REPORT OF THE SUBCOMMITTEE ON RESEARCH EVALUATION, COORDINATION, AND PLANNING: USEFUL ADAPTATIONS OF GENETIC RECORD SYSTEMS FOR COMPUTER ANALYSIS

by Jonathan W. Wright 1/

Computers are becoming so generally available that they can be expected soon to be used for most computational work connected with forest genetic studies. Efficient computer use requires that both measurement and record-system techniques be designed with computer computation in mind. This can save up to 90 percent in data-preparation time. Following are some helpful generalizations which have come out of the past few years' experience.

Data are presented to a computer as numbers in rows aid columns. They can be recorded in the same manner--there is no need for decimal points, slant lines, letter symbols, etc. Records in which it is necessary to look from one page to the next to make a comparison are especially troublesome.

Use of a computer does not lessen the need for assessing the usefulness of data prior to collection. Willy-nilly processing of data without thought leads to large increases in interpretation time.

Port-a-punch cards are probably not useful in genetic studies. A moderately good typist can type tape or cards from written sheets at the rate of 10,000 characters per hour. Thus, the port-a-punch cards can possibly save no more than 1 hour for every 10,000 numbers recorded. Actually, their use involves additional rather than less time. The same principle applies to mark-sense cards. Ruled sheets of paper are apparently the most efficient recording media.

Currently there are several large geographic origin tests which are represented in each of several different localities. The measurements and analyses of the individual plantings probably will be made by the individuals who established them, but it is desirable to make overall summaries also. If the analyses of the individual plantings are complete and properly done, the overall statistical analyses often can be accomplished in one or two days. However, to do that, the method of analyzing and reporting the individual plantings must be thought out carefully.

1/ Member of the RECAP Subcommittee and Associate Professor, Forestry Department, Michigan State University. The Chairman of the RECAP Subcommittee is Paul O. Rudolf, and the other members are Dean Einspahr, H. L. Mitchell, and Stephen H. Spurr. Since there are no other reportable activities, the above paper constitutes the Subcommittee report.

a) Each person should report in columns if the others report in columns. It is desirable but not necessary that the same measurement-set be recorded in the same column by all. It is not necessary for all investigators to use cards or tapes as input for the computers for their results to be combined easily.

b) The variety-sequence in which data are reported should be the same for all. Suppose one investigator included 108 varieties in his planting and the second only 72 of those 108 varieties . The second investigator should report for all 108 even though he inserts dashes for the 36 which are missing from his field test.

c) Investigators need not measure the same traits at the same time to make useful comparisons. However, if they do measure the same traits, they should use the same units and approximately the same accuracy of measurements. Comparisons are very weak if one man recognizes 2 color grades and another person recognizes 10. There is considerable confusion and increase in time if it is necessary to do much conversion from one set of units to another.

d) Except in special circumstances it is necessary when making overall summaries to have access to all the data gathered by each cooperator. This includes the plot means (recorded by variety and trait) and varietal totals (totals certainly, varietal means possibly) for each measurement set (including the degrees of freedom, sums of squares, mean squares, F values), and the complete analyses (b's, r's, variances, covariances) when correlating one trait with another. With all computers there are facilities for easy duplication so that there is no reason why an individual investigator cannot supply all data to the central agency doing the summarizing.

Space is too limited here to give all the reasons for requiring that all these data be supplied to a summarizer. However, it can be pointed out that an investigator supplied with all the data and analyses need not go through any unnecessary duplication of effort when he decides to combine data if he has everything in front of him. For example, varietal totals are needed if one wishes to compute genotype-environment interaction. If they have already been calculated, why go through the work again ?

e) In certain perfectly regular experiments a summarizer can get by with less information. Suppose the same 20 varieties are replicated 10 times in each of 5 field tests. An investigator armed with varietal and replicate totals for each planting; plantation totals; and sums of squares for error, variety, and replicate in each planting can turn out a complete overall analysis of variance in a half-hour or so.

5. A person who plans to use a computer for his analysis work should spend at least one week acquainting himself with the computer and its operation before going to the field to make his measurements. When this is not done, about 90 percent of the measurement-sets later prove unsuitable for analysis or in need of extensive editing before submission to the computer. On the other hand, it is usually possible to take measurements in such a manner that a week's data can be analyzed in 2 hours at the end of the same week.