

SOME STATISTICAL ASPECTS OF PROGENY TESTING

by

Thomas C. Evans, John C. Barber, and A. E. Squillace
Southeastern Forest Experiment Station
Forest Service, U. S. Department of Agriculture

They went to sea in a sieve, they did.

In a sieve, they went to sea.

-- E. Lear

These familiar lines from a famous bit of nonsense verse seem especially appropriate, for upon reviewing the 10-year progress in methodology of progeny testing for southern species in the hope of offering some practical suggestions to progeny testers. I have the unhappy feeling of truly setting sail in a sieve on a sea of conjecture.

Ten years of progeny testing southern species have not produced as much improved methodology as we had originally hoped. Perhaps this should not be surprising. The field, after all, is relatively young, the responses are glacially slow; what appears true at one point of experience seems not necessarily true at another, and the demand for methods of high sensitivity seems always to lie beyond our immediate experience and knowledge.

If there is a fault at all, it is a persistent failure to examine the data we have in the interest of searching out information which might lead to more enlightened methods. The most encouraging fact in the 10-year history is that the whole gamut of methodology has come into sharper focus. More and more mature effort is going into answers to questions which were compromised or simply bypassed a decade ago; What size plot? What precision? Which design? We will not presume to answer these questions. We will attempt to sum up some of the experiences and preferences in the Southeastern Station's program in the hope that the recommendations they imply will be useful in the head-long program of progeny testing we foresee for commercial and industrial tree breeders.

First, we accept the minimum standards suggested by the Sub-Committee on Progeny Testing for Seed Certification as an appropriate starting point. We submit, and we believe the Sub-Committee appreciates, that the standards are only a guide for prudent governing bodies and administrators. What we offer are a few comments, and dogmas, strongly expressed, in favor of more rigorous testing and in the interest of emphasizing some of the Committee's recommendations.

I wish it were possible to make this opening point with categorical emphasis. If it were, I would offer that at this stage of our experience, we should categorically deny any progeny test not containing a statement of the least difference which, in the mind of the investigator, would be meaningful. In the early stages, when there is no knowledge of the expected differences between progeny, it is wise and profitable to indulge tests merely bracketing the range of those differences, provided, of course, that we accrue information on the kind and degree of variability which accompanies the test. This, after all, is the purpose of exploratory trial, for with some idea of the expected differences and with some knowledge of variation, the tree breeder and statistician can work together in choosing design, plot size, number of replications, and other factors which can most effectively detect the differences at any prescribed level of confidence.

But this alone is not sufficient. During exploratory trial, and certainly before definitive testing, the investigator must decide what order of difference would be meaningful. Preliminary trials may show very consistent differences, in which case the test would demonstrate high significance to the joy of the statistician. But if the differences are small, they may have little or no practical value and hence would not be meaningful to the practical tree breeder.

Note the powerful implications of a "meaningful" difference as opposed to the fashionable and often overworked "significant" difference. Whatever the degree of difference, if it is not meaningful, there would be little justification for continuing the test, no matter what its significance. On the other hand, if an observed difference is meaningful and significant, the prayers of the investigator are answered. If it is meaningful and not significant, a more comprehensive test is justified.

Among the progeny testers and geneticists of my limited acquaintance, few are willing to put precise limits on what they would consider a meaningful difference for any single trait. Fewer still would guess at the difference which by itself might be "meaningless" but which in combination with other "meaningless" differences may produce a meaningful difference for the combination. Generally, we take refuge in the accumulating evidence that differences less than ten percent are difficult to detect with our present methods.

If this problem could be inverted, and we could begin with some idea of a least meaningful difference, the net effect would be to markedly simplify initial screening of a large number of progeny, to permit more sensitive designs with fewer progeny, and generally to improve and sharpen the objectives of study. There is a most urgent need for serious consideration of what these differences should be for the thirty-odd traits under examination in various tests of southern species. I am quite aware of the staggering proportions of the task, but this Committee has not been over-awed by other equally profound problems. I

offer that no other advisory body is as well equipped to suggest the guides so badly needed by research as well as by commercial and industrial tree breeders.

Categorically, we deny any design in which the number or size of plots forces blocks to be unreasonably large. This point of caution would be trite if progeny testers would observe the classic warning, but we continue to hear and read of randomized block experiments where the block ranges up to three or four acres in size. We have no great fund of evidence to demonstrate the wisdom or the fault of these designs, but there are some general guides. In theory, if a fertility gradient exists in the planting area, and if that gradient should be assigned logically to random error, the block size should decrease as the gradient increases. The intent is that differences in random fertility would be identified with blocks, and the derived error would be a valid estimate of random positional variation, and a proper error for testing differences among progeny. On the other hand, if blocks are excessively large, differences between blocks tend to vanish, but, even more important, the error tends to resemble a true site-progeny interaction rather than random variation in microsite, and hence tends to make the test insensitive.

It is difficult, if not impossible, at this point of our experience to put boundaries on block size over the range of planting sites in the South. But our experience in the piedmont of Georgia indicates that blocks approximately one acre in size are entirely too large in that area. Barber will discuss this point in some detail. In the coastal plain there is some evidence that blocks may be larger; how much larger we cannot say. The important generality is that size of block can invalidate, and for that matter often does destroy, many otherwise well designed tests, and the only plausible solution in the absence of worthy evidence is to favor small blocks.

Block size may be controlled by manipulating either the size of the plot or the number of plots it contains, or both. I will begin on a point of mild controversy and some documentation.

With few exceptions, such as limited amounts of seed from stock of exceptional value, extreme limitations in availability of land, or complete refusal of statistical inference, we categorically deny the single tree plot. In spite of its reputed efficiency, the conditions to attain that efficiency are notably absent in tests of southern species. In the first place, mortality is common, and often relatively high. Interplanting to repair damaged plots is unsuccessful because interplants remain depressed throughout at least their juvenile development. The alternative dodges of omitting plots, or using missing-plot techniques are equally unsatisfactory in that the test is always weakened, and more often utterly useless. In competition, many if not most studies are set out with the provision that competition within the plots is a necessary condition of the test. Height and diameter growth, for example, are assessed in the presence of competition from individuals of the same clone or progeny group. For this objective, the single tree plot is worthless, and our interest centers on the maximum

amount of information per replicate, not per unit area.

But perhaps the most formidable argument against the single tree plot is the likelihood that the genetic portion of variation from plot to plot is non-normal. It is recognized that variation in progeny tests is of two kinds, that of the variation among trees of the same progeny, and that due to position in the plot. Much of the evidence in agriculture and forestry supports the notion that positional variation tends to be normally distributed, and that if this were the sole source of variation, single tree plots would be quite amenable to analysis of variance. But since the error of progeny tests is a combination of both genetic and positional variation, total variation may or may not be normal, depending on the relative magnitude of the two sources. With so little information on the distribution, the best gamble would seem to be a multiple tree plot, because of the property that means drawn from non-normal populations tend toward normality as sample size increases. It is, of course, anyone's guess how large a plot is required to attain normality or near-normality. We can only speculate, but it seems to be an intelligent speculation that a multiple tree plot is preferred.

Quite apart from the question of normality, optimum plot size has been examined with data from the 3-year heights of the tests at the Ida Cason Calloway Foundation in the piedmont of Georgia, and 10-year height, diameter, bark thickness, and clear length from a spacing study in slash pine of the upper coastal plain, and with data for the same traits in 2-year-old slash pine plantations in the upper coastal plain. The graphs, taken from the 10-year-old spacing study only, illustrate a point which appears in all these tests examined to date; i.e., the coefficient of variation tends to become stable in the neighborhood of 20 trees per plot for all traits observed. In itself, this is but a shred of information, but taken together with our concern with lack of homogeneity of the planting site and the likelihood of non-normal variation in very small plots, the combined evidence has led us to an arbitrary decision to use a 25-tree observation plot.

There remain, then, only the question of how many progeny can be admitted without unduly inflating block size. One-acre blocks will admit about 25 at the most. Most commercial tests will be with fewer progeny and will offer no real problem, but industrial and research programs may quite easily require space for 75 to 100 progeny or even more. Here the only alternatives seem to be:

- (1) Insist upon homogeneous planting site.
- (2) Ignore significance tests in favor of arbitrary and ruthless screening to reduce the number of progeny for definitive testing in a second stage.
- (3) Select more sophisticated designs, such as the incomplete block.

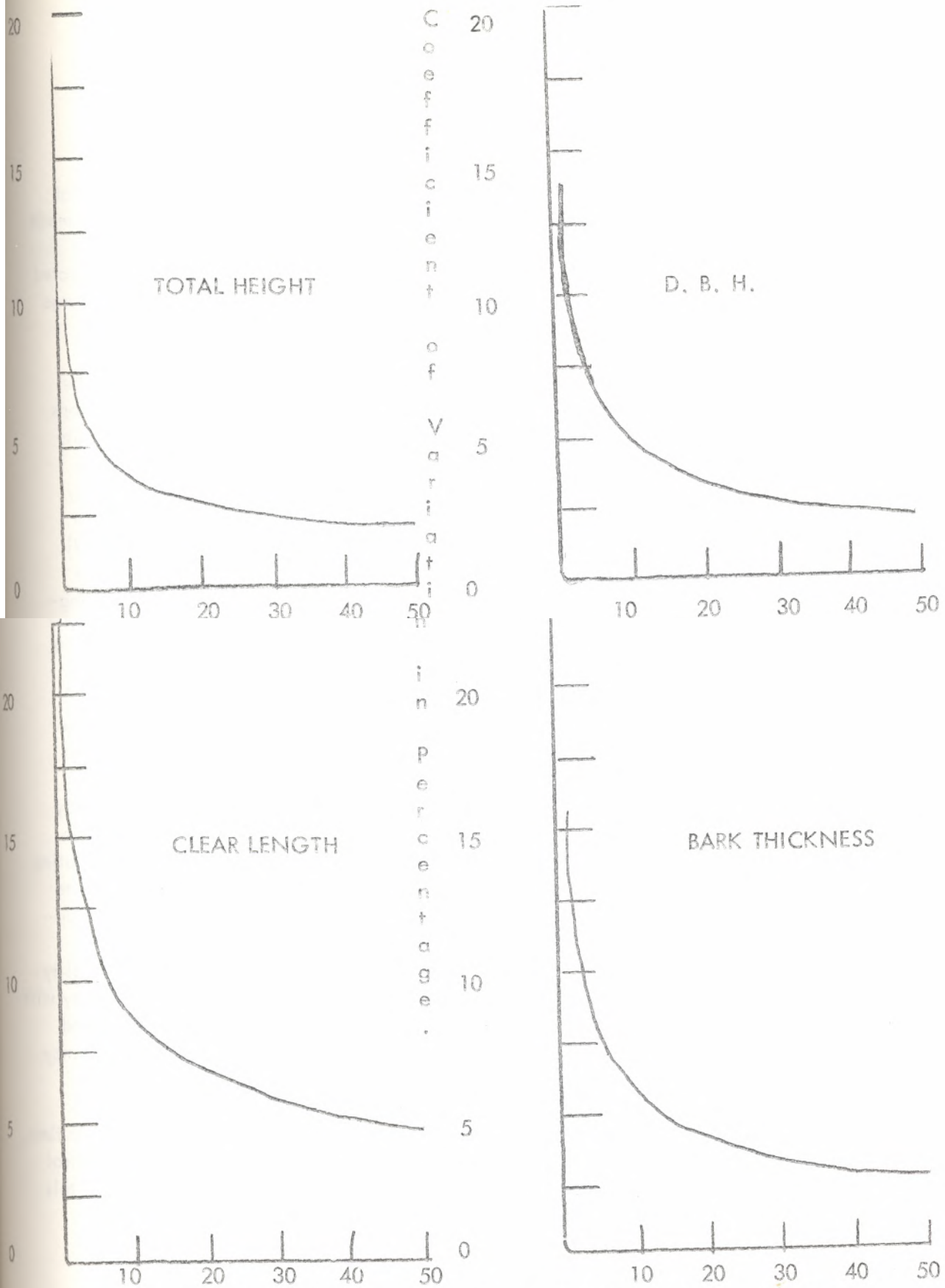


Fig. 1 --Relationship of Coefficient of Variation of Plot Size in 10 Year Slash Pine. Number of Trees Per Plot

Our decision is not yet clear, but at the moment we seem to be persuaded in favor of 2-stage testing of the plantings made during the past ten years. New plantings will be made in a series of smaller tests, each with different progeny but with one or more standard controls in each test. This will permit crude testing at the time of screening. The expected gain may be small, but in our opinion it is preferred to the imprecision of excessively large blocks on the one hand and the insecurity of the more sophisticated designs on the other:

In designing studies for the traits we have discussed and others, I assume you will be calling upon some statistician for advice. If you do, I can assure you one of his questions will be where he can get information on the variation you will contend with for the traits you will test in the area and time where you propose the test. Without this information much of his advice will be fruitless. Tree breeders and geneticists are not without data; our files and your files are full of data, most of it relevant to the point. But little of it has been published, probably because it is only of secondary interest to readers. But all of it is important to progress in the design of progeny tests; the only problem is how the information can be made available to other workers. Most of you have seen the Planning-Factors Manual issued by the Forest Service this year. It, or something like it, is one possible outlet. But whatever the outlet, I urge all of you to promptly analyze and study the data you have, and in some way make it available promptly to the rest of us. If you do, I am sure the effort will be rewarding.

One other point is worth emphasizing. Our persuasion is to refuse all second-stage progeny tests which are not replicated in both time and space. Ideally, this preference should exist for all gents, but when the number of progeny under test is large, as it frequently is in exploratory trials, the enormity of the task makes this replication utterly impossible. There is some evidence of the effect of time and space, and a considerable body of data as yet unanalyzed in our own and other files. But even without substantiating evidence, both variables should be tested; time, in the interest of assessing the response to insects, disease, drought, frost, flooding, or any other factor associated with year-to-year variation in climate; space, in the interest of assessing the response to soil, site, history, land use, or any other factor associated with place-to-place variation in fertility. And, even more important, there is every reason to examine the interactions of progeny with both time and space. The consistency with which both factors interact with species and treatment in other fields of forest research is significant evidence that the interaction may well be the most important single contribution of the tests.

There are many other points I would like to discuss with you; but in the interest of time, I have touched on only three: namely, the place of a meaningful difference, the control of block size, and the need for replication in time and space as the most important at this point in our experience.

If we have been sailing in a sieve, I would remind you that in the nonsense poem the whole nation of Jumbies took to sea in a sieves after the jubilant return of the bold pioneers. Won't you join me?