

A comparison of planned unbalanced designs for estimating heritability in perennial tree crops *

B. G. McCutchan^{1,**}, J. X. Ou² and G. Namkoong³

¹ Forestry Department, P.B. 8002, North Carolina State University, Raleigh, NC 27695-8002, USA

² South China College of Tropical Crops, Hainan Island, Guangdong Province, People's Republic of China

³ Southeastern Forest Experiment Station, USDA Forest Service, Genetics Department, P.B. 7614, North Carolina State University, Raleigh, NC 27695-7614, USA

Accepted July 19, 1985

Communicated by P. M. A. Tigerstedt

Summary. One hundred families of average size 10 are allocated in single-tree plots to 20 blocks in several planned unbalanced designs. Based on the variance of the estimate of heritability from the Minimum Variance Quadratic Unbiased Estimates and 100% survival, the three partially balanced designs are equally efficient. A design with variable family size is more efficient for heritabilities (h^2) generally less than 0.25; an equal family size design is more efficient for h^2 generally greater than 0.25. A design with a large number of small families is more efficient at high h^2 than that with a small number of large families; a design with fewer families of larger size is more efficient at low h^2 . Two-tree plot designs are never more efficient than single-tree plot designs and are also shown to be sensitive to the magnitude of the variance components that generate a given h^2 , whereas the single-tree plot designs are not.

Key words: Minimum Variance Quadratic Unbiased Estimates – Design efficiency – Plot size – Variable family size – Number of families

Introduction

Perennial tree crops present several experimental difficulties for the breeder who is interested in estimating variance components or functions of them, such as heritability. Not only do experiments consume large amounts of space, but the seeds are often difficult or expensive to obtain in reliable quantity. For these

reasons, Shrikhande (1957) developed a method for estimating genetic variances in coconut plantations of randomly dispersed genotypes without using parental information; this method was further developed for forest trees by Sakai and Hatakeyama (1963). Usanis (1972) concluded that the model produced inaccurate estimates with low precision in situations where environmental gradients prevailed; only random variation is accounted for by Shrikhande's model.

In this paper, we address some of the problems of space and variable family size in planted experiments. In particular, we consider that appropriately uniform test sites often exist only in small patches but that a large number of families must be included in such tests. In fact, one of the obstacles to the conduct of genetic experiments in tree crops is the large size of blocks needed for a randomized complete block design. The difficulties of either finding reasonably uniform sites or of including large site heterogeneity within blocks reduce the utility of such experimental plantings. It is often far easier to find several small planting sites than one large site of the same total area. In addition to the problem of space, we also consider that seeds collected from individual trees are often limited in number and may have germination problems such that family sizes are small and variable.

As a guide to the sizes of experiments which are useful with trees, we note in recent publications that around 50 parent trees are used in rubber (Tan 1977), and Oil Palm (Obisesan and Fatunla 1983) with a total of 1,500 to 3,100 individuals measured. Even though the number of families is small, this is roughly similar to forest tree experiments (Namkoong 1979). Some typical experiments such as in Western White Pine (Hanover and Barnes 1962) include only 28 families and 2,700 seedlings while some, as with Ponderosa Pine (Callaham and Duffield 1962), include 81 families and 4,300 seedlings. Larger experiments such as one open-pollinated study with Loblolly Pine with 280 families and 50

* Paper No. 9853 of the Journal series of the North Carolina Agricultural Research Service, Raleigh, NC, USA

** Present address: Westvaco, P.B. 1950, Summerville, SC 29484, USA

seedlings per family are usually too large to manage in complete block replications (Stoneypher 1966). Since the genetic variance estimate depends on the number of families included, it is desirable to include upwards of 100 families in these tests. It is also beneficial, for progeny testing, to include a large number of families when heritability and dominance variance are low (Pepper and Namkoong 1978). With the additional need to keep block size small, balanced designs will rarely be satisfactory to estimate variance components for tree experiments. Under such conditions, we require efficiency in use of materials as well as of space and the analytical procedures with unbalanced field designs (Namkoong 1981).

As a paradigm, we choose an experimental situation in which it is possible to collect a few viable seeds from each of 100 sampled trees. We assume that an average of 10 seedlings per parent tree can be planted and that all 1,000 trees will survive to measurement age. For the distribution of family size, we assume that seed viability varies from 2/3 to 1/4 and appropriate numbers of seed from 15 to 40, respectively, to generate binomial distributions of offspring with mean family size of 10. We also choose an extremely small area for each planting site, as is appropriate for many tree crops. We consider that 1/10 hectare blocks are easy to find and to manage such that survival within a block is easily maintained. We assume that each block can contain only 50 measurement trees laid out as single-tree plots and, hence, for a total experiment size of 1,000 trees, there are 20 such planting sites. These sites are random samples of the distribution within a site type. Our inferences from such a design will be limited to this site type.

Since variance component estimators derived from Analysis of Variance are inefficient with unbalanced data (Swallow and Monahan 1984; Searle 1979), and provide minimum variance estimators only with balanced data, we prefer to use more efficient estimators. Many unbalanced design comparisons are made on the basis of various forms of modified Least Squares procedures such as those by Muse et al. (1982) and others, as reviewed by Anderson (1981). For our analysis, we choose to use the Modified Maximum Likelihood (MML) procedure as detailed by Searle (1979) and programmed by Giesbrecht (1983). Estimates from this procedure are translation invariant, asymptotically normal (Brown 1976), free of fixed effects and have minimum variance in the locality of the true values. The MML estimates result from iterating the Minimum Norm Quadratic Unbiased Estimates (MINQUE) of variance components. We compare efficiency ratios of three different incomplete block designs, and of family size distribution over the full range of heritability on the basis of the variance of a given heritability. The efficiency of designs is expected to differ if evaluated with another comparison criterion, as shown by Muse and Anderson (1978). We also consider the efficiency of designs with fewer

families of larger size and two-tree plots, which are possible alternatives in the 20-block situation if the constraints on family size and number, and use of single-tree plots are relaxed. Muse and Anderson compare parameter estimation from several planned unbalanced designs in which the numbers of "blocks" and "families" are not constrained, as they have been here.

Methodology

The model

The following notation and computational methodology for Minimum Norm Quadratic Unbiased Estimates (MINQUE) are after Giesbrecht (1983), whose procedure for variance component estimation was written as a temporary SASTM program entitled Procedure MIXMOD.

The statistical model for each design considered is:

$$\mathbf{Y} = \mu \mathbf{1} + \mathbf{U}_B \mathbf{e}_B + \mathbf{U}_F \mathbf{e}_F + \mathbf{U}_P \mathbf{e}_P + \mathbf{e}_W$$

$n \times 1$ $n \times 1$ $n \times b \times b \times 1$ $n \times f \times f \times 1$ $n \times s \times s \times 1$ $n \times 1$

where \mathbf{Y} is the column vector of observations; μ is the overall mean; \mathbf{U}_B , \mathbf{U}_F and \mathbf{U}_P are design matrices pertaining to the block, family and plot (with s block-by-family combinations) effects, respectively, with all elements equal to zero or one; for single-tree plots \mathbf{U}_P is the identity matrix; \mathbf{e}_B , \mathbf{e}_F , \mathbf{e}_P and \mathbf{e}_W are independent column vectors of independent random variables, each with mean zero and variance-covariance matrix $\mathbf{I}_b \sigma_B^2$, $\mathbf{I}_f \sigma_F^2$, $\mathbf{I}_s \sigma_P^2$ and $\mathbf{I}_n \sigma_W^2$, respectively. The variance-covariance matrix of \mathbf{Y} is:

$$\mathbf{V}(\mathbf{Y}) = \mathbf{U}_B \mathbf{U}_B' \sigma_B^2 + \mathbf{U}_F \mathbf{U}_F' \sigma_F^2 + \mathbf{U}_P \mathbf{U}_P' \sigma_P^2 + \mathbf{I}_n \sigma_W^2,$$

where σ_B^2 , σ_F^2 , σ_P^2 , and σ_W^2 are the variance components due to the block, family, plot and within plot error effects, respectively. Letting $\mathbf{V}_i = \mathbf{U}_i \mathbf{U}_i'$, and for convenience $\mathbf{V}_W = \mathbf{I}_n$, $\mathbf{V}(\mathbf{Y})$, based on the parameters, can be rewritten as:

$$\mathbf{V}_{\sigma^2} = \mathbf{V}(\mathbf{Y}) = \mathbf{V}_B \sigma_B^2 + \mathbf{V}_F \sigma_F^2 + \mathbf{V}_P \sigma_P^2 + \mathbf{V}_W \sigma_W^2. \quad (1)$$

When the variance components of (1) are unknown, MINQUE of them can be found as solutions to the equations:

$$\{\text{tr}(\mathbf{Q}_\alpha \mathbf{V}_i \mathbf{Q}_\alpha \mathbf{V}_j)\} \{\hat{\sigma}_i^2\} = \{\mathbf{Y}' \mathbf{Q}_\alpha \mathbf{V}_i \mathbf{Q}_\alpha \mathbf{Y}\} \quad i, j = B, F, P, W \quad (2)$$

where $\mathbf{Q}_\alpha = \mathbf{V}_\alpha^{-1} - \mathbf{V}_\alpha^{-1} \mathbf{1} (\mathbf{1}' \mathbf{V}_\alpha^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{V}_\alpha^{-1}$, \mathbf{V}_α is \mathbf{V}_{σ^2} in (1) with the true values (σ_i^2) replaced with prior values (α_i), $\mathbf{1}$ is a $n \times 1$ vector and the symbol $\{\mathbf{a}_k\}$ refers to a matrix whose elements are a_k . Iteration of MINQUE (I-MINQUE) to convergence results in Modified Maximum Likelihood (also referred to as Restricted Maximum Likelihood (REML)) estimates, assuming all the random elements are normally distributed and no negative estimates are encountered in the iteration process.

If the variance components (σ_i^2) were known, then (2) with \mathbf{V}_{σ^2} of (1) replacing \mathbf{V}_α in \mathbf{Q}_α (now \mathbf{Q}_{σ^2}) represents equations for Minimum Variance Quadratic Unbiased Estimates (MIVQUE) (assuming normality). The variance-covariance matrix for the resulting estimates of the components ($\sigma_B^2, \sigma_F^2, \sigma_P^2, \sigma_W^2$)' would be:

$$2 \{\text{tr}(\mathbf{Q}_{\sigma^2} \mathbf{V}_i \mathbf{Q}_{\sigma^2} \mathbf{V}_j)\}^{-1} \quad i, j = B, F, P, W. \quad (3)$$

The dispersion matrix (3) of the variance components is a function of the variance components themselves plus the design matrices (\mathbf{U}_i). It is therefore possible to calculate this dispersion matrix for a given true set of variance component values and a design.

For the calculation of heritability (h^2), we assume that the families are half-sibs and that $4\sigma_F^2 = \sigma_A^2$, where σ_A^2 is the additive genetic variance. Heritability is calculated as:

$$h^2 = 4\sigma_F^2 / (\sigma_F^2 + \sigma_P^2 + \sigma_W^2) = X/Y.$$

Using the variances and covariances from the dispersion matrix (3), we calculate the variance of heritability using a Taylor's series approximation for the variance of a ratio:

$$\text{var}(\hat{h}^2) \approx \text{var}(\hat{X})(1/Y)^2 - 2\text{cov}(\hat{X}, \hat{Y})(1/Y)(X/Y^2) + \text{var}(\hat{Y})(X/Y^2)^2,$$

where values X and Y are based on true values of variance components. Sets of variance components are evaluated for $0 < h^2 \leq 1$ and standardized such that $\sigma_W^2 = 1$ which requires $\sigma_F^2 \leq 1/3$ for $\sigma_P^2 = 0$. We examine two types of the plot error component (σ_P^2): $\sigma_F^2 = \sigma_P^2$ and $\sigma_F^2 > \sigma_P^2 = 0.005$, each over the range of h^2 for a constant $\sigma_B^2 = 2$. These types of variance component sets are compared by an efficiency ratio (E_i) which equals the $\text{var}(\hat{h}^2)$ based on the variance component set type $\sigma_F^2 > \sigma_P^2$ divided by the $\text{var}(\hat{h}^2)$ based on the variance component set type $\sigma_F^2 = \sigma_P^2$ for a given h^2 . The variance of heritability is scale invariant such that the results are generally applicable. Since σ_B^2 affects the variance of other variance components through its inclusion in (3), we examine its effect

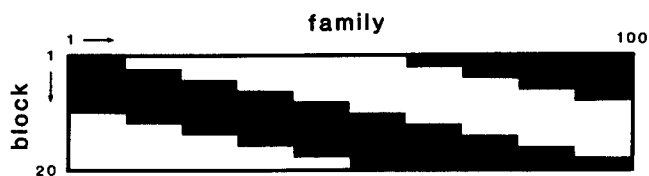
on the magnitude of $\text{var}(\hat{h}^2)$ at $\sigma_B^2 = 0.5$, $\sigma_B^2 = 2$ and $\sigma_B^2 = 4$ by an efficiency ratio (E_i) equal to the $\text{var}(\hat{h}^2)$ based on $\sigma_B^2 = i$ divided by the $\text{var}(\hat{h}^2)$ based on $\sigma_B^2 = 2$, for $i = 0.5, 4$ and a given h^2 .

The field designs and family size distributions

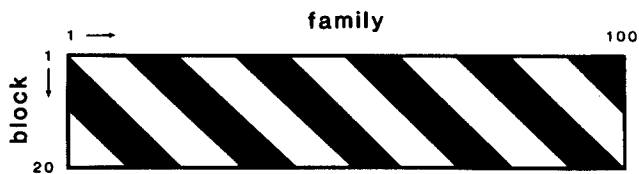
The unbalanced designs examined for allocating 1,000 trees to 2,000 block-family combinations were chosen to fit the constraints of 20 blocks with 50 trees per block and 100 families with an average of 10 trees per family. These designs then differ in the way families are allocated to blocks.

The general forms of these design allocations are a parallel design, a diagonal design and a uniform design (Fig. 1). Upon rearranging the order of the families and the blocks, all of these allocations can be viewed as incomplete block designs with varying patterns of family commonality among blocks (Fig. 2). For the case of ten trees per family, the diagonal design is a series of 10-block by 10-family units with 8 common blocks among "adjacent" units. The parallel design is a series of 10-block by 5-family units with 9 common blocks among "adjacent" units. The uniform design is a series of two 10-block by 50-family units without any common blocks among the units, referred to by Gaylor and Anderson (1960)

DIAGONAL DESIGN



PARALLEL DESIGN



UNIFORM DESIGN

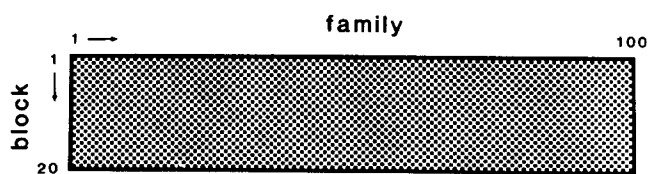
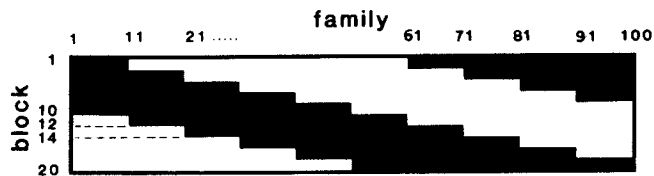
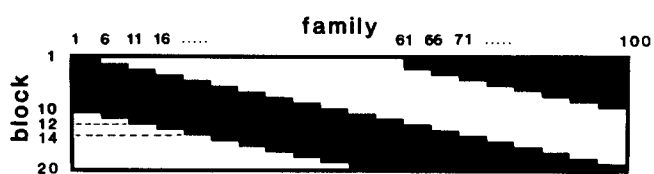


Fig. 1. Schematic diagram of the diagonal, parallel and uniform design allocations, where a shaded block-family combination indicates that a single tree from a given family is randomly allocated within a given block. Each of the 20 blocks has 50 trees, the total number of trees being 1,000. An uniformly ten trees per family distribution is illustrated here

DIAGONAL DESIGN



PARALLEL DESIGN



UNIFORM DESIGN

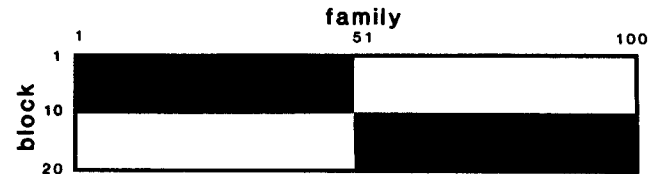


Fig. 2. Schematic diagram of the diagonal, parallel and uniform design allocations, with the order of families and/or blocks rearranged to show the commonality of the families among blocks. Each shaded block-family combination indicates that a single tree from a given family is randomly allocated within a given block. Each of the 20 blocks contains 20 trees and $n = 1,000$. For this diagram, each family has 10 trees

as a disjoint rectangles design. These patterns are complicated with the imposition of family size distribution.

Five distributions were considered for family size: one with a uniform 10 trees per family (U(10)), and four binomial distribution (B(N,p)): B(15,2/3), B(20,1/2), B(30,1/3) and B(40,1/4). For all distributions, the number of trees per plot is either zero or one, for a total of 1,000 filled cells. Since σ_w^2 is confounded with σ_b^2 , only the sum ($\sigma_b^2 + \sigma_w^2$) rather than the individual components is estimable. For calculating $\text{var}(\hat{h}^2)$, only the variance of the sum is required.

Given the same set of environmental effects with 20 blocks, we can also consider the use of 50 families each with 20 trees. In addition, we can compare the efficiency of using two-tree plots (TTP) designs with that of single-tree plots (STP) designs. For these comparisons we examine designs with single-tree and two-tree plots, when we carry 50 families with 20 trees per family, and when we carry 100 families with 10 trees per family. We assume a uniform family size distribution and choose the following designs to fill the four cells of this two-by-two comparison table:

B.20.50.1 A Balanced design with 20 blocks, 50 families using single (1) tree plots.

U.20.50.2 An Unbalanced diagonal design with 20 blocks, 50 families using 2 tree plots. These are arranged as a series of balanced units having common blocks among "adjacent" units: 10 blocks by 10 families having 8 common blocks with the "adjacent" 10-block by 10-family unit, which has 8 common blocks with its adjacent 10-block by 5-family unit, which has 4 common blocks with the following 10-block by 10-family unit ... and the pattern repeats.

U.20.100.1 An Unbalanced diagonal design with 20 blocks, 100 families and single (1) tree plots. This is the diagonal design with a U(10) family size distribution.

U.20.100.2 An Unbalanced diagonal design with 20 blocks, 100 families and 2 tree plots. These are arranged as a series of balanced units having common blocks among "adjacent" units: 5 blocks by 10 families having 3 common blocks with the "adjacent" 5-block by 10-family unit, which itself has 3 common blocks with its "adjacent" unit of 5 blocks by 5 families which then has 4 common blocks with the following 5-block by 10-family unit ... and the pattern repeats 3 more times.

For the two-tree plot designs, U.20.50.2 and U.20.100.2, σ_b^2 and σ_w^2 are estimable separately.

The criterion for comparing design allocations (family size distribution) is an efficiency ratio of the $\text{var}(\hat{h}^2)$ from one design allocation (family size distribution) divided by that from another, for a given h^2 and family size distribution (design allocation). For balanced designs, the MML estimates are ANOVA estimates but not otherwise. The $\text{var}(\hat{h}^2)$ for B.20.50.1 based on MIVQUE is the same as that based on the ANOVA procedure.

Design efficiency profiles (Namkoong and Roberds 1974) are drawn to map the variance of the estimate of heritability ($\text{var}(\hat{h}^2)$) for a particular design allocation-family size distribution combination over the range of heritability. Thus, the efficiency of each design relative to a standard criterion can be compared for any range of heritabilities. One such criterion profile might simply be that based on a constant coefficient of variation ($CV(\hat{h}^2) = (\sqrt{\text{var}(\hat{h}^2)}/h^2) \times 100\%$). Since very low heritabilities would require very high precision if such a criterion is adopted, Namkoong and Roberds (1974) suggest that a composite criterion be adopted, using a constant CV at moderate to high heritabilities and a constant standard error at low heritabilities. We compare our designs against a

$\text{var}(\hat{h}^2)$ profile based on a CV = 50% curve for $h^2 = 0.20$ to 1.0; while for $0 < h^2 \leq 0.20$ we have drawn the profile based on $\text{std}(\hat{h}^2) = 0.10$.

Results and discussion

The $\text{var}(\hat{h}^2)$ profile of U.20.100.1, typical of the form and magnitude found with other design allocations and family size distributions, is shown in Fig. 3. The standard curve allows us to readily see that the unbalanced design has a $\text{var}(\hat{h}^2)$ which is always less than the criterion suggested.

Comparison of design allocations is based on efficiency ratios of each of the parallel and uniform designs to the diagonal design for a given family size distribution (B(30,1/3)) at selected heritabilities (Table 1). The general pattern seen for the B(30,1/3) distribution is found with the other distributions. There is a consistent difference of small magnitude (less than 1%) among designs across h^2 . The diagonal design is the most efficient at high heritabilities and the uniform, the least. However, at $h^2 \leq 0.33$, the

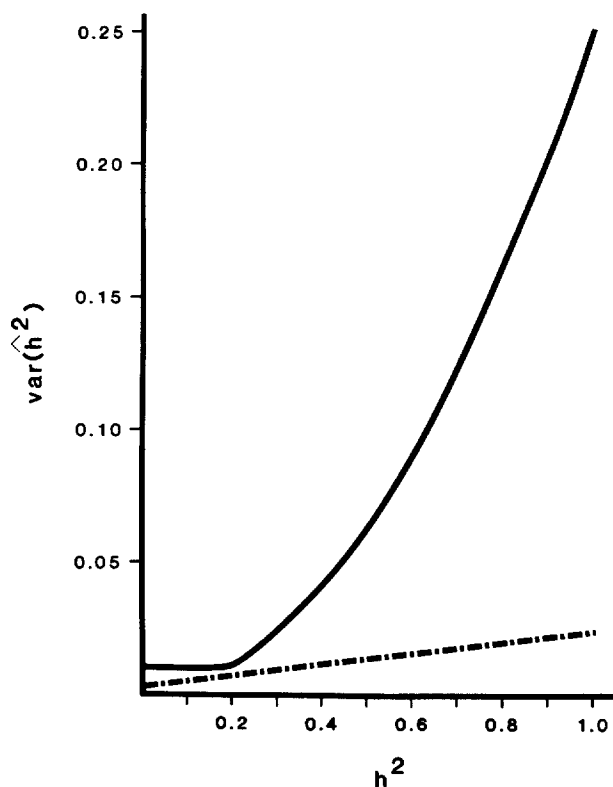


Fig. 3. $\text{var}(\hat{h}^2)$ profile for U.20.100.1 with ten trees per family, given $\sigma_b^2 = 2$, $\sigma_b^2 = \sigma_b^2$ and $\sigma_w^2 = 1$ (broken line) compared with a standard criterion curve (solid line), which is based on a CV(\hat{h}^2) = 50% for $h^2 > 0.2$ and on a $\text{std}(\hat{h}^2) = 0.1$ for $h^2 < 0.2$

uniform design is the most efficient, and at $h^2 \leq 0.26$, the diagonal allocation is the least efficient. The difference between the design allocations is the number of families in the block by family units and the number of common blocks among "adjacent" units. For the U (10) family distribution, the diagonal has units of 10 families having 8 blocks common among "adjacent" units, the parallel has units of 5 families with 9 blocks common, and the uniform has units of 50 without common blocks among "adjacent" units. Neither the effect of the number of common blocks nor the number of families in a unit corresponds with the efficiency rating. How these two factors influence $\text{var}(\hat{h}^2)$ such that we see a change in design efficiency at $h^2 \sim 1/3$ is

Table 1. Design efficiency ratios as affected by design allocation. The ratio is of the $\text{var}(\hat{h}^2)$ from each of the parallel and uniform designs to that from the diagonal design, illustrated using a B(30, 1/3) family size distribution, $\sigma_B^2 = 2$, $\sigma_W^2 = 1$ and $\sigma_F^2 = \sigma_P^2$

$\sigma_F^2 (= \sigma_P^2)$	h^2	Diagonal (D) $\text{var}(\hat{h}^2)$	Parallel (P) E_P^a	Uniform (U) E_U^b
0.5	1.00	0.0219	1.00015	1.00185
0.3	0.75	0.0176	1.00014	1.00142
0.2	0.57	0.0142	1.00013	1.00092
0.15	0.46	0.0120	1.00010	1.00047
0.10	0.33	0.0095	1.00004	0.99979
0.09	0.31	0.0089	1.00002	0.99962
0.075	0.26	0.0080	0.99999	0.99932
0.05	0.18	0.0066	0.99992	0.99869
0.02	0.08	0.0047	0.99978	0.99765
0.01	0.04	0.0040	0.99984	0.99731
0.005	0.02	0.0037	0.99968	0.99694

^a E_P is the $\text{var}(\hat{h}^2)$ from the parallel (P) design divided by the $\text{var}(\hat{h}^2)$ from the diagonal design

^b E_U is the $\text{var}(\hat{h}^2)$ from the uniform (U) design divided by the $\text{var}(\hat{h}^2)$ from the diagonal design

unclear. Swallow and Searle (1978) note that no one design allocation uniformly minimizes the variance of their estimate. In all cases, however, the differences among designs are less than 1% and these allocations are essentially equivalent. Previous knowledge of h^2 is not required to choose an efficient design allocation.

The effects of family size distributions on magnitude of $\text{var}(\hat{h}^2)$ is shown in Table 2 for the diagonal design based on an efficiency ratio of the $\text{var}(\hat{h}^2)$ from a given distribution to that from the U(10) distribution. There is a larger magnitude of differences among designs as affected by family size distributions (up to 6%) than by design allocations. However, these differences are not very large. There is a consistent trend of $\text{var}(\hat{h}^2)$ with variance of the family size distribution (equal to $Np(1-p)$ for a binomial distribution) for a given h^2 . For $h^2 \geq 0.25$ the U(10) design is the most efficient and the designs with variance of family size are less efficient. For $h^2 \leq 0.21$ the most variable family size design is the most efficient and the uniformly 10 trees per family design is the least efficient. Similar results are found for the other design allocations; the most variable family size design is the most efficient at $h^2 \leq 0.18$ for the uniform design. A variable family size distribution which is usually considered undesirable, can be beneficial for estimating h^2 when h^2 is in its lower range.

In assessing design allocations and family size distribution together, we find that the diagonal design with U(10) distribution is the most efficient at high heritabilities and that the uniform design with B(40, 1/4) distribution is most efficient at low heritabilities. The magnitude of these differences suggest that one design or distribution is essentially equivalent across the range of h^2 . However, the trend in the differences suggests that creating a large family variance for

Table 2. Design efficiency ratios as affected by family size distribution. The ratio is the $\text{var}(\hat{h}^2)$ from each binomial distribution (B(N, p)) to that from the uniformly ten trees per family distribution (U (10)), illustrated using the diagonal design with $\sigma_B^2 = 2$, $\sigma_F^2 = \sigma_P^2$ and $\sigma_W^2 = 1$

$\sigma_F^2 (= \sigma_P^2)$	h^2	U (10) $\text{var}(\hat{h}^2)$	B (15, 2/3) E_{15}^b	B (20, 1/2) E_{20}	B (30, 1/3) E_{30}	B (40, 1/4) ^a E_{40}
0.5	1.00	0.0215	1.010	1.015	1.020	1.022
0.3	0.75	0.0173	1.011	1.016	1.021	1.023
0.2	0.57	0.0139	1.010	1.015	1.020	1.021
0.1	0.33	0.0094	1.006	1.009	1.011	1.012
0.07	0.25	0.0077	1.002	1.003	1.003	1.004
0.06	0.21	0.0072	1.000	1.000	0.999	0.999
0.05	0.18	0.0066	0.997	0.996	0.994	0.994
0.02	0.08	0.0048	0.983	0.976	0.969	0.966
0.005	0.02	0.0039	0.971	0.958	0.946	0.941

^a Distributions are arranged in increasing order of family size variance: 0, 3 1/3, 5, 6 2/3, 7 1/2 for the U (10), B (15, 2/3), B (20, 1/2), B (30, 1/3) and B (40, 1/4) distributions, respectively

^b E_N is the $\text{var}(\hat{h}^2)$ from the B (N, p) design divided by the $\text{var}(\hat{h}^2)$ from the U (10) design

Table 3. Design efficiency ratios as affected by the block variance component value. The ratio is of the var(\hat{h}^2) based on each $\sigma_B^2 = 0.5$ and $\sigma_B^2 = 4$ to that based on $\sigma_B^2 = 2$, given the diagonal design with B(30, 1/3) family size distribution and $\sigma_W^2 = 1$

$\sigma_F^2 (= \sigma_B^2)$	h^2	$\sigma_B^2 = 2$ var(\hat{h}^2)	$\sigma_B^2 = 0.5$ $E_{0.5}^a$	$\sigma_B^2 = 4$ E_4^b
0.5	1.00	0.0219	0.9997	1.00006
0.2	0.57	0.0142	0.9997	1.00007
0.1	0.33	0.0095	0.9996	1.00007
0.08	0.28	0.0083	0.9996	1.00006
0.05	0.18	0.0066	0.9996	1.00007
0.04	0.15	0.0059	0.9996	1.00007
0.03	0.11	0.0053	0.9996	1.00007
0.02	0.08	0.0047	0.9996	1.00007

^a $E_{0.5}$ is the var(\hat{h}^2) based on $\sigma_B^2 = 0.5$ divided by the var(\hat{h}^2) based on $\sigma_B^2 = 2$

^b E_4 is the var(\hat{h}^2) based on $\sigma_B^2 = 4$ divided by the var(\hat{h}^2) based on $\sigma_B^2 = 2$

Table 4 a. Design efficiency ratios as affected by plot size. The ratio is of the var(\hat{h}^2) based on a two-tree plot (TTP) design to that of a single-tree plot (STP) design for $\sigma_B^2 = 2$, $\sigma_F^2 = \sigma_B^2$ and $\sigma_W^2 = 1$

$\sigma_F^2 (= \sigma_B^2)$	h^2	Single-tree plots var(\hat{h}^2)	Two-tree plots $E_{TTP} = \frac{\text{var}(\hat{h}^2)_{TTP}}{\text{var}(\hat{h}^2)_{STP}}$
Fifty families each with 20 trees	0.5	1.00	0.0320
	0.1	0.33	0.0096
	0.05	0.18	0.0054
	0.005	0.02	0.0020
One hundred families each with 10 trees	0.5	1.00	0.0215
	0.1	0.33	0.0094
	0.05	0.18	0.0066
	0.005	0.02	0.0039

Table 4 b. Design efficiency ratios as affected by the number of families. The ratio is of the var(\hat{h}^2) from a 50-family design divided by that from a 100-family design, given $\sigma_B^2 = 2$, $\sigma_F^2 = \sigma_B^2$ and $\sigma_W^2 = 1$

$\sigma_F^2 (= \sigma_B^2)$	h^2	100 families var(\hat{h}^2)	50 families $E_{50} = \frac{\text{var}(\hat{h}^2)_{50}}{\text{var}(\hat{h}^2)_{100}}$
Single-tree plots	0.5	1.00	0.0215
	0.1	0.33	0.0094
	0.05	0.18	0.0066
	0.005	0.02	0.0039
Two-tree plots	0.5	1.00	0.0265
	0.1	0.33	0.0111
	0.05	0.18	0.0077
	0.005	0.02	0.0046

estimating heritabilities can be very useful for estimating h^2 at its lower values.

The effect of the block variance component value on var(\hat{h}^2) can be seen in Table 3. The var(\hat{h}^2) increases with the magnitude of the block variance component. We have used $\sigma_B^2 = 2$ throughout the study and see that had we used a higher or lower value, that we would have shifted the var(\hat{h}^2) profile higher or lower, ever so slightly.

In the second part of this study we have removed the constraint of 100 families with 10 trees apiece and single-tree plots (STP) to consider the efficiency of designs with 50 families of 20 trees apiece and two-tree plots (TTP). In Table 4a the efficiency of TTP designs is compared with that of STP designs for each 50 and 100 families. For a constant number of trees per family, it is always more efficient to put out STP designs than TTP designs, in particular for high heritabilities and with 100 families. Differences in STP and TTP designs that may explain the results are that there are fewer combinations of blocks and families with TTP (500) in contrast to STP (1,000); also with 10 trees for each of 100 families, TTP designs reduce the number of plots per family down to 5 compared to 10 with STP designs. Gaylor and Anderson (1960) showed that the variance of “ σ_F^2 ” term is minimized with one observation per plot, based on the method of fitting constants.

In general, it is clear that a design with a greater number of small families is more efficient at high heritabilities compared to heritabilities generally less than one third, where it is more efficient to have a design with fewer but larger families (Table 4b). A design with few large families rather than many small families, in an ANOVA sense is equivalent to allocating a larger number of degrees of freedom to the σ_B^2 (and also σ_W^2 in a design with STP) terms. For small heritabilities, these are the dominant terms. This comparison sheds light on the effect of family size variation. A high variance in family size indicates that there are some families with more offspring, some with respectively fewer. The response of variable family size designs at low h^2 is asymmetric with increased efficiency from the larger families outweighing the loss in efficiency from the smaller families. At high heritabilities, a design with a large number of small families is required to allocate sampling to the numerator ($4 \sigma_F^2$).

The choice of family number for h^2 estimation depends on whether we consider only a portion of the range of h^2 or consider the entire range. The former case involves little difficulty in choosing an appropriate design. The latter case, however, requires deeper consideration and is of interest when there is a lack of prior knowledge of h^2 or when the heritabilities of several traits, which may occur at various points across

the range of h^2 , are to be estimated simultaneously. This would require a design with a large number of families to obtain low variance estimates of h^2 at high h^2 values as well as a large variance on the family size for heritabilities in the lower range.

The two types of variance component sets are evaluated by an efficiency ratio (E_t , "t" for types) and are shown for the factorial arrangement of number of families and plot size in Table 5. Efficiency ratios are also calculated for the design allocations and family size distributions mentioned earlier in this study. Their results are of the 100-family STP design type (U.20.100.1). Single-tree plot designs, regardless of the family size are generally insensitive to the magnitude of the different types of variance component sets used in calculating a given h^2 . Design recommendations can then be based on h^2 itself without requiring knowledge of the actual components. This situation is in contrast with TTP designs, and in general with multiple-tree plot designs, particularly at high heritabilities. Efficiency ratios for TTP designs show up to 16% difference between $\text{var}(\hat{h}^2)$ based on the two variance component set types, with the $\text{var}(\hat{h}^2)$ based on $\sigma_F^2 > \sigma_P^2$ type being consistently less than that based on $\sigma_F^2 = \sigma_P^2$ type. For the allocations examined here, we observe only changes in magnitude, in contrast to that of ranking of the designs. Simplicity of design recommendation is obviously gained if decisions can be based upon h^2 itself rather than the magnitude of the variance components. In Table 6a the effect of a decreased $\text{var}(\hat{h}^2)$ for $\sigma_F^2 > \sigma_P^2$ variance component set type with TTP designs compared with the $\sigma_F^2 > \sigma_P^2$ type can be seen in E_{TTP} . There is a decrease in the design efficiency ratios compared with the analogous ratios in Table 4a. The $\text{var}(\hat{h}^2)$ for the STP design remains essentially the same over variance component set types. The sensitivity of TTP designs to types is particularly noticed for the 100-family design. The 100-family design is more sensitive to the use of TTP compared to STP than the 50-family design (Table 4a).

In Table 6b, the STP design information is the same as in Table 4b. The E_{TTP} ratios have increased compared with those in Table 4b, reflecting the greater sensitivity of the 100-family design compared to the 50-family design employing TTP. Each $\text{var}(\hat{h}^2)$ decreased; however, the $\text{var}(\hat{h}^2)$ due to the 100-family design decreased more than that with the 50-family design.

A natural extension of this research is to examine the robustness of designs to unplanned loss, as is typical of perennial crops and to consider alternative block size and numbers. Of particular interest will be the robustness of a family size distribution with high variance of family size. The calculations we made lead us to suspect that by increasing the variance of family size even higher, we might be able to get lower $\text{var}(\hat{h}^2)$ at low h^2 than that with a lower variance of family

Table 5. Efficiency ratios (E_t) equal to the $\text{var}(\hat{h}^2)$ based on $\sigma_F^2 > \sigma_P^2 = 0.005$ divided by that based on $\sigma_F^2 = \sigma_P^2$, given $\sigma_B^2 = 2$, $\sigma_W^2 = 1$ and a diagonal design allocation for the unbalanced designs

	h^2	Single-tree plots	Two-tree plots
Fifty families each with 20 trees	1.00	1.00000 ^a	0.90781 ^b
	0.33	0.99996	0.93814
	0.18	1.00005	0.95594
One hundred families each with 10 trees	1.00	1.00004 ^c	0.83764 ^d
	0.33	0.99999	0.90836
	0.18	1.00003	0.94147

- ^a This design information is based on B.20.50.1
- ^b This design information is based on U.20.50.2
- ^c This design information is based on U.20.100.1
- ^d This design information is based on U.20.100.2

Table 6a. Design efficiency ratios as affected by plot size. The ratio is of the $\text{var}(\hat{h}^2)$ based on a two-tree plot (TTP) design to that of a single-tree plot (STP) design for $\sigma_B^2 = 2$, $\sigma_P^2 = 0.005$ and $\sigma_W^2 = 1$

	σ_F^2	h^2	Single-tree plots $\text{var}(\hat{h}^2)$	Two-tree plots $E_{TTP} = \frac{\text{var}(\hat{h}^2)_{TTP}}{\text{var}(\hat{h}^2)_{STP}}$
Fifty families each with 20 trees	0.335	1.00	0.0320	1.014
	0.091	0.33	0.0096	1.032
	0.048	0.18	0.0054	1.046
	0.005	0.02	0.0020	1.088
One hundred families each with 10 trees	0.335	1.00	0.0215	1.035
	0.091	0.33	0.0094	1.073
	0.048	0.18	0.0066	1.099
	0.005	0.02	0.0039	1.157

Table 6b. Design efficiency ratios as affected by the number of families. The ratio is of the $\text{var}(\hat{h}^2)$ from a 50-family design to that from a 100-family design, given $\sigma_B^2 = 2$, $\sigma_P^2 = 0.005$ and $\sigma_W^2 = 1$

	σ_F^2	h^2	100 families $\text{var}(\hat{h}^2)$	50 families $E_{50} = \frac{\text{var}(\hat{h}^2)_{50}}{\text{var}(\hat{h}^2)_{100}}$
Single-tree plots	0.335	1.00	0.0215	1.487
	0.091	0.33	0.0094	1.030
	0.048	0.18	0.0066	0.825
	0.005	0.02	0.0039	0.516
Two-tree plots	0.335	1.00	0.0222	1.457
	0.091	0.33	0.0100	0.991
	0.048	0.18	0.0072	0.785
	0.005	0.02	0.0045	0.486

size. However, in a design with high variance of family size, loss of trees may begin to reduce the number of families since there will be some families with only a few trees.

Conclusions

We have considered an estimation situation where the total number of observations is 1,000, and the number of blocks is restricted to 20, each of size 50 trees employing single-tree plots (STP). The objective is to estimate heritability. One hundred families of average size 10 are allocated to these blocks. Given these small blocks, we can design efficient h^2 estimation experiments. We find that the tested patterns of allocating families to blocks is inconsequential. Little difference, as measured by efficiency ratios of the variance of the estimate of heritability, exists among patterns of family associations in and across blocks. We considered designs having variance in family size and found a small but consistent trend. Equal size family designs are most efficient at high heritabilities. A design with a large family size variance (our largest is $7\frac{1}{2}$) is most efficient at low heritabilities.

Relaxing the constraint of 100 families and STP, we find that STP designs are more efficient than two-tree plots (TTP) designs across the range of h^2 , regardless of the number of families. Single-tree plot designs are generally insensitive to the variance component set type used for calculating $\text{var}(\hat{h}^2)$ at a given h^2 , whereas TTP designs are sensitive to the type used. The number and size of families is influential on design efficiency. In general, a design with a large number of small families is desirable for high heritabilities, whereas a design with fewer families of larger size is more efficient at lower heritabilities.

The design of an experiment to estimate h^2 is sensitive to the number and size of families, the variance of family size and the plot size. These results substantiate similar work by Haile and Webster (1975), who examined unbalanced block designs for estimating a row variance component (in a two-way classified design) and concluded that the choice of design is of minor importance. The correct choice of the number of levels for the effects, given a particular magnitude of the components, is where gain, in terms of reducing the variance of the estimator, can be made. Based on the allocations studied and 100% survival of trees, we recommend designs with STP to gain stability over variance component set types as well as efficiency, and the consideration of a large number of families with a high family size variance to provide efficient estimators of h^2 across its range.

The criterion for efficiency ratios has been $\text{var}(\hat{h}^2)$, where h^2 is based on individual performance. De-

pending upon the mating and experimental designs employed, criteria other than h^2 might be preferred. For example, in a breeding program utilizing family selection, the heritability derived with family means (\hat{h}_f^2) is a more appropriate criterion than h^2 . Design comparisons based on $\text{var}(\hat{h}_f^2)$ can be made, in an analysis analogous to ours, where $\hat{h}_f^2 = \sigma_f^2 / \sigma_p^2$ (σ_p^2 being the phenotypic variance among family means).

Acknowledgements. We gratefully acknowledge Dr. Francis Giesbrecht for his guidance and the Statistics Department at North Carolina State University, Raleigh, for providing the computing funds necessary for this study. We extend our appreciation to Ronnie Hise for his drawing of the graphics.

References

- Anderson RL (1981) Recent developments in designs and estimators for variance components. In: Csorgo M, Dawson DA, Rao JNK, Saleh AKMdE (eds) Tech Rep No 118, University of Kentucky. North-Holland, Amsterdam, pp 1–45
- Brown KG (1976) Asymptotic behavior of MINQUE-type estimators of variance components. *Ann Stat* 4:746–754
- Callahan RZ, Duffield JW (1962) Heights of selected ponderosa pine seedlings during 20 years. In: Proc Forest Genet Workshop. Publ No 22, Soc Forest Tree Improv Comm, pp 10–13
- Gaylor, DW, Anderson RL (1960) The construction and evaluation of some designs for the estimation of parameters in random models. Institute of Statistics Mimeo Series No 256, NC State University, Raleigh NC, 88 pp
- Giesbrecht FG (1983) An efficient procedure for computing MINQUE of variance components and generalized least squares estimates of fixed effects. *Commun Stat – Theor Meth* 12:2168–2177
- Haile A, Webster JT (1975) A comparison of four designs for estimating the variances of random effects in a two-way classification. *Technometrics* 17:495–500
- Hanover, JW, Barnes BV (1962) Heritability of height growth in year-old western white pine. In: Proc Forest Genet Workshop. Publ No 22 Soc Forest Tree Improv Comm, pp 71–76
- Muse HD, Anderson RL, Thitakamol B (1982) Additional comparisons of designs to estimate variance components in a two-way classification model. *Commun Stat – Theor Meth* 11:1403–1425
- Muse HD, Anderson RL (1978) Comparison of designs to estimate variance components in a two-way classification model. *Technometrics* 20:159–166
- Namkoong G (1979) Introduction to quantitative genetics in forestry. US Dept Agric Tech Bull 1588. Castle House Publ, Kent, 342 pp
- Namkoong G (1981) Variance component estimation. In: Proc 17th IUFRO World Congr. Jpn Div 6:149–159
- Namkoong G, Roberds JH (1974) Choosing mating designs to efficiently estimate genetic variance components for trees. I. Sampling errors of standard analysis of variance estimators. *Silvae Genet* 23:45–53
- Obisesan IO, Fatunla T (1983) Genotype X environment interaction for bunch yield and its components in the oil palm (*Elaeis guineensis*, Jacq.). *Theor Appl Genet* 64:133–136

- Pepper WD, Namkoong G (1978) Comparing efficiency of balanced mating designs for progeny testing. *Silvae Genet* 27:161–169
- Sakai K-I, Hatakeyama S (1963) Estimation of genetic parameters in forest trees without raising progeny. *Silvae Genet* 12:152–157
- Searle SR (1979) Notes on variance components estimation: a detailed account of maximum likelihood and kindred methodology. Paper No BU-673-M, Biometrics Unit, Cornell University, Ithaca NY, 146 pp
- Shrikhande VJ (1957) Some considerations in designing experiments on coconut trees. *J Indian Soc Agric Stat* 9:82–99
- Stonecypher RW (1966) Estimates of genetic and environmental variances and covariances in a natural population of loblolly pine (*Pinus taeda* L.). PhD Diss, North Carolina State University, Raleigh NC, 169 pp
- Swallow WH, Monahan JF (1984) Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics* 26:47–57
- Swallow WH, Searle SR (1978) Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. *Technometrics* 20:265–272
- Tan H (1977) Estimates of general combining ability in *Hevea* breeding at the Rubber Research Institute of Malaysia. 1. Phases II and IIIA. *Theor Appl Genet* 50:29–34
- Usanis RA (1972) A simulation study on the estimation of random and systematic variance components in a plantation with randomly planted genotypes. PhD Diss, NC State University, Raleigh NC, 99 pp