

# Field Testing and Data Handling in Forest Genetics Research

Roy W. Stonecypher .V

The two subjects I have been assigned to cover in this paper are such that I can only hope to scratch the surface in the time allotted.

I will attempt to present briefly some of the problems associated with field testing and data handling, and hopefully offer a few suggestions which might be used to attack these problems.

## Field Testing

In discussing field testing, I am going to take the liberty of assuming that we have a population of trees (usually a single species) about which we desire information concerning the variation present. The information desired may take the form of estimates of genetic and environmental variances and covariances, test of differences among genetic entries, or both. Although the type of information sought can modify the specific experimental design used, there is usually an orderly sequence of steps to be taken in establishing studies to obtain this desired information.

## Mating Designs

The first step to be taken is to decide what system of matings will be used to develop relatives for the study. The choice of the mating design will depend on the objectives of the experiment, the resources available, and the flexibility of application to forest trees. A list of several possible designs and remarks concerning them is given in table 1.

If the main objective of the study is to estimate additive and non-additive genetic variances as well as some measure of general and specific combining ability, the two factor mating designs of table 1. should be considered. It should be noted that, although control pollinations are required for the one factor full-sib design in table 1, it is not possible to estimate separately additive and non-additive genetic variances. This design would therefore, have little utility for estimating genetic variances.

Most of the mating designs in table 1 have been used for estimating genetic variances in forest tree populations and some information as to the number of families required for reasonably reliable estimates of genetic variances has been published (Goggans, 1962; Stonecypher, 1966).

The choice of mating design and number of crosses used in a well designed study for selection or progeny testing can differ from that of a study which has the major objective of reliably estimating genetic variances.

This difference comes about for two reasons. First, the study properly designed for estimating genetic variances should be concerned with the variances of components of variance while the study for testing or selection should consider the variances of means. Second, some of the populations of interest to the tree breeder can be considered fixed as far as the genetic entries are concerned. In other words, there is more interest in being able to compare accurately the performances of a set of genotypes derived from a number of parents than in making a general interpretation of genetic variances estimated from these entries.

Unfortunately little definitive information has been developed on efficiency of various mating designs and number of crosses required for selection or progeny testing in forest trees. With the increasing availability of reliable estimates of genetic and environmental variances from tree populations, it should be possible to determine the efficiencies of these mating designs for selection. Such information is badly needed.

## Environmental Designs

Having arrived at what appears to be the best mating design for the objectives of the experiment, the next logical step in establishing the study involves the placing of the genetic entries in the environment where they will be tested. This placement must be carried out in such a way that the genetic information desired is not confounded with

---

<sup>1</sup> Research Forester, International Paper Company, Southlands Experiment Forest, Bainbridge, Georgia.

**TABLE I**  
**Some Mating Designs which have been used in Genetic Studies in Forest Trees**

MATING DESIGN	GENERAL DESCRIPTION	GENETIC VARIANCES ESTIMABLE	PROGENY TEST INFORMATION
<b>ONE FACTOR:</b> Clonal	Vegetative Propagation	Total	Cloning Value
Half-Sib	Open-pollinated seed or polycross seed.	Additive	General combining ability
<b>Full-Sib</b>	Control-pollination of pairs of parents	Additive + non-additive	General and specific combining ability
<b>TWO FACTOR:</b> Diallel	Control pollinations in all possible combinations among a group of parents with or without reciprocals or selfs.	Additive and dominance	General and specific combining ability
Design I	Each member of a group of parents is mated to a different group of parents.	do.	do.
Design II	Each of a group of parents used maternally is mated to each of another group used paternally.	do.	do.

the environment and that the degree of reliability of the information is made as high as possible.

The placement of genetic entries into the environment comes under the topic of experimental design which is such a broad field that time limits a detailed discussion in this paper. I will attempt to briefly comment on some design considerations in relation to genetic studies in forest trees.

The choice of the environmental design, like the mating design, will depend upon the objectives of the study, the resources available, and the degree of precision desired. Basic to any valid experimental design are the conditions of randomization and replication which determine to a large degree the precision attained.

If the number of entries is large, which is frequently the case in genetic variance studies, reduction in block size to obtain within block homogeneity should be considered. The reduction in block size can be achieved by reducing plot size or by using an incomplete block design.

Wright and Freeland (1960) concluded that small plots (one to three trees) were statistically the most efficient in uniformity trials of young red pine (*Pinus resinosa*) and yellow-poplar (*Liriodendron tulipifera*). Detailed development by Conkle (1963), however, showed that single-tree plots were most efficient for four and five year old loblolly (*Pinus taeda*) and slash (*Pinus elliotii*) pines, but multiple-tree plots were equally as efficient as one-tree plots for trees eight, nine and twelve years old. Evans et. al (196-1) suggest a plot size of 25

trees. The work at Southlands Experiment Forest (Stonecypher, 1966) on estimates of genetic and environmental variances, indicates a plot size of 10 to 15 trees is most efficient for estimating variances.

If a plot size of ten trees is chosen, the number of entries must be relatively small if block size is to be held at a reasonable level. The size of the block, of course, can depend on the homogeneity of the test site, but it has been our experience that block sizes much greater than one-half acre, are too large on most test areas used for studies of forest trees. If block size is held below one-half acre, the number of entries, with the plot size recommended above, is limited to about 31.

It is to be expected that many tree genetic studies will involve many more than 31 entries. This is particularly true in cases where the objective is to reliably estimate genetic variances. An answer to this problem lies in the use of incomplete block designs. Recently, Snyder (1965) has discussed the use of lattice and compact family block designs in forest genetics research. It should be emphasized, however, that incomplete block designs were developed for reliability of comparisons of means and not for estimates of components of variance (Cockerham, 1963). Therefore, if the objective of the study is estimation of components of variance, the complete confounding of entries in sets proposed by Comstock and Robinson (1952) and used for loblolly pine by Stonecypher (1966) can be employed.

Another problem in genetic studies of forest trees arises because of the difficulty of obtaining progenies from all the desired crosses in any one year. In such cases, the forest researcher is often reluctant to postpone field installation until the mating design is completed. In most instances, it has been the practice to place the material planted in anyone year in a randomized block design and to include controls or checks in order to compare entries planted in different years. Theoretically such a design is likely to be inferior to a comparable incomplete block design if one exists (Cochran and Cox, 1957). It is doubtful, however, that suitable incomplete block designs can very often be used because of the difficulty in obtaining the crosses necessary for any desired degree of balance.

Number of replications used should depend on the variability to be expected, degree of precision required, available resources and size of the experimental unit (Federer, 1955). Therefore, it is difficult to specify the optimum number of replications which should be used in a particular study. Our experience has been that a minimum of four replications per test site should be used in most forest genetic studies.

#### **Field Layout.**

The term layout refers to actual placement of the genetic entries in the field according to the experimental design chosen.

The importance of care in field layout cannot be overemphasized. Much precision can be gained or lost depending on how the material is placed in the field.

In general, blocks should be compact with as little variation as possible within them, so that differences on the test site will be associated with the blocking procedure.

It is essential to obtain an idea of the variation of the test site in order to be able to ally the blocks with the positional variation which may be present. Use of soil samples, plant indicators, or even planting of a special "test crop" can be used in placing blocks. Much information is often lost by careless blocking.

*Testing in Time and Space.* Whether the objectives of a genetic study are estimates of variances, test of means, or both, the application of the results obtained can be severely circumscribed unless the possible effects of genotype-location or genotype-year interactions are considered. Use of estimates of genetic variances or tests of differences among genetic entries obtained from single environments and applied to other environments is extremely risky. Such application is valid only in cases where there is an absence of genotype-environment interaction, knowledge usually not available.

Since generalization of results from genetic studies of forest trees is usually desired, studies

should be designed to adequately sample a range of environments.

*Testing in Special Environments.* It is often possible to evaluate certain characteristics more efficiently in special test areas. For example, in order to evaluate relative resistance of pine to fusiform rust (*Cronartium fusiforme*) it is necessary to insure that test material is exposed to the sporidia of the disease and the fungus has a favorable environment for development. Replication of the test material in time under field conditions will sometimes provide such exposure, but one can never be sure. An alternative, which saves time and space is to plant the test material at close spacing in a special test area which is maintained to insure proper exposure.

Evaluations of other characteristics, such as juvenile wood characteristics (Stonecypher et. al, 1964) and drought resistance (van Buijtenen, 1966), can also be made in special test areas.

Use of such special environments for evaluating certain characteristics can be a low-cost and powerful tool for the tree breeder.

## **DATA HANDLING**

Most genetic studies of forest trees are of a size and complexity which require other than hand calculation methods for complete analyses. For example, if five traits are measured, fifteen analyses are required if it is desired to obtain information on the ten possible relationships among the five traits.

I am not going to spend time on a discussion of the advantages of electronic computers as tools in forest genetics research. The advantages should be self evident to researchers dealing with the mass of data generated by forest genetics studies. I would like to emphasize the word "tools" in the sentence above. Suffice it to say, that no matter how sophisticated an analysis one is able to develop through the use of a computer, the strength of the results is dependent on the soundness of principles discussed above, and on the accuracy of the data generated from the studies. Computer people have a very simple phrase for this: "Garbage in, garbage out."

Much of the discussion below is directly concerned with the development of data for analyses by an electronic computer system.

### **Data Collection**

A complete discussion of data collection should include comments on measurement techniques, but because of time limitations, this section will deal with the mechanics of data collection. It should be emphasized, as Pearce (1953) points out, that measurements should be taken to provide records that will describe to others what was apparent to the researcher at the time, not merely to provide data for statistical analyses.

Whether the data taken are to be analyzed by computers or by desk calculators, much time can be saved by careful design of data forms. Such a form should include adequate space for complete identification of the material being measured as well as for pertinent remarks by the observer. There is nothing more exasperating to an analyst than trying to perform analyses of data which were crudely recorded on note paper with poor documentation and little regard for natural grouping.

At Southlands, we have found that it is most efficient to group the data on each genetic entry on a separate form. We also record all identification information on the form prior to field measurement, and the forms are sorted according to the route to be taken by the field measurement crew. Such a procedure decreases the number of labeling errors and greatly speeds up field measurements.

#### **Data Origination**

Data origination is a specific term related to computers. It is defined as the translation of information from its original form into a machine-sensible form (Sippl, 1966). While this phase of data handling may appear to be outside the area of responsibility of the forest researcher, it is important to realize that one can greatly influence the efficiency and accuracy of this operation by proper attention to form design and use of the latest peripheral input devices, such as optical readers which are becoming more generally available.

All of us are aware of the great technical advances which have been made in the speed and flexibility of the newer generation of computers. Most of these newer computer systems operate in speeds of billionths of a second, and are capable of analyzing masses of data in a few minutes or even seconds. Such speeds are of little consolation to a researcher if it takes days or weeks to develop the data for presentation to the computer.

It has been our experience that the problems associated with data origination present the biggest bottlenecks to efficient computer use. It should be up to us as forest genetics researchers to stay abreast of the rapid developments being made in devices for improving data handling and to use them when feasible.

#### **Data Processing**

It has been disturbing to note that there appears to be a growing dissatisfaction on the part of forest genetics researchers with the results obtained from computers. The dissatisfaction has expressed itself in the form of such statements as: "I could have gotten the results faster from hand analyses", or "I didn't get the information I wanted." In most cases, these difficulties have occurred because of a lack of understanding of the computer or an inability to communicate directly with individuals who are familiar with the computer.

Computer scientists are rapidly developing

user oriented languages which should largely solve this communications problem in the next several years. However, it is up to us as forest researchers to become familiar with computer techniques and capabilities now. The large-scale, costly studies now installed can only yield their maximum potential results through the use of these powerful computing tools.

#### **SUMMARY**

I have attempted to cover in this paper a brief description of the procedures which are usually followed in carrying out forest genetics research. No matter how involved our mating and environmental designs become, or how intricate the analyses of our data, we must not lose sight of the biological implications of our results. The mere reporting of findings is not enough. We must apply these results to the practical goals of the tree breeder. Application of information from well-designed and analyzed genetic studies can lead to an efficient program for producing superior growing stock for our Southern forests. It goes without saying, that this is the goal for all of us attending this meeting.

#### **Literature Cited**

- Cochran, W. G. and G. M. Cox  
‡ 1950. *Experimental Designs*. John Wiley and Sons, Inc., New York
- Cockerham, C. C.  
‡ 1963. *Estimation of Genetic Variances*. *Statistical Genetics and Plant Breeding*. Publication 982. National Academy of Sciences - National Research Council. Washington, D. C. 53-94.
- Comstock, R. E. and H. F. Robinson  
‡ 1952. *Estimation of Average Dominance of Genes*. Heterosis. Edited by John W. Gowen. Hafner Publishing Company, New York 494-516.
- Conkle, M. Thompson  
‡ 1963. *The Determination of Experimental Plot Size and Shape in Loblolly and Slash Pines*. Technical Report No. 17. School of Forestry. North Carolina State College, Raleigh, N. C.
- Evans, Thomas C., John C. Barber, A. E. Squillace  
‡ 1961. *Some Statistical Aspects of Progeny Testing*. *Proceedings Sixth Southern Conference on Forest Tree Improvement*. School of Forestry, University of Florida, Gainesville, Florida 73-79.
- Federer, Walter T.  
1955. *Experimental Design*. The Macmillan Company, New York.
- Goggans, J. F.  
‡ 1962. *The Correlation, Variation and Inheri-*

tance of Wood Properties in Loblolly Pine. Technical Report No. 14, School of Forestry, North Carolina Agriculture Experiment Station, Raleigh, N. C.

Pearce, S. C.

- ‡ 1953. Field Experimentation with Fruit Trees and Other Perennial Plants. Technical Communication No. 23 of the Commonwealth Bureau of Horticulture and Plantation Crops. East Malling, Maidstone, Kent, England.

Sippl, Charles J.

- ‡ 1966. Computer Dictionary and Hand book Howard W. Sams and Co., Inc., The Bobbs-Merrill Company, Inc., Indianapolis. New York.

Snyder, E. Bayne

- ‡ 1965. Lattice and Compact Family Block Designs in Forest Genetics. Second Genetics Workshop of the Society of American Foresters and Seventh Lake States Forest Tree Improvement Conference Oct. 21-23, 1965.

Stonecypher, Roy W., F. C. Cech, and B. J. Zobel

- 1964. Inheritance of Specific Gravity in Two- and Three-Year-Old Seedlings of Loblolly Pine. TAPPI 47(7): 405-407.

Stonecypher, Roy W.

- ‡ 1966. Estimates of Genetic and Environmental Variances and Covariances in a Natural Population of Loblolly Pine (*Pinus taeda* L). Technical Bulletin No. 5, Southlands Experiment Forest, International Paper Company.

van Buijitenen, J. P.

- 1966. Testing Loblolly Pines for Drought Resistance. Technical Report No. 13 Texas Forest Service.

Wright, Jonathan W. and F. Dean Freeland

- ‡ 1960. Plot Size and Experimental Efficiency in Forest Genetic Research. Technical Bulletin 280 Michigan State University, Agriculture Experiment Station, Dept. of Forestry.