LONG-TERM EXPERIMENTS AND THE KONZA PRAIRIE

by

PHILIP K. CLARK

B.S., Kansas State University, 1971

959-5358

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

Approved by:

Arthur D Dayton
Major Professor

LD
2668
R4
1972
C54
c.2
Documents

TABLE OF CONTENTS

LISTING OF TABLES

INTRODUCTION

Agricultural experiments dealing with a factor or groups of factors are often carried out at a number of places and over a number of years. One reason for this is the effects of most factors (fertilizers, varieties, etc.) may differ from place to place and from year to year due to differences in soil, agronomic practices, climatic conditions, and other environmental conditions. The results from one year or one place may not provide enough information for determining the most desirable variety, level of fertility, etc. Perhaps a reason for the few number of papers and discussions in previous years on long-term experiments is due to agricultural experimental stations not wanting to tie-up large areas of land for the many years necessary for such types of experimentation.

The design of long-term experiments is very important. With annual experiments, the defects of a faulty design, if recognized, can be corrected in any repetitions of following years. However, with long-term experiments, the results of several years may be largely vitiated by a poor design at the outset.

There are different types of long-term experiments in regard to the kind of information they yield. These are given by Cochran (1939).

Types of Long-Term Experiments

	TREATMENTS		INFORMATION
FIXED	Applied on the same plots	.. Every year	.. Cumulative effects
		.. 1st year only	.. Residual effects
		.. At fixed intervals	.. Direct and Residual
ROTATING	Applied on different plots in successive years		.. Direct and Residual

X

CROP

Single crop {annual / perennial}

Fixed rotation

Effects of different crops

KONZA PRAIRIE PROJECT

A long-term experiment is to be conducted on the Konza Prairie Research Natural Area. The Prairie, located south of Manhattan along the north side of Interstate Highway 70, is rectangular in shape and about three miles by one-half mile in size and was deeded in 1972 to the Kansas State University Endowment Association from the Nature Conservancy. As far as is known, the land has only been used for cattle grazing, which will be discontinued while research is being conducted. Herbicides have not been used on the area except for spot spraying for the control of musk thistle.

The Konza Prairie is to be kept in a natural state for scientific study of the prairie ecosystem. It is hoped the prairie will revert to the conditions prior to the first settlement of people in the area. Of the factors influencing the ecosystem, climate and available organisms are considered uniform over the area, and substrate and topography vary but are fixed. The area has been burned about three out of four years during the last thirty years. However, it is not known how often the prairie might have been burned in previous years. Therefore, in order to study the effects of burning the following burning treatments were chosen so as to include the extremes with respect to how often the area was burned;

B0 - unburned

B1 - burned annually

B2 - burned every other year

B4 - burned every fourth year

B10 - burned every tenth year

BW - burned each season the rainfall the year before was 1.2 times the median point (median = 30.65 inches).

The burnings are to occur in late April. The area is divided into twenty-four plots, or four replications for each of the six treatments. The plot boundaries are often along ridges in order to provide one treatment per watershed. The average size of the plots is about forty-three acres. To help control the fire, a strip at least thirty meters wide will be mowed in late July around the periphery of the areas. The mowed strips will not stop fires, but there, the fire will be low and easily controlled. A map of the prairie appears in Figure 1.

Various types of research have been and are being planned for the Konza Prairie. Some of the research projects include experiments on the insect population, the collard lizard population, vegetation composition in relation to soil and burning treatments, runoff water quality, soil samples, and bird populations. The samples (or observations) will basically be of two types. For the experiments dealing with runoff water quality and the bird and lizard populations, the whole plot will be used to obtain the observations. For the research dealing with the insect populations, soil samples, and vegetation composition, samples may be obtained so that as many soil types as possible are taken into account. A definite problem exists for measuring runoff water quality. Due to the variable amounts of different soils within a plot, it will be difficult to determine whether runoff water quality is due to the burning treatment or the different soils or both.

It was decided by those conducting research that a uniformity trial would not be used. The value of such a trial was pointed out, as it might have been helpful had one been carried out because the observations taken in the trial years could have been used as a covariate to adjust later year's observations.

An administrative board for the project has been established with the following duties:

1. Control all uses of Konza Prairie.
2. Approve all research proposals.
3. Be responsible for maintenance, services, protection, improvements, and publicity.
4. Maintain records of use to others and encourage integration of separate duties.

ILLEGIBLE DOCUMENT

THE FOLLOWING DOCUMENT(S) IS OF POOR LEGIBILITY IN THE ORIGINAL

THIS IS THE BEST COPY AVAILABLE

P-474

THE FOLLOWING DOCUMENT(S) IS OVERSIZED AND IS BEING FILMED IN SECTIONS TO INSURE COMPLETENESS AND CONTINUITY

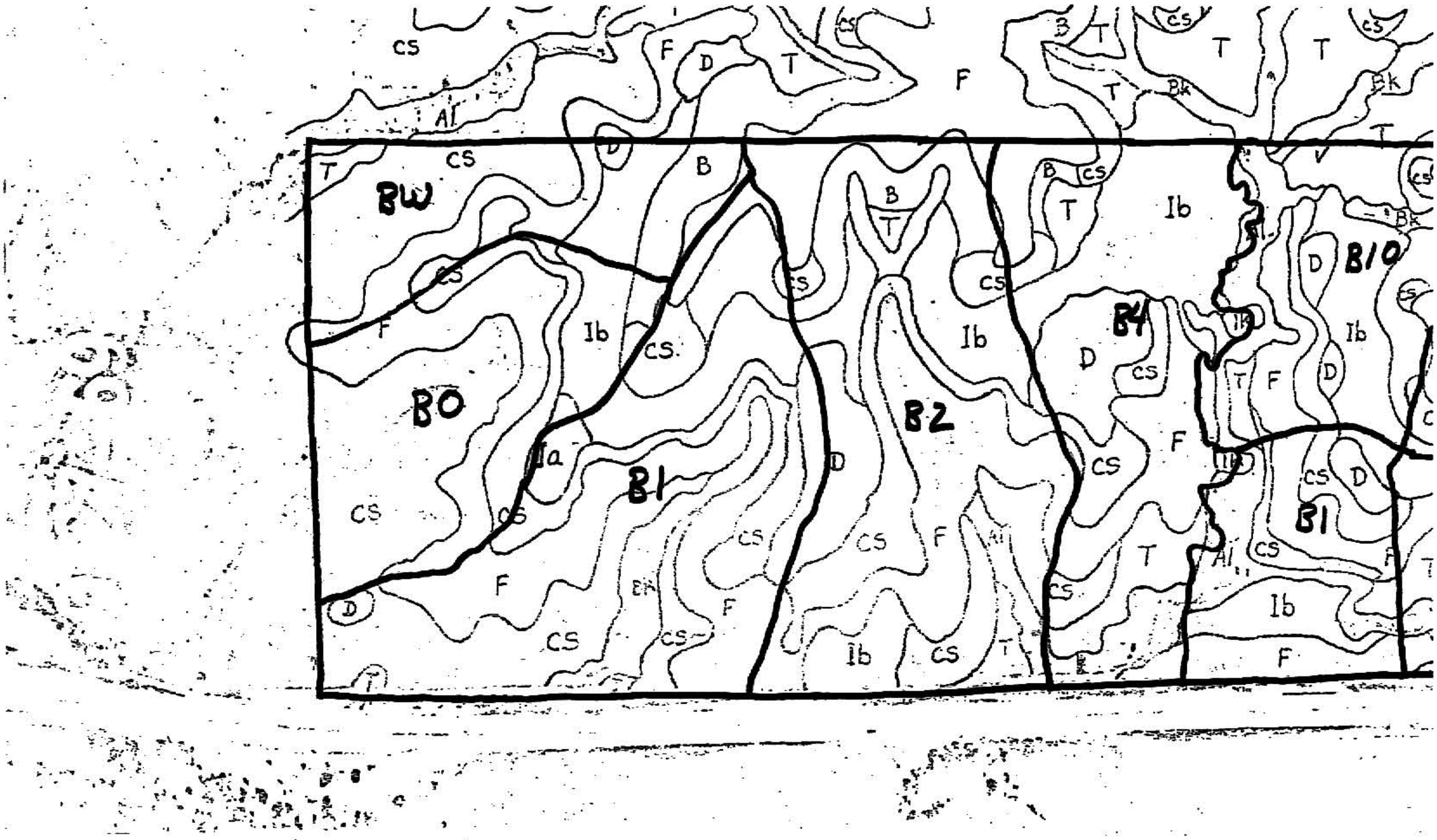


Figure 1.
Map of the Konza Prairie

Map Legend of Soils

Al Alluvial land

B Benfield silty-clay loam, 5 - 20% slopes

Fc Florence cherty silt loam or Florence cherty silty clay loam

Bk Breaks-alluvial land complex

CS Clime-Sogn complex, 5 - 20% slopes

D Dwight silt loam

Ia Irwin silty clay loam, 1 - 4% slopes

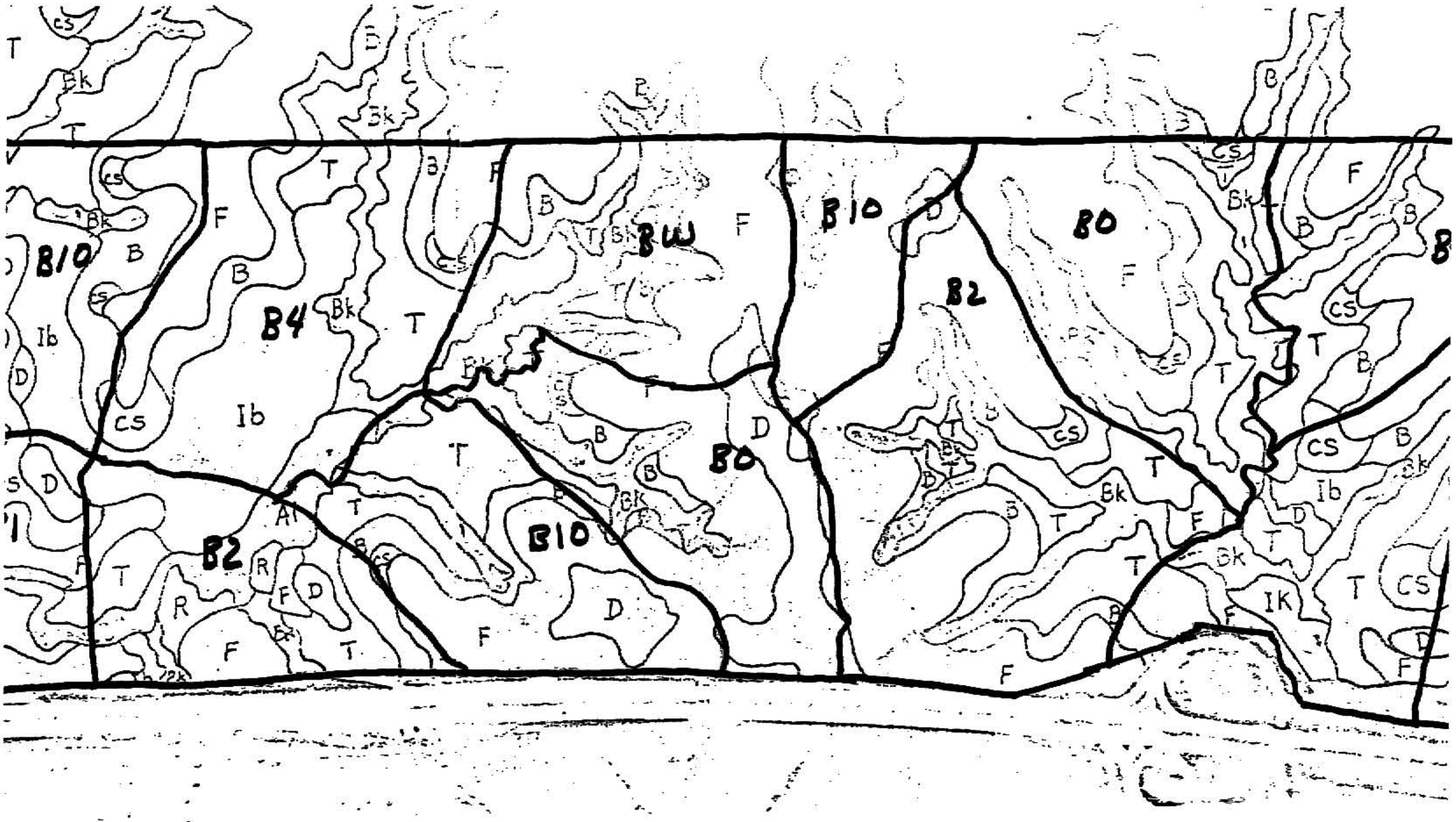
Ib Irwin silty clay loam, 4 - 8% slopes

IK Ivan and Kennebec silt loams

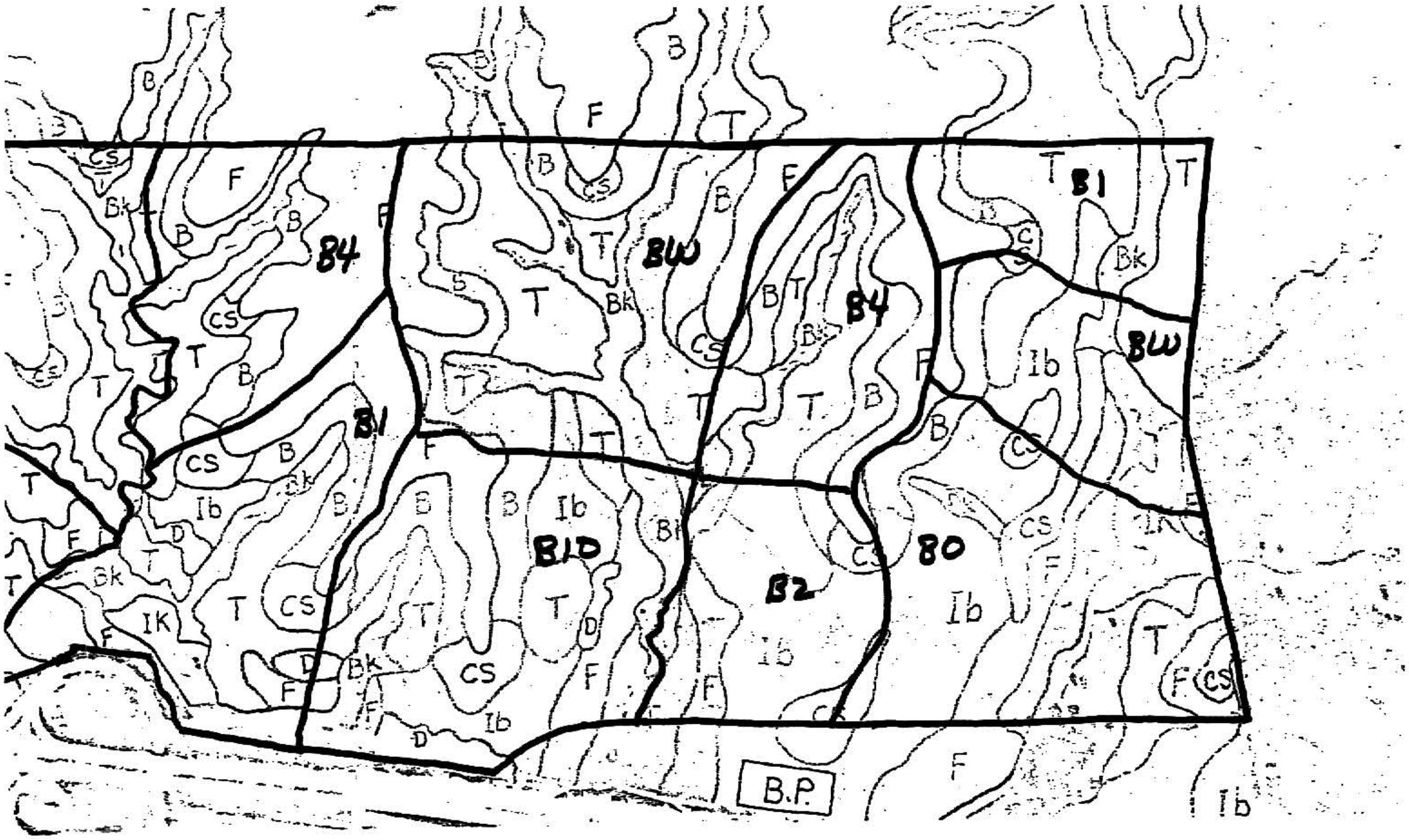
R Reading silt loam, 1 - 3% slopes

T Tully silty clay loam, 4 - 8% slopes

BP Borrow pit



B10
B4
BW
B10
B0
B2
B0
B10
B2


B4
BW
B4
B1
B1
BW
B1D
B2
80
B.P.

END

OF

OVERSIZED

DOCUMENT

LITERATURE REVIEW

There were basically two approaches found in the literature for the analysis of long-term experiments. The earlier papers have taken the view that yield variations with time can be best described by fitting a linear, quadratic, or at most, a low order time polynomial to the data. However, this has been criticized because the coefficients, outside of the linear terms, have little practical meaning. Further, if the errors from such trend lines are not independent in time, the method of least squares may lead to inefficient estimates. The second method, presented by Dutton (1951), describes the yield by a regression on *time* and an autocorrelated error. Maximum likelihood estimators of the parameters involved will be given.

Cochran (1939) suggests the possibility of running the experiment as a uniformity trial for the first year or two. As mentioned above, it is particularly important to lay out the plots to the best advantage and to avoid the use of highly variable sites. A delay of a couple of years in a ten to twenty year experiment is of no great account. Uniformity trials are especially helpful in experiments on new crops or in a new area. If used to group adjacent plots into blocks, the uniformity trial results are used to decide the shape of the blocks and the yields can be used to adjust subsequent years yields by means of covariance.

Cochran notes that the question of the amount of replication is difficult to answer. Much depends on the variability of the material, the duration of the experiment, and the precision with which measurements are taken. Another point to be noted is the fact that treatments remain on the same plots and any persistent differences between plots will not be "smoothed out" by averaging over years.

An example of the statistical analysis of an incomplete experiment on asparagus conducted over a ten year period is given by Snedecor and Haber (1946). Data for only the first ten harvested crops was analyzed since scarcity of labor terminated the experiment. Three cutting dates June 1, June 15, and July 1 comprise the treatments. The experiment was laid out in six randomized blocks, each block receiving each of the three treatments. The yields of asparagus (oz.) for Block 1 are given in Table 1.

TABLE 1

Yields of asparagus (oz.) from
3 plots for 10 years

Year	a	b	c		b-a	2c-(b+a)
1929	210	310	362		100	222
1930	230	296	353		66	180
1931	324	543	594		219	321
1932	512	778	755		266	220
1933	399	644	580		245	117
1934	891	1147	961		256	-116
1935	449	585	535		136	36
1936	595	807	548		212	-306
1937	632	804	565		172	-306
1938	527	749	353		222	-570
Total	4760	6654	5606		1894	-202
				Lin:	1426	-14450
				Quad:	-1233	-3127

If the cutting treatments June 1, June 15, July 1 are denoted by a, b, and c respectively, the following orthogonal treatment comparisons, T_1 and T_2, were made to learn if they contain information about what the future of the experiment might have shown.

CUTTING DATE

		a	b	c
TREATMENT COMPARISON	T_1	-1	1	0
	T_2	-1	-1	2

These quantities, T_1 and T_2, are computed for each block in each year and also shown in the table. For each block linear and quadratic components were calculated by use of orthogonal coefficients. The coefficients were as follows:

LINEAR: -9, -7, -5, -3, -1, 1, 3, 5, 7, 9

QUADRATIC: 6, 2, -1, -3, -4, -4, -3, -1, 2, 6

Thus, for example, in block I the components were

LINEAR: $-9(100) - 7(66) - 5(219) \ldots + 9(222) = 1426$

QUADRATIC: $6(100) + 2(66) - 1(219) \ldots + 6(222) = -1233$

The sum of the six linear components from all the blocks is then $1426 + 4262 + 2751 + 2077 + 3253 + 1126 = 14,895$.

Then the sum of squares attributable to the linear trend over years of the annual differences of the T_1 comparison is

$T_1Y_L = (14,895)^2/2(6)(330) = 56,026$ where $(-9)^2 + (-7)^2 + \ldots + (9)^2 = 330$; the 6 is due to there being six blocks; and the divisor $(-1)^2 + (1)^2 = 2$ converts the sum of squares calculated from differences to an individual plot basis. To test the significance of the linear component of trend, the appropriate estimate of error is calculated from the six block differences, $T_1Y_LB = \frac{(1426)^2 + \ldots + (1126)^2}{2(330)} - T_1Y_L = 10,535$. Then the mean square error for $T_1Y_LB = 10,535/5 = 2107$, the 5 degrees of freedom coming from the blocks (6-1). Testing for significance, $F = 56,026/2107 = 26.59$ which for 1 and 5 degrees of freedom is highly significant. However, Snedecor and Haber indicated that this linear trend did not persist after the first four years, the trend during the last six years being nonsignificant. They concluded the linear trend is not a secure basis for forecasting.

Deviations from the linear trend may be fitted by a second degree curve and these are found in a similar manner for the T_1 comparison. The sum of the six quadratic components is

$-1233 - 40 + 896 - 1076 - 726 - 375 = -2554$. It was noted that, except for Block III, the value of Q is negative, indicating a tendency for comparison T_1 to decrease in the latter part of the ten year period. The sum of squares due to the quadratic trend of annual differences for T_1 is $T_1Y_Q = \frac{(-2554)^2}{(2)(6)(132)} = 4118$ where 132 is the sum of the quadratic coefficients squared. The appropriate error term for testing the quadratic component of trend in T_1 is

$$T_1Y_QB = (-1233)^2 + \ldots + (-375)^2/2(132) - T_1Y_Q = 11,602.$$

The mean square for this error is 11,602/5 = 2302 and the test of significance yields F = 4118/2302 = 1.77 which is not significant for 1 and 5 degrees of freedom.

In contrast, the comparison T_2, which compares cutting to July 1, with the average of the two June cuttings has undoubted trends in the population. Block I shows that these differences changed from positive to negative about half way through the experiment. The computation of T_2 follow the same pattern as T_1 except the first divisor now becomes $(2)^2 + (-1)^2 + (-1)^2 = 6$. For example, for Block I, the linear component is;

$$\text{LIN:}\quad -9(222) - 7(180) - \ldots + 9(-570) = -14{,}450.$$

The six linear components are added yielding

$-14{,}450 - 16{,}504 - 12{,}615 - 12{,}885 - 10{,}261 - 10{,}648 = -77{,}363.$

The sum of squares due to the linear trend over years for this comparison is $T_2Y_L = \frac{(-77{,}363)^2}{6(6)(330)} = 503{,}791$. The appropriate error term for this linear component is $T_2Y_LB = \frac{(-14{,}450)^2 + \ldots + (-10{,}648)^2}{6(330)} - T_2Y_L = 13{,}893$. and the error mean square 13,893/5 = 2779, the 5 degrees of freedom again coming from blocks. Thus the test of significance F = 503,791/2779 = 181 is highly significant. The same procedure is followed for the quadratic trend for the T_2 comparison over years and it too is found to be highly significant. It is clear that cutting to July 1 depletes the vitality of the plants to such an extent that little further yield can be expected from these plots. The complete analysis is presented in Table 2.

TABLE 2

Analysis of variance of yields of asparagus.

Source	DF		SS
Blocks (B)	5		102,532
Treatments (T)	2		756,930
B×T	10		144,647
Years (Y)	9		5,461,360
B×Y	45		122,888
T×Y	18		717,343
T_1Y_L	1	9	56,026
T_1Y_Q	1		4,118
Remainder	7		32,384
T_2Y_L	1	9	503,791
T_2Y_Q	1		63,991
Remainder	7		57,033
B × T × Y	90		104,927
T_1Y_LB	5	45	10,535
T_1Y_QB	5		11,602
Remainder	35		20,279
T_2Y_LB	5	45	13,893
T_2Y_QB	5		18,271
Remainder	35		30,347
TOTAL	179		7,410,627

This example illustrates a method of analyzing the treatment X years interaction by partitioning the sum of squares into its linear and quadratic components for the two orthogonal comparisons of the three treatment cutting dates.

Cochran (1939) illustrates a similar procedure for a sugar-beet experiment and acid land dressings of 0, 1, 2, 3 and 4 tons of chalk applied in 1932 in a 5 × 5 latin square. The plots receiving no chalk gave such small yields that they were omitted and the data was analyzed as an incomplete Latin square. The analysis of variance was carried out for each year and the three degrees of freedom for the treatments in each year were partitioned into their linear, quadratic, and cubic components. The linear component provides an estimate of the average increase to higher dressings over the 1-ton dressing, while the quadratic term tests the falling off in responsiveness at the highest levels of application. Table 3 gives the analysis of variance of the experiment as a whole. The average differences between treatments are tested by analyzing the totals over the four years on each plot.

TABLE 3

Analysis of variance of sugar-beet yields.

Source	DF		SS		MS
Rows	4		46,387		
Columns	4		27,092		
Treatments	3		97,466		
linear		1		75,158	
quadratic		1		13,650	
cubic		1		8,658	
Error (1)	8		22,328		2791
Years	3		43,172		
Rows × Years	12		9,260		
Columns × Years	12		7,862		
Treatments × Years	9		13,747		1527
linear × years		3		11,181	3727
remainder		6		2,566	428
Error (2)	24		7,608		317

Cochran notes that the most interesting term in the interaction of treatments with years is the linear regression of yield on years which tests whether treatment differences are becoming more or less pronounced as the experiment proceeds. A separate error term appropriate to the linear regression may be obtained by calculating the regression separately for each plot and analyzing these figures in the same way as the plot totals were analyzed.

The most important rule governing rotation experiments is that each crop in the rotation must be grown every year. In a four-course rotation there is a choice of growing each of the four crops each year in single replication or each crop every four years in four-fold replication. In the case of the four-fold replication, there is increased accuracy in a single year's results of a particular crop, but the experiment will have to last longer to obtain equal information on long-term effects and if seasonal variations are great compared to variations in a single year, it will have to last almost four times as long. This rule imposes a lower limit on the size of the experiment. If there are t treatments in an r-course rotation, a single replication requires tr plots.

Crowther and Cochran (1942) describe three and four course rotation experiments on the Sudan Gezira in which cotton is the main crop. For the three course rotation there were five different rotations each with cotton as the common crop. One of the principal objectives of these experiments was to examine whether there are indications of long-term differences between rotations. The four degrees of freedom for rotations were partitioned into four comparisons of the five rotations and the rotations X years interaction was similarly partitioned.

Many times it is of interest to know whether the effects of a treatment are confined to the year applied, or whether the effects persist for some years afterward. In this type of long-term experiment Cochran (1939) notes that direct and residual effects may be separated by applying the treatments at fixed intervals. He stresses the idea that all phases of the treatment cycle should be present. For example, if the treatments were applied every third year, then some plots should receive the treatments in the first, fourth, ... years, some plots in the second, fifth, ... years, some plots in the third, sixth, ... years, and so on. If this rule is not followed, then differences between direct and residual effects may be confounded with seasonal (years) variation. However, if all phases are present, in any year there are plots which receive treatment in that year, plots which received treatment in the previous year, and so on.

Following this rule a minimum size of the experiment is set. For example, with eight treatments applied every three years in a four-course rotation, a single replication of the experiment requires 96 plots. With four crops (a, b, c, d) and treatments (t = treated, u = untreated), the first twelve years on a plot receiving treatment in the first year is as follows; 1-at, 2-bu, 3-cu, 4-dt, 5-au, 6-bu, 7-ct, 8-du, 9-au, 10-bt, 11-cu, 12-du. The average of a plot receiving treatment in the present, previous, or second previous year gives the average effect of a three yearly dressing of a treatment and this comparison is only four-fold replication for a given crop since a crop returns to the same plot every four years. To compare treatments in year of application only, or previous year only, the replication is only twelve-fold since the plot with crop a would be treated in the first year and not again until the thirteenth year. The permanent differences between plots are eliminated after three complete rotations, because crop a on the above plot represents the direct effect in the first year, first year residual effect in the fifth year, and second year residual effect in the ninth year. Cochran points out that this happens only when the periods of the crops and treatment cycles are different. If the crop periods and treatment cycles are the same, there will be no "smoothing out" of permanent differences between plots.

In experiments where different varieties are grown or treatments applied over several years, the interaction of varieties (or treatments) X years, if significant, indicates that some varieties (treatments) yielded better in some years whereas others were less productive over the years. LeClerg, et al. (1962) indicates that if the mean squares for varieties (treatments) is significantly greater than the mean square for the interaction of varieties (treatments) X years, then some varieties (treatments) may be superior to others and certain recommendations can be made. In the case of experiments on different varieties being grown at different locations as well as over years, the authors give the appropriate interpretations of the other interactions if they are found to be significant.

Yates and Cochran (1938) analyze an experiment on five varieties of barley at six experimental stations over two years. The period of two years is obviously not long-term but the analysis is suitable for more than two years. The original data and general analysis of variance are in Table 4 and Table 5.

TABLE 4
Yields of 5 varieties of barley in each of 6 locations in 1931 and 1932 (yields are totals of 3 replications).

		Varieties					
Place and	Year	Manchuria	Svansota	Velvet	Trebi	Peatland	Totals
University	1931	81.0	105.4	119.7	109.7	98.3	514.1
Farm	1932	80.7	82.3	80.4	87.2	84.2	414.8
Waseca	1931	146.6	142.0	150.7	191.5	145.9	776.5
	1932	100.4	115.5	112.2	147.7	108.1	583.9
Morris	1931	82.3	77.3	78.4	131.3	89.6	458.9
	1932	103.1	105.1	116.5	139.9	129.6	594.2
Crookston	1931	119.8	121.4	124.0	140.8	124.8	630.8
	1932	98.9	61.9	96.2	125.5	75.7	458.2
Grand	1931	98.9	89.0	69.1	89.3	104.1	450.4
Rapids	1932	66.4	49.9	96.7	61.9	80.3	355.2
Duluth	1931	86.9	77.1	78.9	101.8	96.0	440.7
	1932	67.7	66.7	67.4	91.8	94.1	387.7
TOTAL		1132.7	1093.6	1190.2	1418.4	1230.5	6065.4

TABLE 5
Analysis of variance of barley yields.

Analysis of Variance

Source	DF		SS		MS	
Places (P)	5		7073.64		1414.73	
Years (Y)	1	11	1266.17		1266.17	
P × Y	5		2297.96		459.59	
Varieties (V)	4		1769.99		442.50	
Trebi		1		1463.76		1463.76
Remainder		3		306.23		102.08
V × P	20		1477.67		73.88	
Trebi		5		938.09		187.62
Remainder		15		539.58		35.97
V × Y	4		97.27		24.32	
Trebi		1		7.73		7.73
Remainder		3		89.54		29.85
V × P × Y	20		928.09		46.40	
	59		14910.79			

Experimental error 216 $s^2 = 23.28$

It was observed in this experiment that one variety, Trebi, has a mean yield over all experiments that is substantially higher than the other varieties. The authors partition the sources of variation involving varieties into Trebi versus the remaining varieties. The 216 degrees of freedom for experimental error come from the fact that in the original experiment there were ten varieties. Yates and Cochran used only five in the analysis. An error term for the Places, Years, and Places × Years sources may be found to test these components. There are 6 × 2 = 12 experiments. This leaves 24 degrees of freedom for the within experiments error, which may be found by subtracting the above three sum of squares from the sum of squares due to replications over experiments.

YIELD MODELS INVOLVING AUTOCORRELATED ERROR

Cochrane and Orcutt (1949) indicate that the estimate of the variance from a set of residuals from least squares regression lines when the errors are autocorrelated are biased as compared to the residuals from such lines when the errors are independent. They studied this problem and found that the efficiency of the least squares estimators is low when the errors are autocorrelated.

Dutton (1951) compares the series of yields from a single plot of land to a time series stating there is little hope that the data will exhibit observable regularities to help predict future values. Further, he states that the coefficients of time polynomials has little biological meaning except for the linear term. The true deterioration or amelioration appears to be linear only for a short period of time for long-term systems. However, the most serious criticism has been directed at the independent error assumption. If there is correlation between successive yields from a single plot, Dutton lists two drawbacks to the application of the ordinary least squares methods. 1) The estimates so obtained have a low efficiency in repeated sampling when compared to the efficiency of estimates so obtained when in fact the yields are not correlated. 2) The estimates of variance obtained from residuals in the usual way are biased downwards.

Dutton cites two phenomena that seem to be apparent in examining long-term yield series data. 1) The general yield trend starts at a high level and decreases, usually rapidly at first and more slowly in later years, to a limiting value. In the case of certain cropping systems the general trend may be an amelioration from an initial value to a high final limiting value. 2) There is an oscillatory movement around the general trend, the form of the oscillation varying with the system under study.

Dutton gives the following model for a series of yields from a single crop grown repeatedly on a specified plot,

$y_i = a + bt_i + \eta_i$ [1] where $i = 1, 2, \ldots, n$

and the autocorrelated error model is of the form $\eta_i = \alpha\eta_{i-1} + \varepsilon_i$, [2]. Thus, the full model is given by $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$. The ε_i are independent and $|\alpha| < 1$. The alpha (α) expresses the degree of linear dependence of a given deviation from the trend on the previous deviation. The ε_i are the random yearly deviations. The independent variate t_i denotes the general regression on a single independent variate as well as regression on time. We take $\varepsilon_1 = \eta_1$ since the system starts at time $i = 1$ and there can be no dependency of the first deviation on some previous deviation.

The first part of the model, $a + bt_i$, depends only on the system itself, that is, the fact that a given crop is grown repeatedly under specific treatment on a given plot of ground.

The second part $\alpha\eta_{i-1}$ denotes the portion of yield i that is determined by the magnitude of previous yield (i - 1). Alpha (α) varies with the system. If the activity of the agents of nature which replenish fertility is such that a chance high yield the previous year would lead to less available nutrients in the current year, then α will be negative. If however, the previous year's yield tends to allow high yields in the current year so that the replenishing agents are of the opposite kind, then α will be positive.

As already mentioned, the ε_i are the random yearly deviations. It is the portion of the model that is due to the random seasonal effects.

It was also noted the ε_i are independent and if they are assumed to be normally distributed with zero means and variance σ^2, except ε_1 which has variance $\frac{\sigma^2}{1 - \alpha^2}$, then the deviations η_i will also be normally distributed with mean zero and common variance σ^2_η. It would be simpler to take the variance $\varepsilon_1 = \sigma^2$ but that would mean η_1 would have variance differing from that of the other η_i, $i = 2, 3, \ldots, n$. The η_i are not independent. Their functional form is as follows,

$$\underline{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix} = \begin{bmatrix} y_1 - a - bt_1 \\ y_2 - a - bt_2 \\ \vdots \\ y_n - a - bt_n \end{bmatrix} = \underline{y} - a\underline{j} - b\underline{t}$$

where the characters underlined indicate a vector and $\underline{j} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, a vector of ones.

To find the maximum likelihood estimators for a, b, and α, the transformation is first made;

$$\underline{\xi} = \begin{bmatrix} \sqrt{1 - \alpha^2} & \varepsilon_1 \\ & \varepsilon_2 \\ & \vdots \\ & \varepsilon_n \end{bmatrix} = \underline{\underline{R}}\underline{\varepsilon}$$

where

$$\underline{R} = \begin{bmatrix} \sqrt{1-\alpha^2} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & & 1 \end{bmatrix}$$

Also,

$$\underline{\varepsilon} = \begin{bmatrix} \eta_1 \\ \eta_2 - \alpha\eta_1 \\ \vdots \\ \eta_n - \alpha\eta_{n-1} \end{bmatrix} = \underline{T}\underline{\eta}$$

where

$$\underline{T} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\alpha & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & & -\alpha & 1 \end{bmatrix}$$

Then $\underline{\xi} = \underline{R}\underline{T}\underline{\eta}$ and $\underline{\xi}'\underline{\xi} = \underline{\eta}'\underline{T}'\underline{R}'\underline{R}\underline{T}\underline{\eta} = \underline{\eta}'\underline{A}\underline{\eta}$ where the matrix $\underline{A}$ is

$$\underline{A} = \begin{bmatrix} 1 & -\alpha & 0 & \cdots & 0 & 0 & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \cdots & 0 & 0 & 0 \\ 0 & -\alpha & 1+\alpha^2 & \cdots & 0 & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & \cdots & -\alpha & 1+\alpha^2 & -\alpha \\ 0 & 0 & 0 & \cdots & 0 & -\alpha & 1 \end{bmatrix}$$

Then the distribution of the observational vector $\underline{y}$ is given by

$$f(\underline{y}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \{\exp - \frac{1}{2\sigma^2} (\underline{y} - a\underline{j} - b\underline{t})' \underline{A}(\underline{y} - a\underline{j} - b\underline{t})\}J. \qquad [3]$$

This is a function of the model parameters $L(a, b, \alpha, \sigma^2)$ and the natural logarithm is given by

$$\ln L = - \frac{n}{2} \ln 2\pi - n \ln \sigma + \ln J - \frac{1}{2\sigma^2} \underline{\eta}'\underline{A}\underline{\eta}, \qquad [4].$$

To obtain the estimators Dutton equates to zero the derivative of [4] with respect to each of the parameters. The solutions are given in terms of the other parameters which are usually (and practically) unknown.

For $\hat{a}$, $\frac{\partial \ln L}{\partial a} = \frac{\partial \underline{\eta}'\underline{A}\underline{\eta}}{\partial a} = -2(\underline{y} - a\underline{j} - b\underline{t})' \underline{A}\underline{j} = 0$ or

$$X\hat{a} + Yb = U \qquad [5]$$

where $X = \underline{j}'\underline{A}\underline{j} = (1-\alpha)[2 + (n-2)(1-\alpha)]$

$$Y = \underline{t}'\underline{A}\underline{j} = (1-\alpha)[t_1 + t_n + (1-\alpha) \sum_{i=2}^{n-1} t_i]$$

$$U = \underline{y}'\underline{A}\underline{j} = (1-\alpha)[y_1 + y_n + (1-\alpha) \sum_{i=2}^{n-1} y_i].$$

This can be solved for $\hat{a}$ if b, α are known.

Similarly, for $\hat{b}$, $\frac{\partial \ln L}{\partial b} = \frac{\partial \underline{\eta}'\underline{\underline{A}}\underline{\eta}}{\partial b} = -2(\underline{y} - a\underline{1} - b\underline{t})'\ \underline{\underline{A}}\underline{t} = 0$ or

$Ya + Z\hat{b} = V$ [6]

where $Z = \underline{t}'\underline{\underline{A}}\underline{t} = t_1^2 + t_n^2 - 2\alpha \sum_{i=1}^{n-1} t_i\ t_{i+1} + (1+\alpha^2) \sum_{i=2}^{n-1} t_i^2$

$$V = \underline{y}'\underline{\underline{A}}\underline{t} = y_1t_1 + y_nt_n - \alpha \sum_{i=1}^{n-1} y_it_{i+1} - \alpha \sum_{i=2}^{n} y_it_{i-1} + (1+\alpha^2) \sum_{i=2}^{n-1} y_it_i$$

If a and α are known [6] can be solved for $\hat{b}$ or if only α is known, [5] and [6] can be solved simultaneously for $\hat{a}$ and $\hat{b}$.

To obtain the maximum likelihood estimator for α, Dutton ignores the Jacobian of the transformation noting that if the slightly different assumption that the variance $\varepsilon_1 = \sigma^2$ had been adopted, the Jacobian would just have been 1, a constant. Then the equation to be solved for $\hat{\alpha}$ is

$$\frac{\underline{\eta}'\partial \underline{\underline{A}}\underline{\eta}}{\partial \alpha} = 0 \quad \text{or} \quad \hat{\alpha} = \sum_{i=1}^{n-1} \frac{(y_i - a - bt_i)(y_{i+1} - a - bt_{i+1})}{\sum_{i=2}^{n-1} (y_i - a - bt_i)^2}, \text{ [7]}.$$

Finally, the maximum likelihood estimator for σ^2 is found from

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{\underline{\eta}'\underline{\underline{A}}\underline{\eta}}{2\sigma^4} = 0 \quad \text{or} \quad \hat{\sigma}^2 = \frac{\underline{\eta}'\underline{\underline{A}}\underline{\eta}}{n}$$ if a, b, and α are all known.

In the most practical situation where none of the parameters a, b, α, and σ^2 is known, an iterative method of simultaneous solution can be used. The equations to be solved are

$$\begin{aligned} X(\hat{\alpha})\hat{a} + Y(\hat{\alpha})\hat{b} &= U(\hat{\alpha}) \\ Y(\hat{\alpha})\hat{a} + Z(\hat{\alpha})\hat{b} &= V(\hat{\alpha}) \end{aligned} \qquad [8]$$

$$\hat{\alpha} = \sum_{i=1}^{n-1} \frac{(y_i - \hat{a} - \hat{b}t_i)(y_{i+1} - a - \hat{b}t_{i+1})}{\sum_{i=2}^{n-1} (y_i - \hat{a} - \hat{b}t_i)^2} \qquad [9]$$

where, $X(\hat{\alpha}) = (1-\hat{\alpha})[2 + (n-2)(1-\hat{\alpha})]$

$$Y(\hat{\alpha}) = (1-\hat{\alpha})[t_1 + t_n + (1-\hat{\alpha}) \sum_{i=2}^{n} t_i]$$

$$Z(\hat{\alpha}) = t_1^2 + t_n^2 - 2\hat{\alpha} \sum_{i=1}^{n-1} t_i\, t_{i+1} + (1+\alpha^2) \sum_{i=2}^{n-1} t_i^2$$

$$U(\hat{\alpha}) = (1-\hat{\alpha})[y_1 + y_n + (1-\hat{\alpha}) \sum_{i=2}^{n} y_i]$$

$$V(\hat{\alpha}) = y_1t_1 + y_nt_n - \hat{\alpha} \sum_{i=1}^{n-1} y_i\, t_{i+1} - \hat{\alpha} \sum_{i=2}^{n} y_i\, t_{i-1}$$

$$+ (1+\alpha^2) \sum_{i=2}^{n-1} y_it_i$$

The above equations can be solved iteratively by the following steps Dutton has set up;

1. Select trial values of α, say $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_K$.
2. For each trial value solve the pair of equations [8] simultaneously for $\hat{a}_i$ and $\hat{b}_i$.
3. Substitute each pair $\hat{a}_i$ and $\hat{b}_i$ in [9] and solve for $\hat{\alpha}_i$.
4. If $\tilde{\alpha}_i = \hat{\alpha}_i$ for some value i, the set $\hat{a}_i$, $\hat{b}_i$, and $\hat{\alpha}_i$ will be the solutions.

5. If $\tilde{\alpha}_i \neq \hat{\alpha}_i$ for all i, then by interpolation among the $\tilde{\alpha}_i$ a value of α can be obtained such that the $\hat{\alpha}$ will be $\cong \alpha$ if α is used as a trial value.
6. Either this value can be taken as the approximate value of the estimate and the corresponding estimates for a and b obtained by inverse interpolation, or
7. A second series of trial values of α in the neighborhood of this approximate value can be selected and the process repeated until an $\tilde{\alpha}_i = \hat{\alpha}_i$ is obtained.

It is noted that the estimates obtained are only approximate joint maximum likelihood estimators since the Jacobian is ignored.

In the above situation where none of the parameters are known, the estimate of σ^2 becomes $\hat{\underline{\eta}}'_{ab}\ \hat{\underline{A}}\hat{\underline{\eta}}_{ab}/n-3$ where $\hat{\underline{\eta}}_{ab} = (\underline{y} - \hat{a}\underline{j} - \hat{b}\underline{t})$, and $\hat{\underline{A}}$ indicates that α has been estimated by $\hat{\alpha}$. Then confidence intervals may be constructed as follows

$$\hat{a} - t_{\frac{\gamma}{2},n-3}\left[\frac{\hat{\underline{\eta}}_{ab}\ \hat{\underline{A}}\hat{\underline{\eta}}_{ab}(\hat{Z})}{(n-3)(\hat{X}\hat{Z}-\hat{Y}^2)}\right]^{1/2} < a < \hat{a} + t_{\frac{\gamma}{2},n-3}\left[\frac{\hat{\underline{\eta}}'_{ab}\ \hat{\underline{A}}\hat{\underline{\eta}}_{ab}(\hat{Z})}{(n-3)(\hat{X}\hat{Z}-\hat{Y}^2)}\right]^{1/2}$$

and

$$\hat{b} - t_{\frac{\gamma}{2},n-3}\left[\frac{\hat{\underline{\eta}}'_{ab}\ \hat{\underline{A}}\hat{\underline{\eta}}_{ab}(\hat{X})}{(n-3)(\hat{X}\hat{Z}-\hat{Y}^2)}\right]^{1/2} < b < \hat{b} + t_{\frac{\gamma}{2},n-3}\left[\frac{\hat{\underline{\eta}}'_{ab}\ \hat{\underline{A}}\hat{\underline{\eta}}_{ab}(\hat{X})}{(n-3)(\hat{X}\hat{Z}-\hat{Y}^2)}\right]^{1/2}$$

where $t_{\frac{\gamma}{2},n-3}$ is the upper $(1 - \gamma/2)$ 100% point of student's t distribution and γ is the level of significance.

Dutton extends the procedure to more than one plot (or more than one treatment). The model is $y_{ij} = a_i + b_i t_{ij} + \alpha_i \eta_{i,j-1} + \varepsilon_{ij}$ where $i = 1, \ldots, K$ indicates the plots and $j = 1, \ldots, n$ indicates the years. Transformations are used and approximate estimators are found. However, the procedure and estimation is beyond the scope of this report.

Therefore, the following testing procedures are proposed when there is more than one plot or more than one treatment. The first model given, $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$, $i = 1, 2, \ldots, n$, is used to obtain estimates of a, b, α, and σ^2 for each of t treatments in an experiment. It is of interest to test the hypothesis H_o: b = 0 for each of the treatments, because b_j for treatment j is an estimate of the increase or decrease in yield over the years the experiment was performed. The principle of conditional error can be used to test H_o: b = 0 vs. H_a: b $\neq$ 0. The error sum of squares (ESS) for the full model $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$ is as shown above $\hat{\underline{\eta}}'_{ab} \hat{\underline{A}} \hat{\underline{\eta}}_{ab}$. Now, restrict the model by the null hypothesis, i.e., $y_i = a + \alpha\eta_{i-1} + \varepsilon_i$ and find the error sum of squares conditional (ESS_c), given by $\hat{\underline{\eta}}'_a \hat{\underline{A}} \hat{\underline{\eta}}_a$ where $\hat{\underline{\eta}}_a = (\underline{y} - \hat{a}\underline{j})$. It will be necessary to re-estimate a and α for the particular b_j being tested. This can be done by following the iterative procedure given earlier.

1. Choose trial values of α, say $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_K$.
2. For each trial value of α_i, solve equation [5] for $\hat{a}_i$.
3. Substitute $\hat{a}_i$ and b = 0 into equation [9] and solve for $\hat{\alpha}_i$.

4. If $\tilde{\alpha}_i = \hat{\alpha}_i$ for some value i, then the solutions are $\hat{\alpha}_i$ and $\hat{a}_i$.
5. If $\tilde{\alpha}_i \neq \hat{\alpha}_i$ for all values of i, then follow steps 6 and 7 given previously for the solutions.

The sum of squares due to the hypothesis (SSH_o) is $SSH_o = ESS_c - ESS$ with $(n-2) - (n-3) = 1$ degree of freedom. Then $\dfrac{SSH_o/1}{ESS/n-3}$ is distributed approximately as an F statistic with (1, n-3) degrees of freedom. It is shown in the dissertation by Dutton that $\hat{\underline{\eta}}'_{ab}\,\hat{\underline{A}}\hat{\underline{\eta}}_{ab}/\sigma^2$ and $\hat{\underline{\eta}}'_{a}\,\hat{\underline{A}}\hat{\underline{\eta}}_{a}/\sigma^2$ are distributed as χ^2 random variables with (n-3) and (n-2) degrees of freedom respectively.

STATISTICAL TREATMENT OF KONZA PRAIRIE PROJECTS

Most of the researchers conducting projects on Konza Prairie are in the planning stages of their experiments. Therefore, the statistical comments that follow will have to be general in content.

More discussion is in order concerning the decision to bypass a uniformity trial. As pointed out previously, the delay of two or three years in a 20 year or longer experiment is of minor consequence. It is my opinion that the whole area should have been burned (or unburned) for at least two years. The researchers should have taken observations in the same manner as they plan to do for the experiment itself. From these observations, certain differences between plots may have been present. Further, it may have been necessary to rearrange the plots to account for the differences found in the trial years. At the end of 20 or 30 years, long-term trends may indicate marked differences between the burning treatments. But the question remains as to how much of the differences between plots was already present due to the variable amounts and arrangements of the soil types. Thus, it may be difficult to make valid inference with respect to the burning treatments.

It is also my opinion that plots to be burned every other year and every fourth year should not all be burned in the same year. For example, all plots being annually burned were done so in April, 1972. It is suggested that of the plots to be burned every two years, two should be burned in April, 1973 and the remaining two burned in April, 1974. A similar procedure of burning may be followed for the four year burning treatment as one may be burned in April, 1976, one in 1977, and so on. If this timing procedure is followed, all phases of the treatment cycle will be present and it will help eliminate yearly variation due to environment. The phasing of the treatment cycles should be followed unless the researchers feel they cannot afford the delay on these plots in the early years of experimentation.

One project on the Konza Prairie is concerned with the insect population. The burning treatments of interest are no burning and annual burning. The number of insects found in 8 to 12 drop trap samples per season along with the total biomass are to be measured. Later, a classification of the insects into families is planned. Initially, it is not known whether soil types will be a factor. Had the researcher been able it would have been better to sample more than one soil type within the two burning treatments, but owing to a limitation of time and labor only one soil type can be sampled. Also, it is not known whether differences would arise as to when the samples were taken with respect to time of burning. It was suggested that the researcher take his samples at the same time each season. This particular study is to be carried out in conjunction with another entomology project at another location (Donaldson) in the Flint Hills, where different fertility treatments have been applied. The same soil type should be sampled in each location in order to make valid comparisons. However, it should be noted that the Konza projects are just starting while application of treatments and experimentation have been carried out for several years at Donaldson.

The researcher has indicated he would sample only two plots, one of each of the treatments B0 and B1. Therefore, with only two treatments and a common soil type, to test the hypothesis that there is no difference in burning treatments, i.e., H_o: $\mu_{B0} = \mu_{B1}$, one can use the t-test in comparing the two treatment means.

If it should be decided after a few years that the researcher can sample more than one soil type, a split-plot design could be set up with burning treatments as the whole plots and soil types as the subplots. The analysis of variance follows;

Analysis of Variance

Source	DF
Replication	(r-1)
Burn Treatments	1
R × BT (error a)	(r-1)
Soils	(s-1)
BT × S	(s-1)
Error (b)	2(r-1)(s-1)

To test the hypothesis H_o: B0 = B1, the test statistic would be $F = \frac{MS_{BT}}{MS_{R \times BT}}$ which is distributed as an F statistic with 1 and (r - 1) degrees of freedom.

The long-term changes of burning treatments can be examined by using the procedure outlined in the section on autocorrelated errors where the model was $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$, $i = 1, \ldots, n$, the terms in the model having already been defined. The mean value for a given year of a particular burning treatment will make up the series of yield data, i.e., the observations y_i. The parameters of the model can be estimated, confidence intervals constructed from the estimates, and a test of the hypothesis H_o: $b = 0$ can be performed for the burning treatments.

Another research project at Konza Prairie involves the vegetation composition in relation to the burning treatment and soil types. This study will be more extensive but similar to the entomology study. The researcher uses a coding system to designate the number of a particular species on a burn treatment-soil type combination. He is interested not only in how a species changes over years but also the competition between species on the same plot. The researcher should find an appropriate model that describes the competition between species to obtain an index. This index, calculated in some manner each year, could then be used as the observations, or y_i values, that constitute the series of yield data for the model $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$. The hypothesis H_0: b = 0 tests if the rate of change in the competition among species is zero. The conclusions and inference will depend on the model of competition as to what b is estimating. If the researcher wishes to examine within year differences of burning treatments and soil types, a split-plot design might be suitable. This project is to include all burning treatments, thus, the sum of squares due to treatments and soils should be partitioned into comparisons of interest.

A third research project planned, concerning runoff water quality, may be more difficult to analyze. The researcher was unavailable for consultation, therefore, the following comments will be general with respect to this project. The runoff water is to be sampled at the point where a particular plot is drained, and it will be difficult (if not impossible) to determine whether differences in quality of runoff water are due to the burning treatments or the soil types. That is to say, burning treatments and soil types will be confounded because of the variable amounts and uneven distribution of the soil types on the plots. One (partial) solution might be to establish microwatersheds (.5 hectare $\doteq$ 1.235 acre) so that only one or two soil types are included. A map of soil types indicates that this is possible if the researcher agrees. However, the area of prairie immediately north of Konza may have runoff effect on the northernmost plots, and in turn they may have an effect on the southern plots. Although the soils map shows that several of the plots contain nearly the same soil types, the variable amounts, uneven distribution, and unlike ordering within plots could invalidate the study if runoff is measured at the drainage point of the plots. By using the microwatersheds the contributions to runoff from the burning treatments and soil types might be better assessed. The time of observations may also be a factor. Therefore, it is important that the researcher take measurements at the same time each season, whether it be before burning, right after, or near the end of the summer.

To examine the long-term effects of a burning treatment, the mean of the observations for a particular burning treatment (the observation itself if only one per treatment) in a year should be used as the y_i values in the model $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$. The parameter b measures the rate of change in the runoff water quality over the years the experiment is conducted. But, it should be kept in mind that if observations are taken on the entire plot, any conclusions drawn may be invalid for the reasons explained above.

Research is also planned involving the bird population. The birds are to be observed every week from April to August and every other week from September to March. The researcher will visually count the number of birds seen in such a manner that species are also noted. Two plots were chosen to be used, one each of the unburned and annually burned. The two plots were chosen so that soil types were as similar as possible.

This allows for a close resemblance of vegatation. It is of interest to determine how the population numbers will change over the years of experimentation. The regression model with autocorrelated error, $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$, can be used. The researcher indicated that the mean number of birds for a season would be of interest. Also to be recorded annually was the maximum number to determine how it was changing. Due to migratory practices of the birds, certain other measurements might also be of interest. These measurements can make up the y_i values for the series of yield data. In this research, b measures the rate of change in the population number per season.

A study on the population of lizards will be conducted only in the rocky areas of Konza Prairie and this includes only two or three plots. The researcher has indicated he anticipates no influence from the burning treatments, and thus has no interest in them as they will not affect the rocky areas. The researcher does plan to take preliminary observations in the first two or three years to learn what is on the prairie initially. While this is not a uniformity trial, it is a step in that direction.

SUMMARY

Long-term experimentation has not been extensively used due to the fact that experimental stations have not been willing to involve large areas of land for several years for this purpose. The patience, labor, and costs required are sometimes too great.

There were basically two approaches found in the literature concerning the analysis of long-term experiments. One approach was the analysis of variance technique where certain sums of squares were partitioned into time trends. The other approach was that of a regression with an autocorrelated error structure.

Various long-term research projects have been planned for the Konza Prairie. There are six burning treatments to be applied to the 24 plots. The plots have different amounts, arrangements, and types of soils making it important to examine the effects of these factors. However, owing to a shortage of time and labor, many of the research projects will not be of a large scale, but rather will involve only a small part of the land area. The purpose of this paper is to make it known that some thought was given to the Konza Prairie from a statistical standpoint. The methods developed by Dutton and outlined in this paper can be used to examine the long term trends.

ACKNOWLEDGEMENTS

The writer wishes to express his gratitude to Dr. Arthur D. Dayton for his guidance, assistance, and encouragement during this report. Deeply appreciated are the many helpful suggestions provided by him during the entire period of graduate study.

REFERENCES

Cochran, W. B.
Long-term agricultural experiments. Journal of the Royal Statistical Society, Supplement 6:104-148, 1939.

Cochrane, D. and G. H. Orcutt.
Application of least-squares regression to relationships containing autocorrelated error terms. Journal of the American Statistical Association, 44:32-61, 1949.

Crowther, F. and W. G. Cochran.
Rotation experiments with cotton in the Sudan Gezira. Journal of Agricultural Science, 32:390-405, 1942.

Dutton, A. M.
Statistical analysis of long-term experiments. Unpublished Ph.D. thesis. Iowa State College, 1951.

LeClerg, E. L., W. H. Leonard and A. G. Clark.
Field plot technique. Second edition. Minneapolis: Burgess Publishing Company, 1962.

Snedecor, G. W. and E. S. Haber.
Statistical methods for an incomplete experiment on a perennial crop. Biometrics, 2:61-67, 1946.

Yates, F., and W. G. Cochran.
The analysis of groups of experiments. Journal of Agricultural Science, 28:556-580, 1938.

LONG-TERM EXPERIMENTS AND THE KONZA PRAIRIE

by

PHILIP K. CLARK

B.S., Kansas State University, 1971

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

Experiments dealing with a group of factors are often carried out at a number of places and over a number of years because their effects may differ from place to place and from year to year. Thus, the design of long-term experiments is very important. Some ideas on the construction, design, and analysis of long-term experiments are discussed in this report. The advantages of a uniformity trial are also considered as it is helpful in assigning treatments or arranging blocks and the observations taken during the trial year(s) can be used as a covariate(s).

The two basic approaches to the analysis of long-term experiments is presented. One method is an analysis of variance technique including three illustrative examples that are given. The second method uses a regression model with an autocorrelated error structure. The model is $y_i = a + bt_i + \alpha\eta_{i-1} + \varepsilon_i$. Estimates of the parameters of the model, confidence intervals, and a test of the hypothesis H_o: $b = 0$ are all given in the report.

A long-term experiment is being conducted on the Konza Prairie, a strip of land south of Manhattan, Kansas. Various research projects that have been and are being planned for the prairie are discussed in the paper. Also, comments are made in regard to the physical aspects of the land area and their relation to the experiments. Additional remarks are made concerning a uniformity trial on the prairie and the phasing of the treatment cycles.