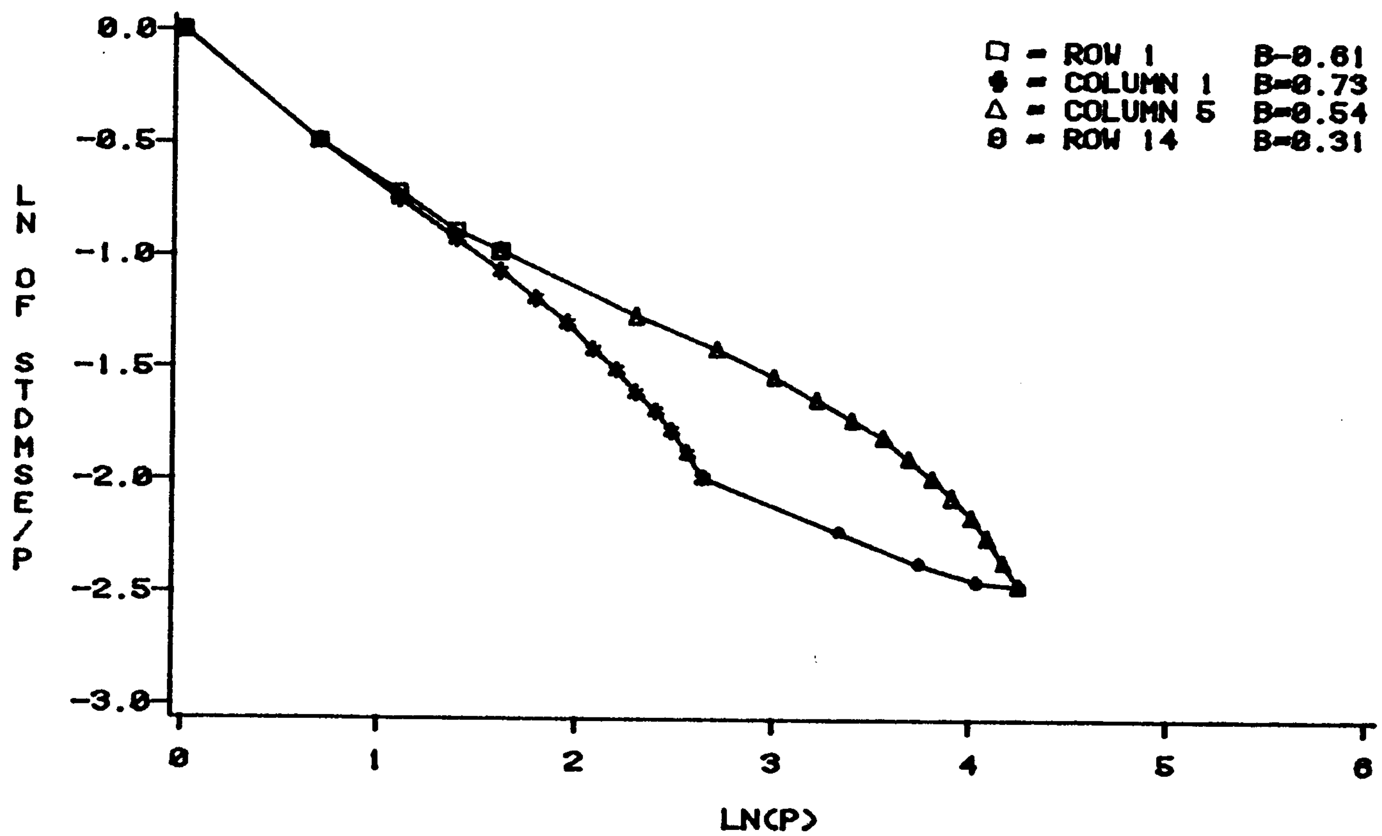


Figure 8. Plot of ℓn(StdMSE/p) versus ℓn(p) using the first and last rows and columns of the StdMSE matrix from the walnut data showing Smith's B

reflecting the stabilization of the StdMSE's at plot widths of 9 to 11 and the negative correlations of the larger lags.

The mangold field appears to have no dominant row or column fertility pattern but to be patchy instead. Yield, ranging from 267 to 384 pounds per plot, was highest in the middle of the right half of the field, in the upper right hand corner, and two-thirds of the way up tht left side of the field. An area of low yield ran across the upper left side of the field (Figure 9).

Lag (0,k) and lag (1,k) correlations decrease from 1 to 0.08 at $\ell = 4$ and increase to 0.37 at $\ell = 9$, giving a column cycle length of about 8 (Table 10). Lag (0,k) correlations decrease to k = 4. No regular pattern is seen in the rest of the correlation matrix. The P+ and P- correlations are not symmetric with respect to each other and suggest that the existing fertility patterns are not at 0 and 90 degrees to the sides of the field. This asymmetry is also revealed by the singular value decomposition done on P+ and P-. The eigenvectors of P+ and P- plotted against row number and column number give two very different lines in each case. The primary row and column effects are not the same for the (ℓ,k) and $(\ell,-k)$ lags (Figure 10).

StdMSE's increase at approximately the same rate along both row and column. Plots of size 1x5 have a StdMSE of 2.34 and 5x1 plots of 2.25; 2x3 plots of 2.71, and 3x2 plots of 2.71; 2x5 plots of 3.55, and 5x2 plots of 3.32 (Table 11).

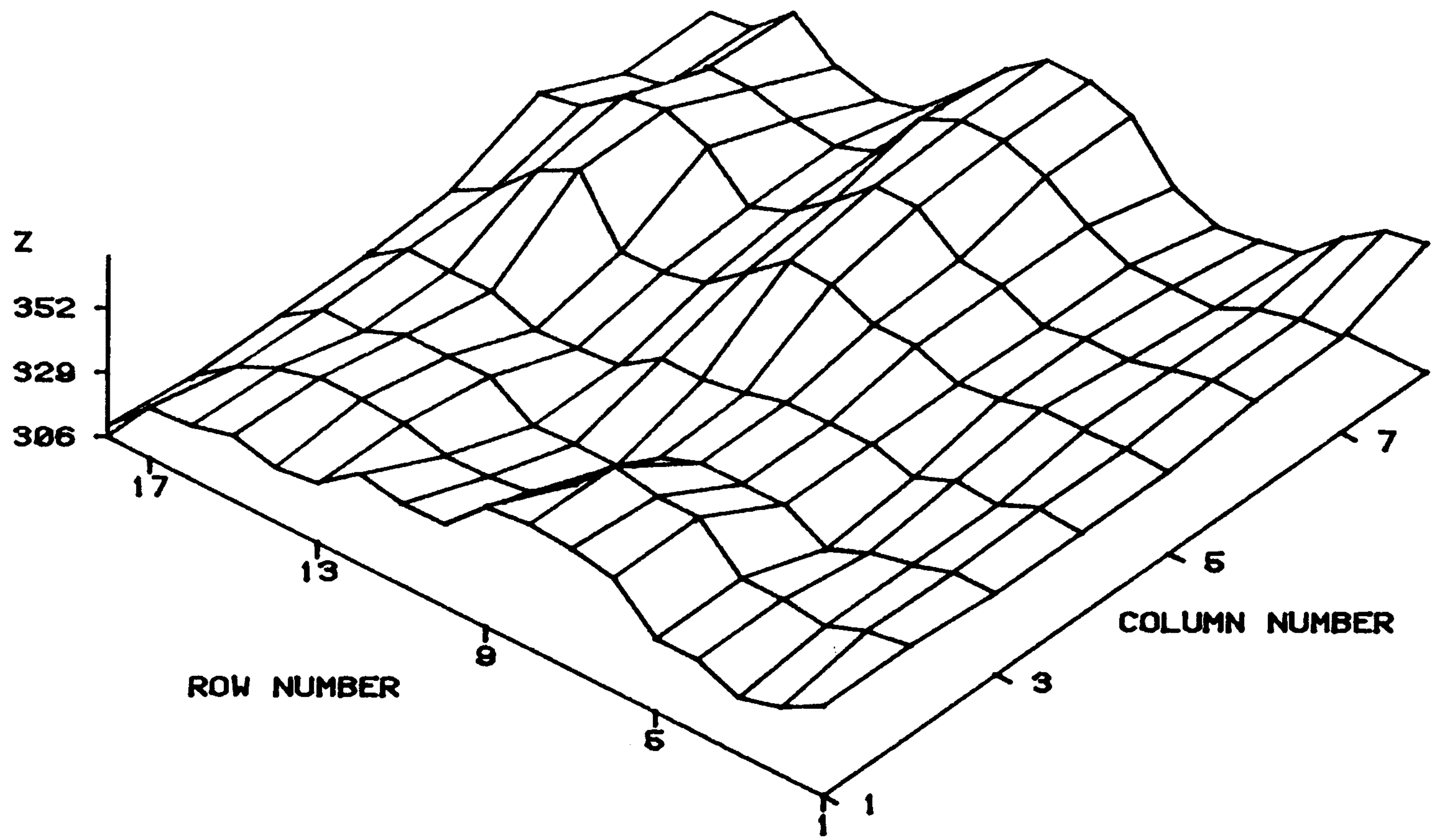


Figure 9. Smoothed response surface using moving medians of all possible 3x3 plots of pounds of mangold roots per plot for the 20x10 mangold field (Mercer and Hall, 1911)

Table 10. Spatial correlations computed from yield of mangold roots from the 20x10 mangold field (Mercer and Hall, 1911)

	k										
ℓ	-5	-4	-3	-2	-1	0	1	2	3	4	5
0	0.32	0.19	0.23	0.31	0.47	1.00	0.47	0.31	0.23	0.19	0.32
1	0.27	0.14	0.07	0.06	0.25	0.46	0.33	0.21	0.19	0.25	0.42
2	0.11	-0.04	0.00	-0.02	0.11	0.33	0.14	0.08	0.12	0.21	0.39
3	0.03	-0.04	0.03	-0.07	0.09	0.18	0.04	-0.04	0.06	0.10	0.29
4	0.19	0.10	0.01	-0.02	0.01	0.08	-0.09	-0.08	-0.15	0.11	0.26
5	0.14	0.15	0.23	0.13	0.18	0.24	0.06	0.01	-0.05	0.01	0.16
6	0.23	0.13	0.28	0.24	0.20	0.28	0.00	-0.12	-0.07	0.01	0.20
7	0.19	0.04	0.06	0.13	0.17	0.25	0.08	0.04	0.02	0.14	0.34
8	-0.08	-0.06	0.13	0.18	0.18	0.37	0.09	0.11	0.10	0.23	0.40
9	-0.28	-0.16	0.03	-0.02	-0.01	0.14	0.06	0.04	0.05	0.29	0.45

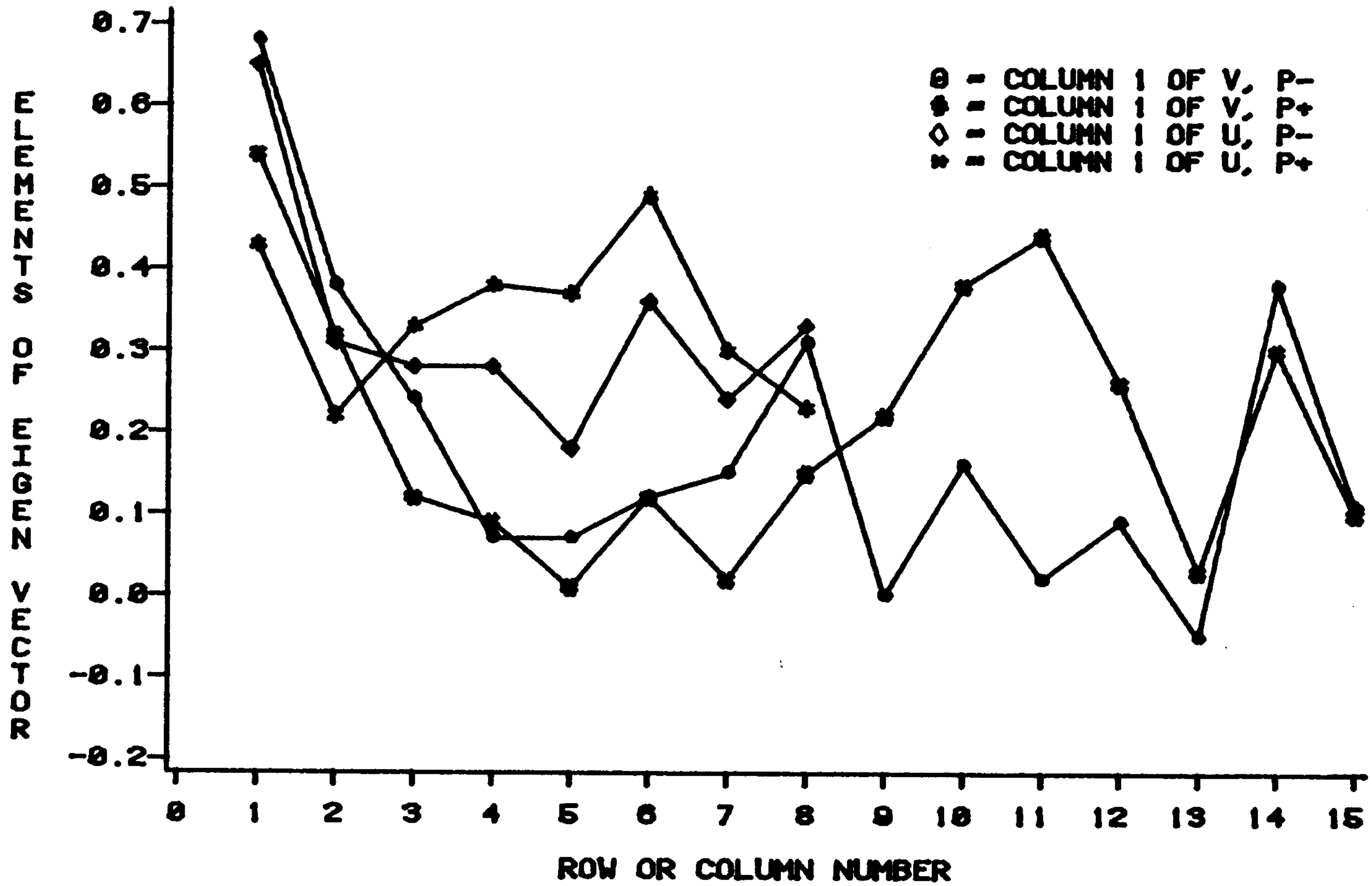


Figure 10. Elements of the eigenvectors corresponding to the largest root of the singular value decomposition of the spatial correlations of the mangold data

Table 11. Standardized mean square errors computed from the mangold data (Mercer and Hall, 1911)

	Number of Unit Plots				
	1	2	3	4	5
1	1.00	1.46	1.82	2.10	2.34
2	1.44	2.18	2.71	3.14	3.55
3	1.80	2.71	3.33	3.85	4.35
4	2.06	3.08	3.75	4.31	4.88
5	2.25	3.32	4.00	4.57	5.17
6	2.45	3.60	4.30	4.91	5.56
7	2.67	3.89	4.64	5.30	6.00
8	2.89	4.19	5.00	5.70	6.47
9	3.13	4.53	5.40	6.17	7.01
10	3.35	4.83	5.75	6.58	7.48

For a given plot of size 2 ($p = 2$) in blocks of 2 treatments ($t = 2$) the MSE(2|4)'s are:

Plot Shape	Block Shape	MSE(2|4)
1x2	1x4	0.83
	2x2	0.74
2x1	2x2	0.71
	4x1	0.83

Plots of size 2x1 in 2x2 blocks are slightly more efficient than 1x2 plots in 2x2 blocks. Plots of size 1x2 and 2x1 in 1x4 and 4x1 blocks give equal efficiencies.

Regression of ℓn(StdMSE/p) on ℓn(p) fits Smith's model very well. No matter what nesting or orientation of plots is used, the slopes are all close to -0.50 (Figure 11). Thus, it appears that a randomly patchy field can be analyzed with Smith's model with no loss of information or underestimation of b, but that a field with definite fertility patterns suffers under his equation.

The highest yields in the 1000-tree orange grove occurred in the upper left one-third of the grove (Figure 12). The rest of the grove alternated between low and intermediate yields, with two strips of each, running across columns. The lower one-third of this area contained a region of very low yield running down columns. The high-, low- and intermediate-yield areas seem to be strips running across columns on the average while the very low-yield area runs down columns.

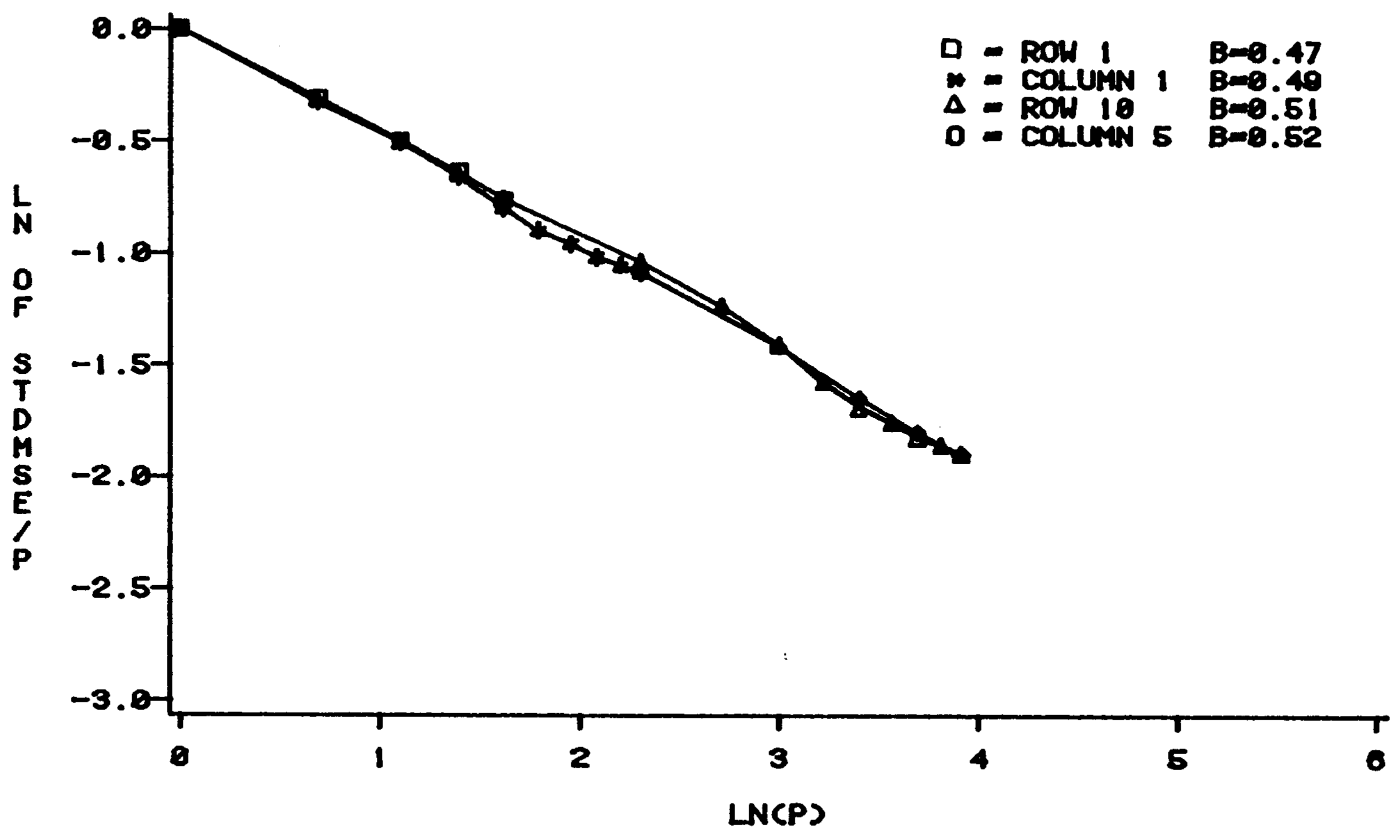


Figure 11. Plot of ℓn(StdMSE/p) versus ℓn(p) using the first and last runs and columns of the StdMSE matrix from the mangold root data showing Smith's b

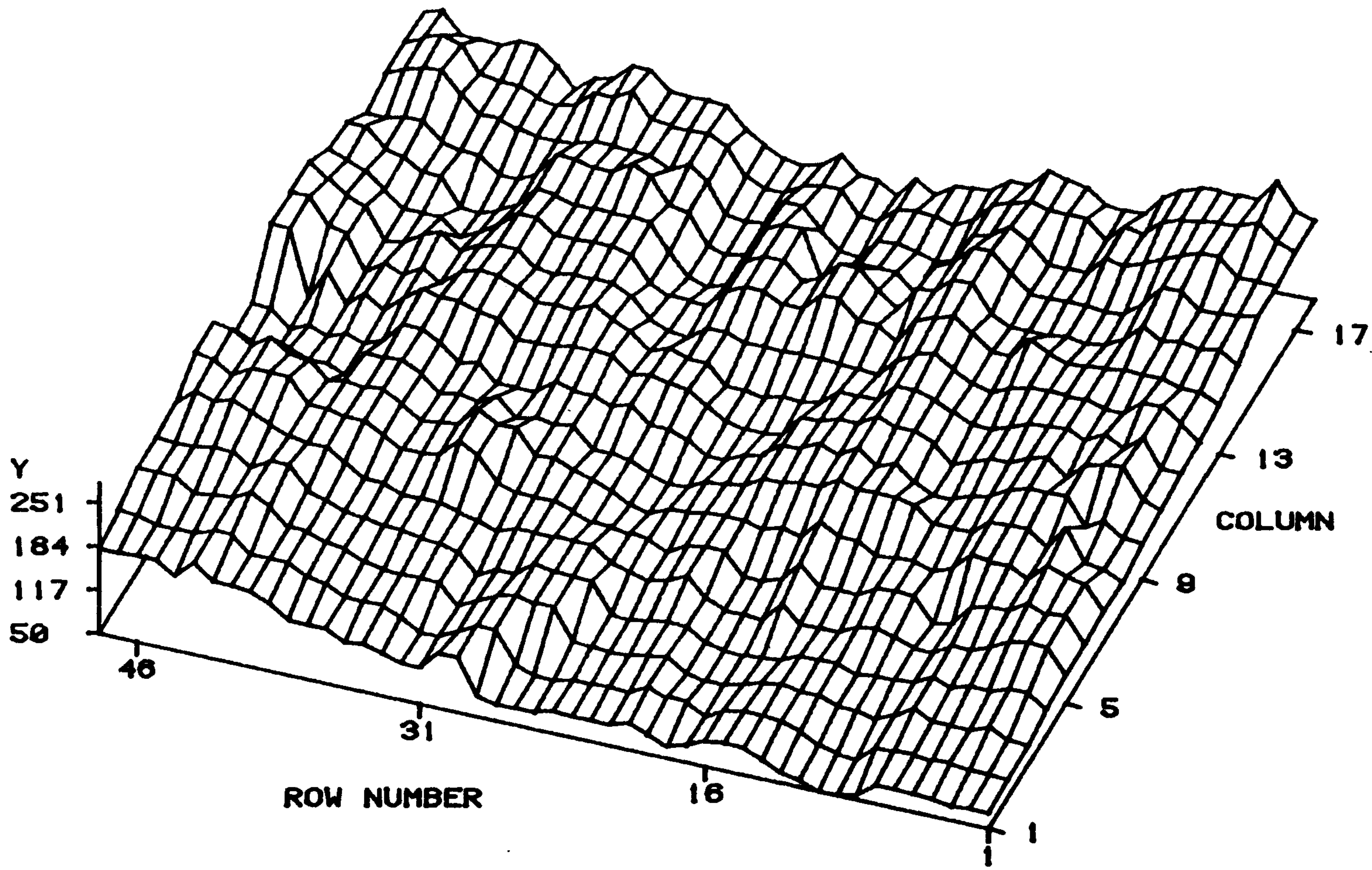


Figure 12. Smoothed response surface using moving medians of all possible 3x3 plots of pounds of oranges per tree from the 50x20 orange grove (Batchelor and Reed, 1918)

Lag $(\ell,0)$ correlations decrease very slowly and do not reach 0 until $\ell = 18$ (Table 12). Lag $(0,k)$ correlations also decrease slowly and at $k = 9$ are 0.26. Correlations tend to remain high within both row and column and are greater than 0 for lag (ℓ,k) where ℓ is less than or equal to 16 and k less than or equal to 9. Both P+ and P- appear to be similar to each other, again reflecting a 0 and 90 degree orientation of fertility gradients with respect to the axes of the grove.

StdMSE's increase at approximately the same rate along a row and column (Table 13). For example, the 1x10 plot has a StdMSE of 4.86 and a 10x1 plot of 4.91; a 2x10 plot of 8.69 and a 10x2 plot of 8.71; a 3x10 plot of 12.38 and a 10x3 plot of 12.21. Plots of size 4x5 and 5x4 with StdMSE's of 9.07 and 9.05, respectively, are less efficient than the 10x2 and 2x10 plots. Orientation of rectangular plots appears not to change the efficiency. Where square plots can be compared to rectangular plots of the same size, they are found to be less efficient. For example, 1x9 and 9x1 plots have StdMSE's of 4.51 and 4.56, respectively, whereas a 3x3 plot has a StdMSE of 4.72. Adjusting the MSE of plots of size 3 ($p = 3$) for blocks of 3 treatments ($t = 3$) gives the following MSE$(3|9)$'s:

Table 12. Spatial correlations of yield in pounds per tree from the 50x20 orange grove (Batchelor and Reed, 1918)

ℓ	k: -9	-8	-7	-6	-5	-4	-3	-2	-1	0
0	0.26	0.36	0.35	0.38	0.39	0.39	0.46	0.47	0.55	1.00
1	0.23	0.28	0.27	0.32	0.39	0.39	0.40	0.42	0.47	0.54
2	0.22	0.27	0.30	0.30	0.36	0.38	0.39	0.40	0.43	0.51
3	0.18	0.22	0.25	0.27	0.34	0.36	0.37	0.35	0.39	0.48
4	0.15	0.20	0.21	0.27	0.30	0.32	0.25	0.32	0.39	0.46
5	0.18	0.17	0.17	0.21	0.27	0.29	0.34	0.32	0.38	0.42
6	0.13	0.17	0.21	0.23	0.24	0.25	0.31	0.33	0.35	0.40
7	0.08	0.16	0.16	0.19	0.24	0.25	0.28	0.28	0.29	0.33
8	0.06	0.08	0.14	0.14	0.19	0.24	0.29	0.27	0.26	0.32
9	0.06	0.09	0.10	0.16	0.18	0.21	0.29	0.28	0.26	0.32
10	0.02	0.12	0.12	0.15	0.19	0.19	0.26	0.26	0.21	0.32
11	0.01	0.11	0.12	0.13	0.14	0.15	0.22	0.22	0.24	0.29
12	0.01	0.09	0.10	0.12	0.13	0.15	0.22	0.19	0.16	0.25
13	-0.02	0.04	0.09	0.08	0.09	0.14	0.20	0.19	0.15	0.19
14	-0.09	-0.06	0.02	0.05	0.09	0.13	0.16	0.12	0.05	0.15
15	-0.13	-0.09	-0.02	-0.01	0.02	0.09	0.11	0.06	0.04	0.11
16	-0.14	-0.12	-0.07	-0.03	0.00	0.04	0.08	0.07	0.03	0.09
17	-0.20	-0.15	-0.10	-0.10	-0.07	-0.02	0.02	0.00	-0.02	0.05
18	-0.25	-0.20	-0.13	-0.09	-0.06	-0.02	-0.03	-0.02	-0.04	0.00
19	-0.25	-0.23	-0.17	-0.19	-0.11	-0.03	-0.04	-0.07	-0.03	0.00
20	-0.26	-0.25	-0.23	-0.20	-0.09	-0.12	-0.07	-0.06	-0.10	-0.07
21	-0.27	-0.22	-0.15	-0.13	-0.09	-0.10	-0.02	-0.01	-0.07	-0.03
22	-0.21	-0.12	-0.09	-0.12	-0.07	-0.07	-0.04	0.03	-0.07	-0.05
23	-0.17	-0.13	-0.07	-0.08	-0.08	-0.13	-0.11	-0.02	-0.05	-0.03
24	-0.16	-0.11	-0.06	-0.15	-0.09	-0.16	-0.06	-0.02	-0.09	-0.11

continued

Table 12 (continued)

ℓ	k: 0	1	2	3	4	5	6	7	8	9
0	1.00	0.55	0.47	0.46	0.39	0.39	0.38	0.35	0.36	0.26
1	0.54	0.45	0.39	0.42	0.36	0.34	0.31	0.30	0.29	0.25
2	0.51	0.43	0.38	0.37	0.39	0.36	0.30	0.31	0.28	0.20
3	0.48	0.38	0.35	0.38	0.34	0.34	0.30	0.26	0.26	0.21
4	0.46	0.39	0.36	0.35	0.35	0.32	0.33	0.33	0.34	0.26
5	0.42	0.36	0.35	0.33	0.35	0.32	0.31	0.30	0.28	0.23
6	0.40	0.37	0.34	0.34	0.33	0.30	0.26	0.39	0.31	0.25
7	0.33	0.30	0.29	0.28	0.29	0.28	0.24	0.25	0.24	0.21
8	0.32	0.30	0.28	0.24	0.27	0.25	0.20	0.25	0.25	0.18
9	0.32	0.27	0.22	0.28	0.23	0.24	0.19	0.22	0.21	0.20
10	0.32	0.28	0.24	0.25	0.22	0.23	0.18	0.19	0.25	0.19
11	0.29	0.27	0.20	0.22	0.22	0.17	0.16	0.14	0.22	0.15
12	0.25	0.21	0.16	0.16	0.15	0.14	0.10	0.15	0.15	0.11
13	0.19	0.15	0.08	0.08	0.12	0.10	0.11	0.11	0.09	0.09
14	0.15	0.10	0.07	0.05	0.09	0.08	0.03	0.03	0.07	0.08
15	0.11	0.06	0.05	0.04	0.11	0.06	0.06	0.04	0.04	0.03
16	0.09	0.06	0.04	0.04	0.08	0.07	0.03	0.04	0.04	0.07
17	0.05	0.04	-0.03	-0.04	-0.01	-0.01	-0.05	0.02	0.07	0.05
18	0.00	-0.03	-0.01	-0.04	-0.05	-0.01	-0.02	-0.02	0.05	0.02
19	0.00	-0.02	-0.04	0.01	0.00	-0.01	-0.04	0.05	0.07	0.06
20	-0.07	-0.08	-0.05	-0.03	-0.02	-0.05	-0.05	-0.03	0.06	0.02
21	-0.03	-0.08	-0.08	-0.06	-0.08	-0.09	-0.13	-0.08	0.01	-0.03
22	-0.05	-0.09	-0.08	-0.06	-0.11	-0.07	-0.12	-0.09	0.00	-0.02
23	-0.03	-0.06	-0.10	-0.11	-0.08	-0.16	-0.13	-0.11	-0.01	0.01
24	-0.11	-0.14	-0.15	-0.11	-0.14	-0.13	-0.15	-0.14	-0.17	-0.02

Table 13. Standardized mean square errors calculated from the orange data of Batchelor and Reed (1918)

	Number of Unit Plots									
	1	2	3	4	5	6	7	8	9	10
1	1.00	1.54	2.02	2.49	2.92	3.33	3.74	4.13	4.51	4.86
2	1.53	2.52	3.40	4.26	5.08	5.86	6.62	7.34	8.04	8.69
3	2.04	3.44	4.72	5.96	7.14	8.28	9.38	10.42	11.44	12.38
4	2.52	4.31	5.94	7.54	9.07	10.55	11.97	13.32	14.62	15.84
5	2.97	5.14	7.12	9.05	10.91	12.72	14.44	16.10	17.68	19.15
6	3.41	5.94	8.24	10.50	12.68	14.80	16.82	18.75	20.60	22.33
7	3.82	6.10	9.32	11.90	14.39	16.80	19.10	21.30	23.41	25.38
8	4.20	7.40	10.34	13.21	15.99	18.69	21.26	23.71	26.05	28.25
9	4.56	8.07	11.30	14.46	17.51	20.48	23.30	25.98	28.56	30.97
10	4.91	8.71	12.21	15.64	18.96	22.18	25.24	28.16	30.95	33.56
11	5.24	9.32	13.09	16.78	20.36	23.82	27.10	30.23	33.23	36.03
12	5.56	9.91	13.93	17.87	21.69	25.38	28.88	32.21	35.40	38.38
13	5.86	10.47	14.71	18.90	22.94	26.84	30.54	34.07	37.44	40.59
14	6.15	10.98	15.45	19.85	24.10	28.20	32.09	35.79	39.33	42.63
15	6.41	11.56	16.13	20.73	25.17	29.45	33.50	37.36	41.06	44.49
16	6.65	11.89	16.74	21.52	26.14	30.59	34.80	38.80	42.63	46.18
17	6.87	12.28	17.30	22.25	27.02	31.63	35.97	40.10	44.05	47.72
18	7.08	12.64	17.80	22.89	27.81	32.55	37.02	41.26	45.31	49.06
19	7.26	12.96	18.24	23.47	28.51	33.36	37.94	42.28	46.41	50.24
20	7.42	13.24	18.64	23.97	29.13	34.09	38.75	43.17	47.38	51.26
21	7.56	13.48	18.98	24.41	29.66	34.70	39.44	43.92	48.19	52.13
22	7.69	13.69	19.27	24.79	30.13	35.24	40.04	44.57	48.89	52.86
23	7.80	13.87	19.53	25.12	30.53	35.71	40.55	45.13	49.48	53.48
24	7.90	14.04	19.75	25.41	30.87	36.10	40.99	45.60	49.97	53.98
25	7.99	14.17	19.94	25.64	31.15	36.42	41.34	45.97	50.36	54.38

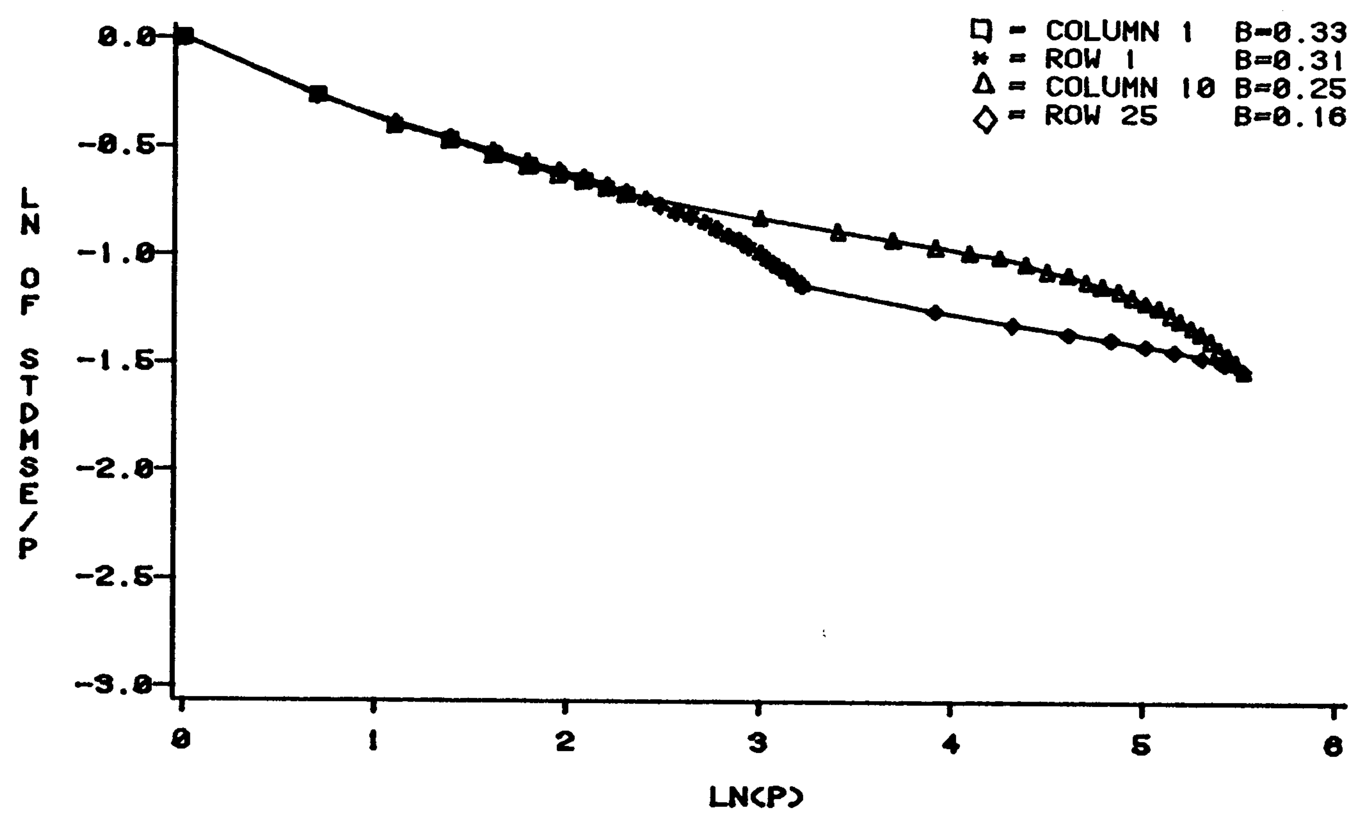


Figure 13. Plot of $\ell n(\mathrm{StdMSE}/p)$ versus $\ell n(p)$ using the first and last rows and columns of the StdMSE matrix from the orange grove data showing Smith's b

Plot Shape	Block Shape	MSE(3|9)
1x3	1x9	0.78
	3x3	0.68
3x1	3x3	0.70
	9x1	0.77

The greatest efficiency is obtained from 1x3 or 3x1 plots in blocks of size 3x3, and next from 1x3 or 3x1 plots in rectangular 1x9 or 9x1 blocks, with blocks of the same shape having approximately equal efficiencies. Orientation of rectangular plots and blocks had little effect on efficiencies.

Smith's b ranged from -0.16 to -0.33 in this example, the plot of $\ln(\text{StdMSE}/p)$ vs $\ln(p)$ being similar to the same plot for the walnut data (Figure 13). Each column of the StdMSE matrix yielded a straight line with a slight downward curve at the extreme end because of the negative correlations of lags greater than (6,9), with the concavity increasing as p increased. The 1x1, 2x1, 3x1, . . . series has a slope of -0.33 and 1x10, 2x10, 3x10, . . . series of -0.25.

To investigate repeatability of the results of a spatial analysis the upper left quarter of the orange grove was analyzed by itself. It appears to be a slight bowl with respect to fertility (Figure 11). Yield decreased and then increased along a column and along a row. Yield on the average was high, with the lowest yield being slightly below the center of the quarter. There was also an area of lower yield across the top of the quarter (rows 25 to 28, columns 10 to 20).

The correlations also decrease to about lag (7,k) and then increase slightly for all k (Table 14). Lag (ℓ,k) correlations for ℓ = 1,2,3,4 and k = 1,2,3,4,5 remain high and decrease slowly. Independence is never reached in either direction. P+ and P- are fairly similar for all ℓ and suggest that the gradients are at 0 and 90 degrees with respect to the axes of the grove.

The StdMSE's increase more rapidly along rows than along columns (Table 15). For example, a 1x5 plot has a StdMSE of 2.75 and a 5x1 plot of 2.42. Rectangular plots should therefore be oriented parallel to columns. Square plots seem to be intermediate to the two rectangular plots, i.e., a 1x4 plot has a StdMSE of 2.35, a 2x2 plot of 2.23, and a 4x1 plot of 2.13. Adjusting plots of size 2 for blocks of 2 treatments gives the following MSE(2|4)'s:

Plot Shape	Block Shape	MSE(2|4)
1x2	1x4	0.60
	2x2	0.72
2x1	2x2	0.66
	4x1	0.75

In this situation the 1x2 plots in 1x4 blocks give the highest efficiency, 2x1 plots in 2x2 blocks are next, 1x2 plots in 2x2 blocks are next, and 2x1 plots in 4x1 blocks are last.

Table 14. Spatial correlations of yield in pounds per tree from the southeast quarter of the 50x20 orange grove (Batchelor and Reed, 1918)

	k								
ℓ	-4	-3	-2	-1	0	1	2	3	4
0	0.34	0.51	0.43	0.49	1.00	0.49	0.43	0.51	0.34
1	0.33	0.33	0.31	0.35	0.46	0.31	0.34	0.27	0.21
2	0.36	0.28	0.33	0.27	0.31	0.31	0.27	0.28	0.28
3	0.31	0.24	0.27	0.20	0.37	0.22	0.20	0.32	0.14
4	0.22	0.16	0.12	0.12	0.23	0.14	0.16	0.11	0.10
5	0.13	0.09	0.12	0.10	0.17	0.15	0.15	0.10	0.18
6	-0.01	0.08	0.12	0.04	0.17	0.14	0.05	0.10	0.04
7	0.04	0.14	0.04	0.12	0.22	0.07	0.09	0.03	-0.03
8	-0.06	0.12	0.05	0.01	0.12	0.10	0.06	0.05	0.10
9	0.06	0.26	0.16	0.13	0.18	0.13	0.09	0.14	0.16
10	0.06	0.17	0.10	0.07	0.30	0.15	0.11	0.15	0.02
11	0.18	0.31	0.23	0.16	0.37	0.38	0.23	0.31	0.31

Table 15. Standardized mean square errors computed from the southeast quarter of the orange grove (Batchelor and Reed, 1918)

	Number of Unit Plots				
	1	2	3	4	5
1	1.00	1.47	1.90	2.35	2.75
2	1.44	2.23	2.95	3.69	4.36
3	1.79	2.85	3.85	4.85	5.80
4	2.13	3.43	4.66	5.91	7.10
5	2.41	3.89	5.32	6.74	8.14
6	2.65	4.28	5.86	7.46	9.00
7	2.86	4.62	6.33	8.06	9.73
8	3.06	4.94	6.76	8.60	10.38
9	3.23	5.21	7.13	9.07	10.94
10	3.40	5.48	7.50	9.55	11.51
11	3.59	5.76	7.88	10.03	12.09
12	3.79	6.09	8.31	10.59	12.76

The regression of $\ell n(\text{StdMSE}/p)$ on $\ell n(p)$ gives b's ranging from -0.46 to -0.39 for the 1x1, 2x1, 3x1, . . . and 1x5, 2x5, 3x5, . . . series; and from -0.37 to -0.25 for the 1x1, 1x2, 1x3, . . . and 12x1, 12x2, 12x3, . . . series (Figure 14). The b's for the entire orange grove ranged from -0.16 to -0.33. This discrepancy between the b's of the whole grove and a part of the grove reinforces what researchers have been saying since the early 1900's, that optimum plot size cannot be generalized for a crop and is also field specific. Any series of nested plots would give a fairly linear plot and would not show too much deviation from Smith's model.

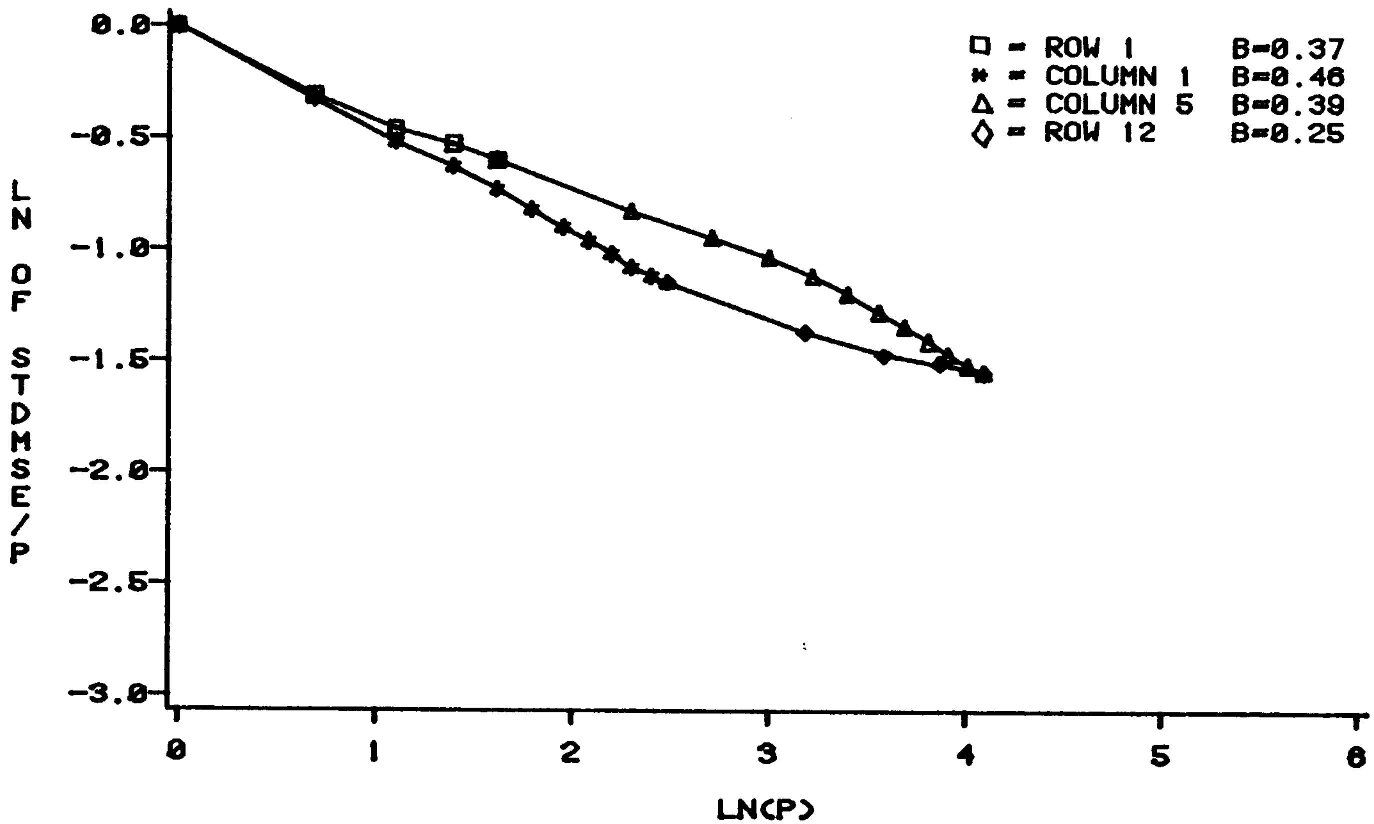


Figure 14. Plot of ℓn(StdMSE/p) versus ℓn(p) using the first and last rows and columns of the StdMSE matrix of the southeast quarter of the 50x20 orange grove showing Smith's b

DISCUSSION

The results indicate that the generality and greater flexibility of the correlation model for the analysis of uniformity data give more information than does Smith's model. From the spatial correlations one can see the nature and orientation of the fertility gradients within the field. One can also determine if the gradients are symmetric within the field. From the StdMSE's one can quickly determine if Smith's one-dimensional model holds. If the rate of change along rows is the same as that along columns, $b_1 = b_2 = b$, and b can be calculated easily from the plot of any row or column. One can also see if Smith's two-dimensional model holds. The first and last rows and columns of the StdMSE matrix will change linearly and give estimates of b that form a polygon within which all other estimates of b will fall. The largest b should be used in computing optimum plot size. If neither of Smith's models holds, the StdMSE's can be used directly to determine optimum plot size and shape and optimum blocking. No matter what model holds, the information is available for adjusting the MSE's of a given plot size p for blocks of a smaller size than the field. All possible sizes and shapes can be compared to find the combination giving the highest efficiency.

The primary disadvantages with this method are the greater computational effort initially and its negative bias. Once the computer program is set up, very little time or cost is involved in analyzing a uniformity trial.

Several different computational methods were tried in efforts to eliminate or decrease the bias introduced by adjusting for the mean. None of them were less biased than computing the covariances directly and combining them to form the variances of the different-sized plots.

Whittle (1954) reported autocorrelations for the orange grove for lags (ℓ,k) where $\ell = 0,1,2,\ldots 9$ and $k = 0,1,2,3,4$. His values are identical to those calculated using the product moment correlation coefficient that was used in the spatial correlation approach (Tables 12 and 16).

Christidis said that plot shape was theoretically immaterial in patchy fields as long as all plots were the same size. The mangold root data appeared to be such a field and the StdMSE's confirm his conclusion. Orientation of rectangular plots changed efficiency very little and square plots were only slightly less efficient than rectangular plots.

Smith's one-dimensional model held for the mangold root data. The spatial correlation approach gave a b of approximately 0.50. Smith reported a value of 0.39 in 1938. There was only one listing for mangolds so it is not known whether he analyzed the root or leaf data. The lemons followed Smith's two-dimensional model. Results of the spatial correlation analysis gave b's ranging from

Table 16. Autocorrelations for the orange data as reported by Whittle (1954)

	k									
ℓ	0	1	2	3	4	5	6	7	8	9
-4	0.39	0.39	0.37	0.36	0.37	0.29	0.25	0.25	0.24	0.21
-3	0.46	0.40	0.39	0.37	0.35	0.34	0.31	0.28	0.29	0.29
-2	0.47	0.42	0.40	0.35	0.32	0.32	0.32	0.28	0.27	0.28
-1	0.55	0.47	0.43	0.39	0.49	0.38	0.35	0.29	0.26	0.26
0	1.00	0.54	0.50	0.48	0.46	0.42	0.40	0.32	0.32	0.32
1	0.55	0.44	0.43	0.38	0.39	0.36	0.37	0.29	0.30	0.29
2	0.47	0.38	0.38	0.35	0.36	0.35	0.34	0.29	0.28	0.22
3	0.46	0.42	0.37	0.38	0.35	0.33	0.34	0.28	0.24	0.27
4	0.39	0.35	0.39	0.34	0.35	0.34	0.33	0.29	0.27	0.23

0.21 to 0.79 while Smith reported a value of 0.34, a clear underestimation of optimum plot size in a situation where the maximum b should be used. The remaining three data sets did not follow Smith's model, i.e., the plots of $\ell n(StdMSE/p)$ versus $\ell n(p)$ were not linear. The spatial correlation approach gave b's from 0.31 to 0.73 for walnuts and from 0.16 to 0.33 for the orange grove. Smith reported values of 0.47 to 0.10, respectively. In this situation Smith's model was too inflexible to allow for the kind of spatial correlation patterns observed in the walnut and orange groves.

Thus, it appears that the spatial correlation approach is a useful addition to the methods of determining optimum plot size. Because of its generality in imposing no restrictions on the behavior of the correlations, more information is retained. And, in those cases where Smith's model does hold, the results readily simplify to Smith's model. But before the spatial correlation approach can be used effectively, the bias should be eliminated from spatial correlations computed from experimental data and not only from null data sets.

LIST OF REFERENCES

Batchelor, L. D. and H. S. Reed. 1918. Relation of the variability of yields of fruit trees to the accuracy of field trials. Journal of Agricultural Research 7(5):245-283.

Christidis, B. G. 1931. The importance of the shape of plots in field experimentation. Journal of Agricultural Science 21:14-37.

Christidis, B. G. 1939. Variability of plots of various shapes as affected by plot orientation. Empire Journal of Experimental Agriculture 7:330-342.

Harris, J. W. 1915. On a criterion of a substratum homogeneity (or heterogeneity) in field experiments. American Naturalist 49:430-454.

Hatheway, W. H. and E. J. Williams. 1958. Efficient estimation of the relationship between plot size and the variability of crop yields. Biometrics 14(2): 207-222.

Koch, E. J. and J. A. Rigney. 1951. A method of estimating optimum plot size from experimental data. Agronomy Journal 43:17-21.

Li, J. C. R. and K. R. Keller. 1951. An application of serial correlation in field experiments. Agronomy Journal 43:201-204.

Mercer, W. B. and A. D. Hall. 1911. The experimental error of field trials. Journal of Agricultural Science 4:107-132.

Pearce, S. C. 1955. Some considerations in deciding plot size in field trials with trees and bushes. India Society of Agricultural Statistics 7:23-26.

Pearce, S. C. 1976. An examination of Fairfield Smith's law of environmental variation. Journal of Agricultural Science 87:21-24.

Smith, H. F. 1938. An empirical law describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science, Cambridge, 28:1-23.

Whittle, P. 1954. On stationary processes in the plane. Biometrika 41:434-449.