Reprinted from the *Soil Science Society of America Journal*
Volume 47, no. 3, May–June 1983
677 South Segoe Rd., Madison, WI 53711 USA

Replications and Subsamples Needed to Show Treatment Responses on Forest Soils of the Coastal Plain[1]

F. THOMAS LLOYD AND WILLIAM H. MCKEE, JR.[2]

ABSTRACT

This research illustrates how randomized, complete-block designed field studies can be insensitive to detecting treatment-caused differences in the concentration of soil nutrients. An example shows that variation without compositing can be so large that 727 blocks are needed to be 90% confident of detecting a difference in total nitrogen of 100 $\mu g\ g^{-1}$. Graphs are presented for determining the optimum numbers of blocks and chemical determinations per plot. The results show that compositing does not affect the number of chemical determinations per plot, but does reduce the number of blocks in direct proportion to the reduction in variance resulting from compositing. A hypothetical example shows that the number of blocks can be reduced from 727 to 15 by constructing blocks with less natural variation between treatment plots and by compositing 20 sample cores for each chemical determination.

Additional Index Words: **sample size, type-II error, experimental design, variance, field trials, nitrogen.**

Lloyd, F.T., and W.H. McKee, Jr. 1983. Replications and subsamples needed to show treatment responses on forest soils of the Coastal Plain. Soil Sci. Soc. Am. J. 47:587–590.

NONSIGNIFICANT ANALYSES of variance (ANOVA's) occur with discomforting frequency in soil nutrient research, no doubt due in part to the substantial variability inherent in nutrient concentrations determined from small sample cores. Of course, small sample cores are necessary because of both the destructive nature of the nutrient analyses and the laboratory techniques for determining nutrient concentration. Such variability is particularly apparent for a variable like the $\mu g\ g^{-1}$ of total nitrogen in a forest soil (the response used to illustrate this procedure), but can also occur when large-scale uniformity is otherwise assumed.

Dealing with the problem of nonsignificant ANOVA's requires planning for that outcome before installing the experiment. Our goal is to supply aids for doing this. We have focussed attention on the randomized, complete block (RCB) experimental design because of its wide application in field studies. For application purposes, our solution is presented in graphic form. The figures allow us to estimate the number of blocks (interchangeably called replicates) and the number of chemical determinations per treatment plot needed for an efficient RCB design. By efficient, we mean the experiment has a high probability (equal to 1 minus the Type-II error rate) of detecting biologically meaningful treatment differences when they are in fact present.

[1] Contribution from the USDA, Forest Service, Southeastern Forest Exp. Stn., Asheville, N.C. Received 5 Apr. 1982. Approved 20 Dec. 1982.

[2] Biometrician and Principal Soil Scientist, Forestry Sciences Lab., 2730 Savannah Highway, Charleston, SC 29407.

RESULTS

Mixing (also called compositing) groups of sample cores is a natural response to this variability problem. We are assuming that the effect of thoroughly mixing r sample cores is similar to the effect averaging has on the variance of a mean, that is, to reduce variation by a factor of $1/r$. The assumption could be (and probably should be) checked by calibration studies designed for that purpose. We will consider this assumption a sufficiently good first approximation for the purposes of this research.

We need to know (i) how to balance the choice between the number of blocks and the number of chemical determinations, and (ii) the effect mixing has on these choices. Figure 1 shows that the number of chemical determinations is a function of a variance ratio called ρ. This is the ratio of the plots-in-blocks variance component (σ_δ^2) over the determinations-in-plots variance component (σ_ϵ^2). Values of these variance components are obtained either from previous RCB experiments in the same population of blocks or from an analysis of nutrient

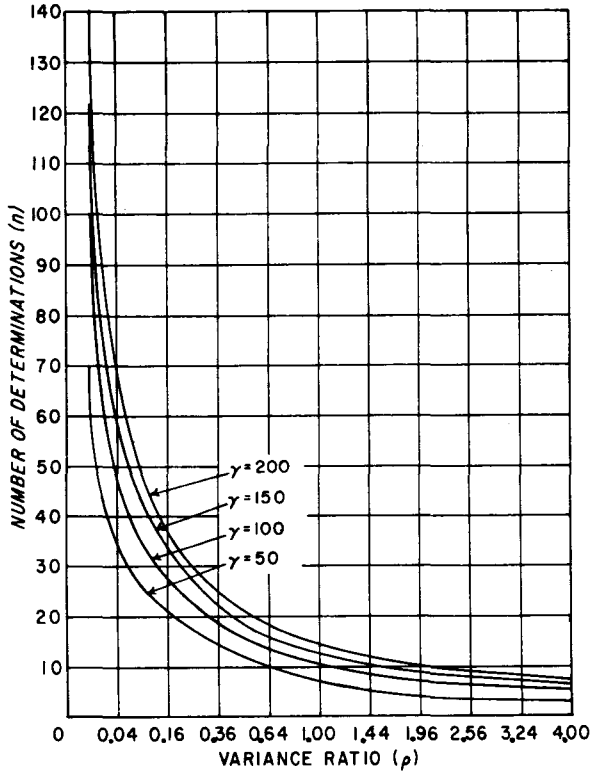

Fig. 1—The relationship of the number of chemical determinations (n) to the variance ratio (ρ) and cost ratios (γ) where $n = (\gamma/\rho)^{1/2}$ (Eq. [13]) with $\gamma = c_\delta/c_\epsilon$ and $\rho = \sigma_\delta^2/\sigma_\epsilon^2$.

concentrations from a preliminary sampling of the study area. The effect on these variance components of mixing small soil cores is to reduce both of them proportionally the same, which means the effect cancels in the variance ratio ρ. This says that compositing does not affect the number of chemical determinations per plot.

A cost ratio (γ) is also used in Fig. 1 for estimating the number of chemical determinations. It is the cost (either in time or dollars) of locating and establishing a single block divided by the cost of making a chemical determination. You would expect these quantities to have large values. Equation [13] in the statistical theory section at the end of the paper could be used to estimate the number of chemical determinations for γ's other than those on the graph. Cost is considered in our solution because it was used in the optimization process for balancing the choice between the number of determinations with the number of blocks. It is not necessary to understand this optimization process to use the results, but for those who might be interested, it is spelled out in the statistical theory section.

The number of blocks is determined using Fig. 2 and 3, where ρ and γ are as defined above. To estimate the number of blocks, the f factor from Fig. 2 is multiplied by the ratio $\sigma_\epsilon^2/(\theta\Delta^2)$, where Δ is the biologically meaningful difference the experimenter wants to detect with a stated measure of confidence called the power of the test. The experimenter must decide on both the value of Δ and the power of the test. Given the selected power and the α-level for the ANOVA, θ can be obtained from Fig. 3. Because the multiplier for f in Fig. 2 contains the variance component σ_ϵ^2, it is clear that mixing reduces the number of blocks in direct proportion to its reduction in variation. A further reduction in the number of blocks can be affected by reducing the value of ρ. This is accomplished by selecting blocks that have less between-plot variability in the nutrient being investigated. As with chemical determinations, Eq. [11] and Eq. [12] can be used to estimate the number of blocks for γ's not on the graph.

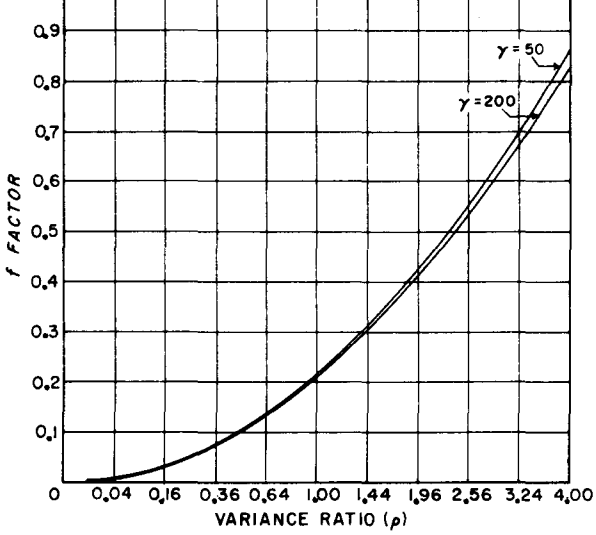

Fig. 2—The blow-up factor f depicts the effect of the variance component ratio (ρ) and the cost ratio (γ) on the number of blocks (Eq. [12]).

Two additional points of application are worth emphasis. First, the RCB model assumes that blocks are randomly sampled from some clearly delimited population of blocks. It frequently occurs that this condition is not met because we are limited by the availability of study areas and by time and money, thus causing uncertainty about the limits of our inferences. Any effort to define the population of blocks in advance, with commensurate attempts to randomly sample it, strengthens the inferences. Second, variance component estimates using the characteristically small data sets from RCB designs have large variances themselves, so use of ρ in Fig. 1 and 2 introduces an unavoidable element of instability in the sample size estimates. As a consequence, they should be interpreted as ball-park values. We believe they are precise enough to help us design more sensitive experiments.

EXAMPLE

The method was tested on soil nitrogen data taken from three prescribed burn studies installed over a range of forest soils in the southeastern Coastal Plain (McKee, 1982).

A South Carolina study, comprising a series of burn treatments initiated 30 years before the time of sampling (McKee, 1982) is located on a Pleistocene terrace of the Lower Coastal Plain. The terrain is nearly level and drainage ranges from somewhat poorly to poorly drained. Plots are 0.10 ha in size with three replications located on the Santee Experimental Forest near Charleston, and two replications located on Westvaco Corporation land near Andrews. Soils are described as Bayboro, Bladen, Coxville, Dunbar, Duplen, Goldsboro, and Lynchburg. Treatments selected for testing were a no-burn control and an annual summer burn.

An Alabama study located on the Escambia Experimental Forest near Brewton, Ala., on Upper Coastal Plain terrain (McKee, 1982) has plots 0.16 ha in size with eight replicated blocks. The soils are classed as Benndale, complexed with

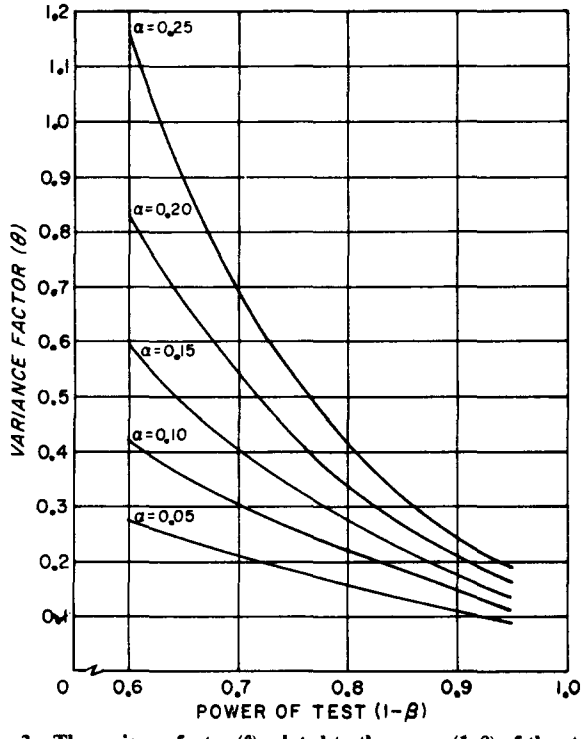

Fig. 3—The variance factor (θ) related to the power ($1-\beta$) of the statistical test for a given Type-I error rate (α).

Orangeburg, Troup, and Saffell. Treatments sampled were a no-burn control and a biennial winter burn. The last winter burn was applied approximately 4 months before sampling. The study was initiated in 1970 and the burn treatment had been subjected to five prior fires.

A Florida study located on the Osceola National Forest in north-central Florida flatwoods on level Coastal Plain terrain (McKee, 1982) has plots that are 0.81 ha in size with six replications in a block design. The soil on the study area is classed as Leon sand with an organic pan between 46 and 61 cm deep. The treatments consist of a no-burn control and an annual winter burn. The annual winter burn treatment was not imposed until 6 years after initiation of the study, and at the time of sampling there had been 14 burns.

Each plot in the three studies was sampled at 40 points by taking 10 sample cores composited to make four determinations per plot for each layer or horizon except on the South Carolina site, where all 40 sample points were kept separate. Two 2.5-cm soil cores were collected at each sample point and were segmented into 0- to 5-, 5- to 10-, 15- to 20-, 30- to 35-, and 45- to 50-cm depths on the South Carolina plots and 0- to 8-, 8- to 16-, and 32- to 40-cm depths on the Alabama and Florida plots. On the heavy South Carolina soils the 0- to 5- and 5- to 10-cm sample depths represent the A1 horizon and the 15- to 20-cm depth represents the A2 horizon. The 30- to 35-and 45- to 50-cm depths were selected to sample the B1 and upper B2 horizons. On the sandier soils in Florida and Alabama, the 0- to 8-cm sample represents the A1 horizon and the 8- to 16- and 32- to 40-cm depths represent the A2 horizon.

Soil samples were air-dried and crushed to pass through a 2-mm sieve. Total soil nitrogen was determined by micro-Kjeldahl digestion of a 1.0-g sample, and ammonia was determined by the salicylate–cyanurate method (Nelson and Somers, 1973). Results are expressed in concentration of nitrogen ($\mu g\ g^{-1}$) in the soil horizons.

The variance component ratios for these three experiments range from 0.043 to 1.284 (Table 1). Only 1 of the 11 variance ratios is > 1 which means that heterogeneity of nitrogen concentration in the profile is generally greater at the sample core level than aggregate variance of the groups of cores between plots. It is clear from Table 1 that plots within blocks are more alike in the Florida and Alabama studies both in terms of the absolute values of σ_s^2 and in that the relative variation of plots-in-blocks to that of determinations-in-plots is smaller (smaller ρ values).

Table 1—Unburned (control) and annually burned treatment differences in total nitrogen concentration ($\mu g\ g^{-1}$) are compared for three long-term prescribed burn studies.

Soil sample depth	Variance components: Plots (σ_b^2)	Variance components: Subsamples (σ_e^2)	Variance ratio (ϱ)	Treatment differences	F statistic
cm	($\mu g\ g^{-1}$)^2			$\mu g\ g^{-1}$	
		South Carolina			
0–5	512,863	399,504	1.284	109	0.06
5–10	83,929	127,801	0.657	199	1.14
15–20	8,556	35,168	0.243	27	0.19
30–35	26,521	32,023	0.828	50	0.23
45–50	4,508	24,906	0.181	12	0.06
		Florida			
0–8	3,092	71,248	0.043	288	6.12***
8–16	5,290	7,017	0.754	45	0.46
32–40	10,905	78,994	0.138	113	0.68
		Alabama			
0–8	5,241	33,791	0.155	125	3.90
8–16	1,257	20,902	0.060	107	7.58***
32–40	1,031	3,661	0.282	21	1.15

*** Statistically significant at the 90% confidence level.

AN APPLICATION

Estimated sample sizes for the number of determinations and number of blocks are presented in Table 2 using Δ equal to 100 $\mu g\ g^{-1}$, γ equal to 200, and Type-I (α) and Type-II (β) error rates of 0.1. The calculations show that all three designs are inadequate in at least one category. The Florida and Alabama studies have enough blocks but not enough determinations per plot. The estimated number of replications for the South Carolina data is prohibitively large and explains why a difference as large as 200 $\mu g\ g^{-1}$ (5- to 10-cm depth) was not statistically significant.

The following hypothetical example illustrates how this procedure could be used with variance component data to improve the experimental design. If blocks in the South Carolina study could have been constructed so that ρ for the 0- to 5-cm depth was reduced from 1.284 to 0.5 then the number of blocks would reduce from 727 to 293; a significant improvement, but still too large. Keeping ρ at 1.284 and compositing 20 cores reduces the number of blocks to 36 (727/20). The combined effect of more uniform blocks and compositing groups of 20 sample cores was to reduce the number of blocks from 727 to 15.

Fifteen blocks is still a large number relative to our usual three, four, or five but then this methodology is not being advanced as a panacea for our experimental design problems. Following the procedure will, however, force us to look more closely at facets of the experiment that significantly affect the usefulness of the results. If it turns out we cannot install the number of blocks called for, we will either have to sample larger composites, construct more uniform blocks, or accept a less sensitive experiment (that is, a larger Δ). Intuition suggests that there is likely a limit to the size of the composite, at which point the reduction in variance will begin to weaken from the assumed $1/r$ factor. Quantification of this effect is left to future applications. We believe the "averaging" assumption is sufficient as a first approximation.

Finally, when applying the results keep in mind the distinction between chemical determinations and sample cores. For example, in the above calculations for the 0- to 5-cm depth data in the South Carolina study, the number of sample cores increased from 12 to 400 per plot, but only 20 chemical determinations were made from the 400 cores.

STATISTICAL THEORY

In order to understand the details of the solution, it is necessary to be familiar with the research reported by Marcuse (1949). Although we leave this to the reader, we will outline

Table 2—Estimated numbers of blocks (b) and subsamples (n) for the three studies when $\alpha = 0.1$, $\beta = 0.1$, $\Delta = 100$, $\gamma = 200$, and the ϱ values from Table 1.

South Carolina			Florida			Alabama		
	Blocks	Sub-samples		Blocks	Sub-samples		Blocks	Sub-samples
Depth	(b)	(n)	Depth	(b)	(n)	Depth	(b)	(n)
cm	no.		cm	no.		cm	no.	
0–5	727	12	0–8	5	68	0–8	8	36
5–10	122	17	8–16	8	16	8–16	2	58
15–20	13	29	32–40	17	38	32–40	2	27
30–35	38	16	–	–	–	–	–	–
45–50	7	33	–	–	–	–	–	–

Marcuse's approach. She uses a minimization technique called Lagrange's multipliers to establish the combination of sample sizes for a 3-level, nested ANOVA that minimizes the cost of the experiment subject to the constraint that the variance of the overall mean is fixed. Our work differs from Marcuse's work in three ways. First, we have used the RCB design rather than a nested design. Second, we work with the variance of the difference between a pair of treatments, rather than the variance of the overall mean of the experiment. Third, and most importantly, we show how to select the target (or fixed) value for the variance so that we control the level of Type-II error.

The form of the RCB model we worked with is

$$Y_{ijk} = \beta_i + \mu_j + \delta_{ij} + \epsilon_{ijk}, \quad [1]$$

for ($i=1,2,\ldots,b$), ($j=1,2,\ldots,t$), and ($k=1,2,\ldots,n$), where β_i, δ_{ij}, and ϵ_{ijk} are random variables representing block, plots-in-block, and determinations-in-plot effects, respectively. The effects are assumed to be normally distributed with zero-valued means and covariances, and variances (or variance components) σ_β^2, σ_δ^2, and σ_ϵ^2. The parameters μ_j are unknown treatment constants. We are interested in testing the null hypothesis H_o:$\tau = 0$ against the alternative H_A:$\tau \geq \Delta$ where the parameter

$$\tau = \mu_j - \mu_{j'} \ (j \neq j') \quad [2]$$

for any pair of the treatments.

We use the random variable

$$T = \bar{y}_j - \bar{y}_{j'} \quad [3]$$

as the estimator of the treatment difference τ, where $\bar{y}_j$ is the observed mean nutrient concentration for the j^{th} treatment. The variance of the random variable T is

$$\mathrm{Var}(T) = 2\left\{\frac{\sigma_\delta^2}{b} + \frac{\sigma_\epsilon^2}{bn}\right\}. \quad [4]$$

Assuming the variance component parameters are known, Type-II error is defined as

$$\beta = \int_{-\infty}^{z_\alpha\sqrt{\mathrm{Var}(T)}} f(T)\,dT, \quad [5]$$

where, under the alternative hypothesis (H_A), $T \sim N[\Delta, \mathrm{Var}(T)] = f(T)$ and z_α equals the one-tailed value of a standard normal variable corresponding to the selected value of Type-I error (α). Equation (5) establishes a relationship between Type-II error (β) and Var(T), but unfortunately there is an infinite number of these relationships because a different one exists for every value of Δ. To eliminate this cumbersome problem we defined the random variable

$$G = T/\Delta, \quad [6]$$

which means that under H_A, $G \sim N(1,\theta) = g(G)$ and

$$\theta = \mathrm{Var}(T)/\Delta^2. \quad [7]$$

In this case Type-II error is defined as

$$\beta = \int_{-\infty}^{z_\alpha\sqrt{\theta}} g(G)\,dG. \quad [8]$$

This produces one relationship (for each α) of the variance factor θ over power of the test (1-β). These curves are presented in Fig. 3 for selected values of α. We use one-sided values of z_α on the premise that if the experimenter knows enough to specify Δ, then it is likely the expected ordering of treatments would also be known.

The pair of treatments for which Δ is chosen might be the two expected to be the least different or the two being the most different; or, the parameter Δ might simply be chosen as the difference that any pair of treatments must differ by to be of practical value. At this point the user-selected power (1-β) and Type-I error rate (α) is used in conjunction with Fig. 3 to determine the variance factor θ. Then θ and Δ are used to compute

$$\hat{\mathrm{Var}}(T) = \theta\Delta^2, \quad [9]$$

where $\hat{\mathrm{Var}}(T)$ is the value of the fixed variance used in Marcuse's minimization process.

The final step is minimization of the variable cost of the experiment

$$E = c_\delta b + c_\epsilon bn, \quad [10]$$

where c_δ and c_ϵ are the unit costs used in the previously defined γ ratio. The minimization is mathematically constrained using Lagrange's multipliers so that

$$2\left\{\frac{\sigma_\delta^2}{b} + \frac{\sigma_\epsilon^2}{bn}\right\} - \theta\Delta^2 = 0.$$

The solution estimates the number of blocks with

$$b = [\sigma_\epsilon^2/(\theta\Delta^2)]f, \quad [11]$$

where

$$f = 2[\rho + (\rho/\gamma)^{1/2}] \quad [12]$$

and the number of chemical determinations per plot with

$$n = (\gamma/\rho)^{1/2}. \quad [13]$$

So far in the development of this theory we have assumed the variance components (σ_β^2, σ_δ^2, and σ_ϵ^2) to be known. Steel and Torrie (1980, p. 118) present an adjustment factor ($df+3$)/($df+1$) to be used when variance is estimated, where df equals the error degrees of freedom. Incorporating this into our solution would require iterative approximations, a process we did not pursue on the premise that for the size experiments we anticipate, df would be large enough to render ($df+3$)/($df+1$) sufficiently near 1 to justify ignoring its effect.

REFERENCES

1. Nelson, D.W., and L.E. Somers. 1973. Determination of total nitrogen in plant material. Agron. J. 65:109–112.
2. Marcuse, S. 1949. Optimum allocation and variance components in nested sampling with an application to chemical analysis. Biometrics 5:189–206.
3. McKee, W.H., Jr. 1982. Changes in soil fertility following prescribed burning on Coastal Plain pine sites. USDA Forest Service, Southeastern Forest Experiment Station, Res. Paper SE-234.
4. Steel, R.G.D., and J.H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. McGraw-Hill Book Company, New York.