Problems in Estimating Genetic Variance by Shrikhande's Method[1]

By G. Namkoong and A. E. Squillace[2]

(Received for publication August 25, 1968)

A quick and inexpensive method for estimating broad-sense heritability has been proposed by Shrikhande (1957) and by Sakai and Hatakeyama (1963), but has not been critically examined. The method involves estimating the variance among plots in a stand of randomly planted genotypes and measuring the change in average plot variances when plot size is changed. The effect of environmental variability is expected to be reflected in the rate at which the plot variances change with changes in plot size in the stand. It is asssumed that the effect of all site factors and cultural practices can be integrated into a simple function of plot size alone. For example, very strong and rapidly changing soil factors or strong effects of competition are expected to decrease intertree correlations rapidly with increasing intertree distances. The result would be a rapid change in plot variance with increases in plot size. Alternatively, uniform site treatments would tend to homogenize the site and cause change in plot variances to be more gradual. The empirical function developed by Smith (1938) has satisfactorily approximated the response of height and diameter growth in plot-size studies on loblolly pine by Conkle (1962) and in genetic studies by Shrikhande (1957), Sakai and Hatakeyama (1963), and Armitage and Burrows (1966). If the effect of environmental variation on the change in plot variances with plot size can thus be accounted for, and if random genetic variation among adjacent trees exists, then the genetic contribution to plot variation can change at a different rate than the environmental source of variation. This difference is exploited by Shrikhande's method.

In the recommended procedure, all trees in an appropriate stand are first measured and mapped. Then artificial plot boundaries are drawn to simulate an experimental planting having plots of size x. This is done separately for x = 1 up to n. Next the variance between plot means for each of the plot-size "experiments" is computed and the change in plot-to-plot variance is charted over the change in plot size. The procedures described below are then applied to derive estimates of the environmental and genetic contributions to the total variance.

Detailed examples of the method are given in the references cited at the end of the present paper. Since it may be highly useful in providing rapid and inexpensive heritability estimates, the method deserves empirical testing. The specific problems which concerned us were whether the genetic variance components were estimated without bias and with low error and whether the recommended "least squares" procedure gave estimates that actually minimized the error sum of squares. If these practical difficulties can be avoided, only the applicability of the material chosen for study and problems of competition would inhibit the utility of this method. A large-scale application is reported by Armitage and Burrows (1966), who applied Stevens' (1951) procedure for estimating the components. We use a similar procedure described by Namkoong and Miller (1968) to examine some of the practical aspects with data on a 19-year-old plantation of slash pine (*Pinus elliottii* Engelm.) growing near Olustee, Florida.

Fitting the Model

The method is based on the assumption that the variance among plot means due to environmental effects changes in a predictable manner as plot size increases. If we assume that this change follows Smith's (1938) empirical law then,

$$V_x = V_0 x^{-b}$$

where V_0 = environmental variance on an individual basis
V_x = environmental variance on a plot mean basis when plots are composed of x individuals
b = Smith's empiric regression coefficient.

The change in V_x due to V_e will follow a logarithmic function of the plot size. If it is also assumed that there is no correlation of genetic value with the location of the individual trees, then the change in V_x due to genetic effects will be a direct inverse function of plot size. Shrikhande (1957) formulates these relations as:

$$V_x = \frac{V\,(\text{genetic})}{x} + \frac{V\,(\text{environment})}{x^b}$$

Given computations of V_x for a sequence of plot sizes, one can easily obtain least-squares estimates of V_g, V_e, and b. It can be seen that as b approaches 1, V_g and V_e would have identical effects on V_x and would be statistically indistinguishable when b = 1. Hence the method is unusable for traits in which b is close to 1, as might occur in traits which are influenced by competitive or other age-dependent effects such as found for height and diameter growth of loblolly pine by Conkle (1962).

To evaluate the procedure, a test should be made of the consistency with which the response of V_x to plot size fits Smith's logarithmic model. The possibility that the response to environmental variation is not logarithmic should be examined, since forests are often planted on marginal sites of high variability. If the soil variation shows a mosaic pattern of fertility levels rather than a gradient, a failure of the model to fit the data would also contribute to estimation error and make interpretation difficult. In addition, estimates of soil heterogeneity may be biased by competitive effects. Therefore, an examination of the errors associated with the fitted model is necessary for assurance that the model for environmental effects is a good one.

In estimating broad-sense heritabilities from clones or parents alone, the inclusion of variances due to special clonal or other nongenetic parental effects as well as to dominance and epistatic effects is both bothersome and unavoidable. Since the magnitude of these effects is specific for each experiment, they will not be considered in this paper though they must be recognized as detrimental

---

[1] Published with the approval of the Director of Research, North Carolina Agricultural Experiment Station, as Paper No. 2422 of the Journal Series. It is a contribution of the Genetics Department of North Carolina State University at Raleigh and the Southern Forest Experiment Station, Forest Service, U. S. Department of Agriculture. The computing services were supported by N. I. H. Grant No. FR-00011, held by the Institute of Statistics, North Carolina State University at Raleigh.

[2] G. Namkoong is Geneticist, Institute of Forest Genetics, Southern Forest Experiment Station, Forest Service, U. S. Department of Agriculture, and the Department of Genetics and School of Forestry, North Carolina State University at Raleigh. A. E. Squillace is Principal Plant Geneticist, Southeastern Forest Experiment Station, Forest Service, U. S. Department of Agriculture, Olustee, Florida.

to the utility of the material in this procedure. In addition, competitive effects among randomly planted genotypes may be sufficiently different (as compared with the competition among trees more or less similar in growth potentials) to cause some bias in component estimation.

Sakai and Mukaide (1967) propose an extension of Shrikhande's method to include competition effects. Where intergenotypic competition effects of the kind they define are important and the definitions and assumptions they require are valid, their extension may be very useful. Since these conditions do not always exist, and because the simpler model may often suffice, the developments of these authors are not considered in this investigation.

Estimation Procedures

Since the equations derived for fitting are nonlinear, there is no assurance that the stepwise procedures recommended by Shrikhande will either give a minimum error sum of squares or that any minimum found is unique. Hence, simultaneous fitting procedures should be employed and searches made to assure that other parameter combinations will not give estimates with less error. Fortunately, the equations lend themselves to a fairly direct iterative procedure for solution of the least-squares estimators. The general equation is:

$$xV_x = V_g + V_e x^{1-b} = MS_x$$

where MS_x is the mean square attributable to variance between plots of size x in an analysis of variance of between- and within-plot sources of variance. As in the Gauss-Newton iterative procedure employed by Stevens (1951) for a different equation, one uses initial estimates of (1—b), V_g, and V_e to get second estimates of these parameters. In turn, these estimates are used to derive a third set, etc., till the estimates no longer change from one set to the next. The procedure involves taking the first element of a Taylor's series expansion for the solution of the roots of an equation, to provide the best simple estimate of the solution vector. It is also actually a multivariate extension of Newton's method for iteratively solving for the roots of equations. Namkoong and Miller (1968) detail the method for estimating coefficients in regression equations of the form: $y = A + Bx^C$. Their procedure can be used to estimate V_g and V_e by making the substitutions:

$$V_g = A$$
$$V_e = B$$
$$1\text{—}b = C$$
$$x_i V_{x_i} = Yi$$

Modifications suggested by Hartley (1961) proved helpful, but were not necessary for convergence.

This estimation procedure is relatively simple. It minimizes the error sum of squares and provides unbiased estimates of the model's parameters but ignores the fact that the errors are correlated because the same trees are used in the various computations of plot size. If the experiences of Elston and Grizzle (1962) are applicable to this nonlinear estimation case, the parameter estimates are good but the errors of estimate are not; and any confidence intervals derived are probably too small. For our purposes, the simple procedure seemed sufficient and more complicated multivariate procedures were not tried.

In addition to these estimates, a range in the space defined by the A—B—C variable parameters was graphically explored by a charting program written for an IBM-360 model 30 plotter. The graphs allow visual examination of the general sums of squares of error surface over a wide range of combinations of parameters and enable one to determine the significance of irregularities in this surface, thereby facilitating the location of more than one local minimum. The support plane confidence interval is

$$\hat{O} \pm \sqrt{pF_{(p,\,n\text{-}p)} \cdot \frac{SSE}{n\text{-}p} \cdot P^{-1}}\,,$$

where n = number of plot sizes sampled and p = number of variables (3), and P is the appropriate element of the error covariance matrix.

Materials and Methods

Data were obtained and the model fitted by the procedures of Shrikhande (1957) and Sakai and Hatakeyama (1963).[3] Nine tenth-acre blocks of trees were used. Since mortality had caused uneven spacing, only those trees which were still within zones with the original 8- × 8-foot spacing were used. The 216 trees which fell within these zones were divided into 1-tree plots. The two-tree plots were drawn first in the direction of the rows and all computations made. Then the operation was repeated for two-tree plots in the direction of the columns and the variance estimates pooled. The three-tree plots were similarly handled, with variation in plot shape permitted by arranging 3 × 1 and 1 × 3 plots as well as 2 × 1 plus a tree in the adjacent row. Similar variations in plot shape with a pooling of variance estimates were made for plots of 4, 5, and 6 trees. No larger plots were possible.

Results

Pooled variance data on height, diameter, and gum yield (adjusted for diameter variation) are displayed as sample

---

[3] The aid of Mr. S. Hatakeyama in plotting the trees and defining groups is gratefully acknowledged.

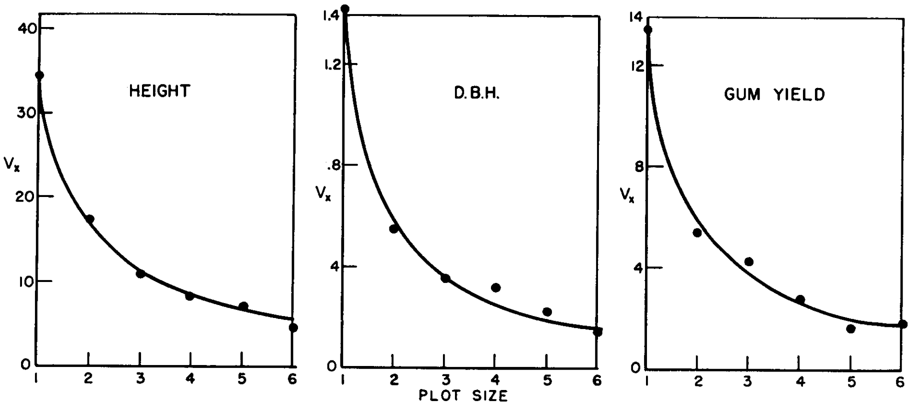

Figure 1. — Variance among plots of size x and the fitted curves.

points in *Figure 1*. The minimum variance curve as fitted by our least-squares iterative procedure is also drawn to indicate the closeness of the fit. No pattern in the lack of fit was discernible.

Height data gave such anomalous results (large negative estimates of variance components) that the error surface was more thoroughly examined. It is apparent that C approaches zero closely in this case (equivalent to Smith's heterogeneity coefficient being 1) and, hence, that the V_g and V_e are almost inseparable. Thus, the error surface would approach having an infinite number of minima — the equivalent of being flat over some region. Graphical solutions show that a very flat region does exist when $b \rightarrow 1$. The iterative procedure was very rapid for converging to a circle of points. However, the surface is apparently in the shape of a broad valley with steep walls and with a local maximum in the center. Solutions for local minima lie in the second and fourth quadrants of the V_g and V_e space. In this particular case, the solutions are uninterpretable genetically. The minimum-error estimates and the 50 percent C.I. for height are 239 ± 32 for V_g and -204 ± 31 for V_e. The estimate of b is 1.014.

The minimum-error estimates (and 50 percent C.I.) of V_g and V_e for gum yield were .151 (±.09) and —.020 (±.08), and the estimate of b = .45 resulted from an error surface which has a well-defined minimum. The negative estimate of V_e is disturbing, it may be caused by sampling error; since the estimate is small, the environmental variance is likely to be small and the heritability high for this trait. A high heritability estimate would agree with the indications of strong heritability for gum yield in slash pine reported by Squillace and Bengtson (1961).

In contrast to the values for height and gum yield, the diameter variance measurements gave an estimate of b = 1.529, and a moderately well-defined unique minimum for V_g and V_e at .789 (±.6) and .596 (±.5), respectively *(Figure 2)*. The estimate of Smith's heterogeneity coefficient of 1.529 indicates either an extremely high microsite variability or significant competition effects, or both. The heritability estimate for diameter was .58, which is in the upper range of that estimated by Sqiullace and Bengtson (1961) on the same species grown in the same area. The C.I.'s are quite large, even by this estimation procedure. By trial and error it was determined that, if the variability were consistent, nine plot sizes would be required to greatly reduce the C.I.

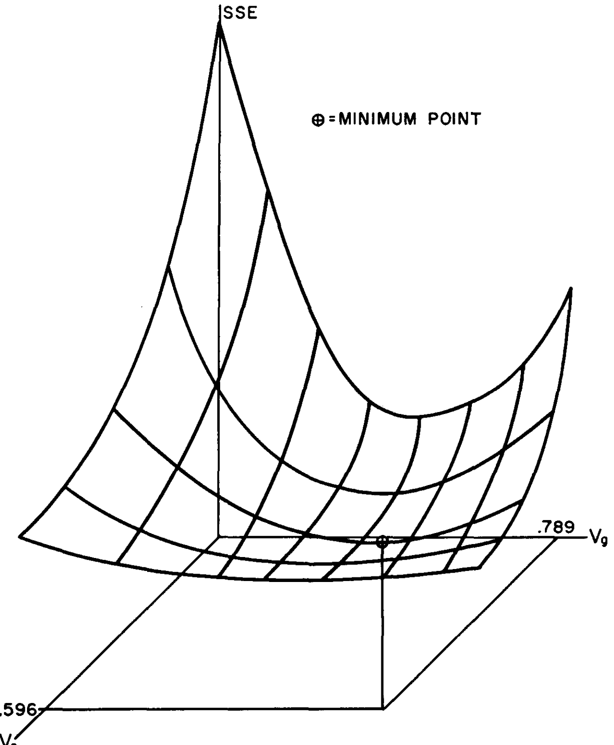

Figure 2. — Sum-of-squares error surface at c = .529 for d.b.h.

Since it is clearly possible that hidden difficulties exist in estimating the least-squares solution and that one might still face interpretative problems, the data on height ("Asahigawa" data) listed by Sakai and Hatakeyama were analyzed by the iterative method. The estimates of within-clonal variance converged rapidly to the same estimates that these authors derived by first finding b, then V_g and V_e. However, the analysis of seedling height data showed that lower error sums of squares are obtained with estimates of $V_g = .04$ and $V_e = 1.82$ and $b = .66$. The error sum of squares is .136 by our estimates and .243 with their parameter estimates. Thus, our procedure would lead us to more reliably estimate a heritability of .02 instead of their .41 and hence to reach quite different decisions on breeding procedures. In both of these cases, the existence of other local minima cannot yet be completely ruled out. At b = .51 as they estimated, two local minima occurred. Therefore, even when b is not close to 1, it is necessary to use proper minimization methods to derive good estimates.

To provide further tests of the value of the iterative method, the data of Shrikhande were used. They resulted in parameter estimates very close to those Shrikhande obtained in both of his trials, and the graphical solution showed that the error surfaces were well curved and had clear, unique minima. Using the "Pilicode" data, he obtained $V_g = 125$, $V_e = 64$, $b = .25$, with an error sum of squares of 1406.3. By the iterative method we obtained 121, 67, .27, and an error sum of squares of 1406.2 — clearly an improvement of no consequence. For the "Kasargod" data, his estimates were 173, 162, .39, and ours were 170, 164, .39, with essentially the same error.

By way of conclusion, it seems clear that the value of the estimation technique must be protected by the analyst's judgment as to whether Smith's model fits the situation. The greatest difficulties in using the technique occur at values of b near 1. For values of b away from 1, a well-defined, unique minimum would appear to exist. The stepwise procedures of first finding one coefficient and then the remainder would always be difficult to trust since even a b of .51 led Sakai and Hatakeyama to estimators with higher errors and high heritability estimates. It is also clear that at least nine sizes of sample plots should be used to obtain reasonably small confidence intervals. Since b values obtained provide a wide sampling of what may occur in other experiments, we may conclude that Shrikhande's experimental method is useful and dependable as long as one chooses data which fit Smith's law, b is not close to 1, and estimation procedures are adequate (Stevens' procedure or other more complete procedures).

Summary

Estimating broad-sense heritabilities by Shrikhande's method can be both convenient and inexpensive. The various assumptions required for the successful use of the method are examined. The analyzed traits must have plot variances which change according to a logarithmic function of changes in plot size, and the genetic effects must not be unduly confounded by competitive effects. In addition, it

is found that the heterogeneity coefficient must not be close to unity and that adequate estimation procedures, such as an iterative least-squares, must be used.

Literature Cited

Armitage, F. B., and Burrows, P. M.: Preliminary heritability estimates for *Pinus patula* in Rhodesia. Rhodesia Zambia Malawi J. Agr. Res. 4: 111—117 (1966). — Conkle, M. T.: The determination of experimental plot size and shape based on the variation of tree heights and diameters in plantation grown loblolly (*Pinus taeda* L.) and slash (*Pinus elliottii* Engelm. pine. M. S. Thesis. North Carolina State Univ. Raleigh, 78 pp. (1962). — Elston, R. C., and Grizzle, J. E.: Estimation of time-response curves and their confidence bands. Biometrics 18: 148—159 (1962). —Hartley, H. O.: The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. Technometrics 3: 269—280 (1961). — Namkoong, G., and Miller, D. L.: Estimation of non-linear parameters for a non-asymptotic function. Biometrics 24: 439—440 (1968). — Sakai, K., and Hatakeyama, S.: Estimation of genetic parameters in forest trees without raising progeny. Silvae Genetica 12: 152—157 (1963). — Sakai, K., and Mukaide, H.: Estimation of genetic, environmental, and competitional variances in standing forests. Silvae Genetica 16: 149—152 (1967). — Shrikhande, V. J.: Some considerations in designing experiments on coconut trees. J. Ind. Soc. Agr. Stat. 9: 82—99 (1957). — Smith, H. F.: An empirical law describing heterogeneity in the yield of agricultural crops. J. Agr. Sci. 28: 1—23 (1938). — Squillace, A. E., and Bengtson, G. W.: Inheritance of gum yield and other characteristics of slash pine. Sixth South. Forest Tree Impr. Conf. Proc. 1961: 85—96 (1961). — Stevens, W. L.: Asymptotic regression. Biometrics 7: 247—267 (1951).

Untersuchungen zur Konkurrenz zwischen verschiedenen Genotypen in Pflanzenbeständen

III. Das Korrelationsmuster eines Bestandes

Von M. Hühn*)

A) Einleitung

Daß die gegenseitige Beeinflussung verschiedener Individuen, die als Nachbarn im Bestand aufwachsen, eine große biologische Rolle spielt und das Aussehen von Pflanzenbeständen mitbestimmt, ist seit langem aus vielen experimentellen Untersuchungen bekannt.

Statistische Maßzahlen zur Beschreibung von Abhängigkeiten zwischen Zufallsgrößen sind z. B. Korrelationskoeffizienten. Es ist daher nicht zu verwundern, daß eine in der Literatur häufig angewandte Methode, Maßzahlen zur quantitativen Untersuchung und Einschätzung von Konkurrenzeffekten in Pflanzenbeständen zu erhalten in der Schätzung der Korrelationen zwischen Nachbarn besteht. Man hat versucht, diese Komponente der phänotypischen Varianz aus der phänotypischen Korrelation zwischen einer Pflanze und ihren umgebenden Nachbarn zu schätzen. Diese Korrelationen erweisen sich meist als negativ, d. h. kleinere Pflanzen stehen meist neben größeren und umgekehrt (Stern 1966, Sakai, Hayashi und Mukaide 1966, Kennel 1966, Sakai und Mukaide 1967, Lichter 1967, Mead 1967, Lichter 1968, Sakai, Mukaide und Tomita 1968, Stern 1968).

Diese Verfahren wurden in der letzten Zeit nach vielen Seiten hin verfeinert: z. B. arbeitet man nicht direkt mit der Korrelation, sondern man nimmt die Differenz zwischen dem Phänotyp der Zentralpflanze und der gewogenen Summe der Erträge ihrer Nachbarn. Dadurch gelingt es sicher, etwas die positiven Bodenkorrelationen auszuschalten, die zwischen Nachbarn in Pflanzenbeständen immer existieren, da sie ja ähnliche Bodenverhältnisse gemeinsam haben: Betrachtet man z. B. hexagonale Anordnungen der konkurrierenden Nachbarn, ist m das Gesamtmittel, y_i der phänotypische Wert der Zentralpflanze, und macht man für die $y_i' = y_i - m$ die Annahme, daß die

$$e_i = y_i' - \lambda \sum_{j(i)} y_j' \qquad (1)$$

eine $N(O, \sigma^2)$-Verteilung haben (dabei läuft die Summation j(i) über die Nachbarn von i), so kann man den Konkurrenzparameter λ mit der Maximum-Likelihood-Methode schätzen. Durch Heranziehung von Monte Carlo Methoden ist es möglich, die Verteilung von λ für die verschiedenen geometrischen Nachbarschaftsmuster zu bestimmen und theoretisch zu untersuchen; ebenso den Einfluß von Schiefe und Exzeß der Ausgangsverteilung auf diese λ-Verteilung. Durch Einführung verschiedener Konkurrenzkoeffizienten λ_i kann man dieses Modell auf den Fall der Konkurrenz zwischen verschiedenen Genotypen erweitern. Doch liegen über diese Verallgemeinerung des Modells von Mead (1967) noch kaum Untersuchungen vor.

Diese experimentell meßbaren phänotypischen Korrelationen sind jedoch kein genaues Maß für die Effekte, die durch Konkurrenz zwischen verschiedenen benachbarten Genotypen entstehen, da sie ja stets ein Bruttoresultat aus vielen wirkenden Korrelationsursachen sind. Die den Pflanzenzüchter und Genetiker interessierenden Größen können also auf diesem Wege nicht — oder nur sehr ungenau — ermittelt werden, und es bleibt bei allen bisher angewandten biometrischen Methoden ein breiter Raum für subjektive Interpretationen.

Man weiß z. B., daß Dichtstandsbedingungen die Einheitlichkeit von Populationen ganz entscheidend mitbeeinflussen können: „So hat man z. B. festgestellt, daß mit zunehmendem Dichtstand des Pflanzenbestandes der Bestand insgesamt einheitlicher wurde, daß man also die größten Unterschiede zwischen den Pflanzen fand, wenn man sie ohne Konkurrenz aufwachsen ließ. Konkurrenz muß also nicht immer zum Beherrschen und Unterdrücken führen, sondern kann durchaus ein intermediäres Resultat bringen: Insgesamt wird der Bestand einheitlicher, aber trotzdem existieren innerhalb des Bestandes die genannten negativen Korrelationen“ (Stern 1968).

Wenn diese phänotypischen Korrelationen auch keine genauen Aufschlüsse über bestehende Konkurrenzeffekte geben, so können sie doch als Parameter für die Beschreibung von Pflanzenbeständen dienen, die ja nicht einfach als Summe ihrer Mitglieder aufzufassen sind, sondern durch diese gegenseitigen Abhängigkeiten wesentlich mitbestimmt werden. Will man einen solchen Pflanzenbestand mit statistischen Maßzahlen beschreiben, so braucht man nach Stern (1966) mindestens die folgenden Statistiken:

---

*) Institut für Forstgenetik und Forstpflanzenzüchtung in Schmalenbeck der Bundesforschungsanstalt für Forst- und Holzwirtschaft.