An Improved Record System for Forest Genetic Nursery and Plantation Studies

By Jonathan W. Wright[1])

(Received for publication September 10, 1968)

Introduction

Record keeping is one essential of a long-term tree improvement program. Most experiments require many years to complete and the person who plans the work often retires before the final report is written. A good record system should include enough detailed information to permit a researcher to make a meaningful analysis several decades after the trees are planted. But it should not be so cumbersome that the researcher loses patience looking for facts which were never recorded.

This is an account of the record system used in Michigan State University's tree improvement program. It was developed to keep track of a large volume of measurement data (sometimes 100 sets for one experiment) for more than 25 provenance or progeny-test experiments and for more than 120 permanent test plantations. It has been successful in that the measurement work has not become burdensome even though the measurement load increases each year. Also, it has been possible to keep records current so that any experiment can be summarized quickly. Some of the system's features are: completeness with minimal recording of useless data, minimal recopying, ease of information recovery, provision for prompt statistical analysis, provision for easy duplication of the most important data, and conciseness.

I am indebted to E. J. Schreiner of the Northeastern Forest Experiment Station for many of the ideas. Also to Scott S. Pauley of the University of Minnesota who originated the practice of mass dissemination of certain origin and plantation records which need to be made only once and remain valid for many years thereafter.

The successful gathering of data is partly a matter of the format in which the records are maintained. Equally important, however, are experimental design and the manner in which the measurements are made. So I will describe all these phases as they affect the record system in toto.

Effect of Experimental Design on Record Taking

Arrangement of trees and plots affects the ease of record taking. The type of design does not, however. I have phrased all examples in terms of a randomized complete block design with 4 (nursery) or 10 replicates (permanent field tests) but minor changes in fornat can be made to accomodate other designs.

Nursery experiments. — We perform nursery experiments to supply juvenile data and also planting stock for permanent tests. It has proven best to separate the two functions. Measurements are made on a replicated experiment with small plots of uniform size. This reduces the amount of measurement and usually improves quality of data. If extra stock is needed (as when supplying trees for plantations in several states), it has been grown in separate seedbeds and not measured.

In a measured nursery experiment it is advisable to lay out carefully (within ±½ inch) long seedbeds, to establish plots transverse to the seedbed axis (we usually use a single 4-foot row as a plot), and to locate the plots accurately (again within ±½ inch) within the seedbeds. The plots are mapped in terms of distance rather than sequence.

The schedule when preparing a 4-replicated nursery experiment is: (1) number four envelopes or bottles for each seedlot and check numbers, (2) place seed in each container, (3) allot one container for each seedlot to each replicate and randomize within replicstes, (4) maintain containers in this sequence and on each place a location-within-seedbed number (in feet), (5) prepare nursery maps in triplicate, (6) prepare seedbed labels, (7) prepare seedbeds and distribute seedbed labels, (8) distribute seed containers to proper rows and make final identity check in comparison with the nursery map, (9) sow seed. The nursery "maps" are in reality lists on ruled paper, one seedlot per line. They serve also as measurement forms.

This schedule requires no sorting once randomization is completed. Also, it permits work to be done by several persons but checked by one. Insistence on accuracy in seedbed and plot location usually requires little extra effort.

Permanent test plantations. — Many of our test plantations have irregular outlines or irregularly shaped replications, in order to make the environmental conditions within replications uniform. This has not affected mapping or record taking. But we do require that the trees be spaced regularly (within ±6 inches) and that every plot contain the same number of trees or tree-spaces. This can be done by marking lines on the ground or by guiding upon two stakes at the end of each row and column. It can not be done by using long string lines or by inexperienced planters who aline on each other. Failure to maintain proper alinement can increase measurement time 2- to 20-fold.

Shortcuts in Measurement and Recording

The success of records depends on the way in which measurements are made and recorded in the field. No matter how efficient a record system is in theory, it degenerates in practice unless used skillfully.

Measurement teams. — It is more important to consider the time needed for field measurement *and* office analysis *and* interpretation than to make the field measurements in the cheapest possible manner. It has usually been poor economy to rely heavily on unskilled labor. At best such persons need close supervision to produce meaningful results. Otherwise they miss important observations or produce data which is difficult to analyze.

One man working alone is often able to work as well and as fast as a pair. However, two- or three-man teams have sometimes been used, partly for efficiency and partly as a matter of training. When working as a team it is best to have each person do the same work throughout to eliminate observer bias as a source of error.

Selection oj characteristics for measurement. — In treating this subject I make no brief for either theoretical or practical data but consider that new (preferably positive) information should be the goal. It is often a waste of time to make a detailed set of measurements which merely prove that statistically significant differences do not exist or

[1]) *Professor of Forestry, Michigan State University, E. Lansing, Michigan. This paper has been accepted* as *Journal* Article *No. 4487 from the Michigan Agricultural Experiment Station.*

that differences are the same as observed previously. Avoidance of such wasted measurements is the best single way to save time. Following are five guidelines.

(1) Observe plantations frequently but avoid a measurement schedule which requires measurements at stated intervals. Instead, measure when differences appear.

(2) Measurement may be unwarranted if there are no easily visible differences among at least some adjacent pairs of plots.

(3) Do part of the statistical analyses promptly and let each day's measurement be guided by results obtained the previous day. Suppose that *a priori* evidence indicates a low probability of significant differences in each of three traits and that actual measurement confirms the absence of such differences in one of them. Important differences in the other two are unlikely.

(4) A probability analysis of qualitative data obtained from a rapid traverse through a plantation can be helpful. As an example, record the N tallest seedlots (of a total of T seedlots) in each of R replicates. The chance probability that one seedlot is selected in all cases is $T(N/T)^R$. If this is less than .05, height differences are probably significant and measurement is warranted. With practice this probability technique can be used on easily observed traits to forecast the approximate significance level to be expected after measurement.

(5) Determine whether new measurements are warranted by observation of extreme seedlots whose previous history is known.

One additional point can be made. Measurements of "branchiness" or "log quality" are frequently worthless genetically whereas measurements of "branch size", "branch number", etc. may have considerable value. Genetic data becomes more meaningful the closer it comes to describing traits controlled by single genes.

Use of plot means only. — Approximately 95 per cent of the data gathered on variability within plots is neither desired nor used. Its recording can be avoided by the "deviations-from-assumed-mean" technique, as follows: assume a mean, measure deviations and add them mentally, divide by the number of trees per plot, and add to the assumed mean.

There are other variations. When working with 4-tree plots, measure in 4-inch units, total the measurements and record the total as the average in inches. When working with bud color assign a gradenumber to each tree, add and record the total. No confusion results if the grade numbers are defined in terms of 0, 4, 8, etc. rather than 0, 1, 2.

Use of whole numbers only. — For three reasons, we require that data be recorded in terms of whole numbers only (or dashes for missing observations). There can be no decimals, fractions, slant lines, or letters. First, it is simple to do. Second, data must be presented to a computer in this manner. Third, this requires forethought about analysis and interpretation.

This may be accomplished in several ways — record the number of dead trees per plot, measure height in tenths of feet rather than feet, assign grade numbers to color or amount of insect attack, and record separate but related data in different columns.

Measurement of one trait at a time. — The "deviations-from-assumed-mean" technique and some other shortcuts can be used only if one trait is measured at a time. This practice may also result in greater accuracy. That is true when working with color because all observations can be made quickly before the light changes. It is also true when measuring cone set or insect attack because the observer looks for one thing only and does not become forgetful.

Concentration on one trait at a time requires more time to walk between trees. In one plantation which contains 150 plots (= 600 trees) one must walk approximately 4800 feet to visit each tree. That requires about 15 minutes, less time than is needed to change from a diameter tape to a height-measuring rod on each tree.

Choice of measurement unit. — A useful rule-of-thumb, learned from Dr. Sheldon Reed of the University of Minnesota, is to select a measurement unit equal to 1/10th to 1/20th of the range between extremes. Some statisticians prefer 1/24th of the range. This means less than usual accuracy when measuring height, diameter, leaf length, and many chemical traits. This results in smaller numbers and a saving in time, particularly in statistical analysis. Rarely is there a loss in useful information, because of the inherent variability in biological material.

This rule also means greater than usual accuracy when measuring color, cone set, amount of pest damage, etc. Two to four categories are usually used but 10 to 20 would be better, even when differences are small and difficult to recognize.

When dealing with purely qualitative traits (death, presence of cones, etc.) it is most convenient to choose a unit so that the majority of the numbers are small. Thus, count the number of dead rather than living trees, the number of trees with cones rather than those without cones, etc.

Should damage after a late spring frost be recorded in terms of growth loss, percentage of foliage damaged, number or per cent of buds damaged, or amount of damage per leaf? The only general rule applicable to such a situation is to make sample interpretations from a few data obtained with each of several different measurement units.

Proper interpretation and convenience are most important in choosing measurement units; readability is most important when presenting data for publication. Conversions, as from feet to centimeters, are easy, and it is unnecessary to measure and publish in the same units.

Use of transformations in the field. — It is simple to record "number of trees per N-tree plot" instead of "number of trees per plot". That is, record a "4" in a 4-tree-per-plot experiment if 3 of 3 living trees exhibit a trait. This allows direct comparisons between plot or seedlot totals without conversion to a percentage basis.

Other transformations are successful if they result in more normally distributed data or if they reflect the genic basis for differences better than direct measurement. Some are simple enough to use directly in the field. That was the case when measuring fruitfulness in a *Pinus sylvestris* experiment. We recorded the square root (± 2) of the estimated number of cones per tree instead of counting them. That was done primarily to save time but also resulted in more useful information inasmuch as a transformation would have been needed on the cone counts.

Recording visual evidence rather than cause. — Such a thing as "winter damage" may consist of bud death, cambium death, necrotic areas on the leaves, or merely a reduction in growth. But there are other causes for each of these symptoms. If data are recorded only in terms of cause, they become worthless if the guess as to cause is wrong. They are much more useful if recorded in terms of what is actually observed.

Measurement of a few trees per plot. — When variation follows a normal curve, the height of the tallest tree bears a close relation to the average height of that plot and this relationship becomes more reliable, the greater the plot size. The same is true of other traits. The validity of such a relationship must be determined anew for each experiment and trait. This can be done by measuring all trees on 50 plots and a few trees (the one tallest, the two tallest, the tallest on each quarter-plot) on the same plots, and calculating the correlation between the two sets of data. We have tested the shortcut several times and generally found a 1 to 2 per cent loss in information when dealing with a quantitative trait such as height. The loss of information has been much greater when dealing with qualitative traits.

Record neatly and completely. — It is — but should not be — necessary to emphasize neatness.

It has proven advantageous to insert abbreviated column headings on each page of the field records. This facilitates checks on the previous history of a plot measured some years before and helps to insure that data are recorded in the proper columns. It has also proven desirable to require the insertion of an entry in every box. This is either a number (often a zero) or a dash (no opportunity to observe).

The Plant-Material Accession Record

Some 5,000 lots of plant material are included in Michigan State University's tree improvement project. To maintain their identities each is numbered and described in an accession record. The original pages are prepared directly on mimeograph stencils, after which 125 copies are made. Some complete sets are distributed to fellow workers as insurance against loss. Partial sets applicable to single experiments are distributed with planting stock and are placed in the folders for individual experiments. This enables any one who has occasion to measure a plantation to interpret the data meaningfully.

The accession record is made as complete as possible, to include *all information* applicable to the plant material. In theory at least, reference to correspondence should be unnecessary. Among other things, completeness means name and address of collector and sender, date of collection, accurate source location. In the case of control-pollinated seed it also means name and number of both parents, date of pollination and seed yield per flower. In the case of provenance-test material it may mean information on soils, climate, and elevation. The need for completeness can hardly be overemphasized. Future investigators are at the mercy of the accession-record compiler.

The University of Minnesota and Petawawa Forest Experiment Station follow similar standards in their accession records. Theirs, however, are made and distributed separately for each experiment.

The types of plant materials vary and no one format suffices. However, a few general rules aid in preparation and use of the accession record. (1) Assign blocks of numbers to particular experiments so that similar materials can be numbered consecutively. This facilitates interpretation and makes it possible to drop one or two digits when preparing maps or labels. (2) Describe general features of a set of material in one place only before proceeding to detailed descriptions of individual accessions. (3) Present data in a concise tabular form, when possible. When there is too much information to use a tabular form, use letters A, B, C, etc. to denote classes of information such as state of origin, county of origin, town of origin, etc. (4) To identify accessions, use single numbers, and insist that the originally assigned numbers be used wherever the material is tested. (5) Duplicate in such a manner that the records can be handled conveniently in the field and office. Large fold-out tables and extra-length paper are unwieldy.

Plantation Maps and Establishment Reports

The functions ordinarily assigned to two separate records are combined into a 1-page (sometimes 3 or 4 pages for a large plantation) combination plantation map-establishment report. In one place this gives another person enough information to locate the plantation, measure it, and interpret the data meaningfully without reference to anything else except the accession record. The original copy is prepared *on the day of planting* and within a week 125 copies are prepared by means of mimeograph stencil.

Each establishment report contains the following information: (1) plantation number, species, and purpose; (2) age and history of planting stock; (3) exact location and instructions on how to reach the plantation with the aid of a generally available road map only; (4) time of day, date, and number of hours; (5) personnel — laborers as well as supervisors; (6) description of weather conditions, soil, condition of planting stock, vegetative cover, and all other factors which may influence tree growth; (7) statement of plot size and orientation and of spacing in both directions; and (8) statements of planting methods and equipment used. We attempt to err on the side of completeness because only a few extra minutes are required. As a general rule the poorest plantations are accompanied by the fewest data on establishment.

We use the following rules to guide the preparation of plantation maps. (1) Include a directional arrow. (2) Map all trees by seedlot number, including border and filler stock. (3) Map after planting, making no attempt to plant certain seedlots in pre-designated locations or to keep track of plots as they are planted. Those procedures are confusing. (4) Map the plantation as it is, not as it should have been. (5) Designate every tree-row (including borders) by a number and every tree-column by a letter (or vice versa). This is the easiest way to designate single trees which may be used later for special purposes such as breeding. (6) Map plots rather than single trees. A 4-tree row-plot may be shown as occupying Row 2, Columns B-though-E, etc. (7) Show replicate boundaries and numbers directly on the map, not by arrows alongside. (8) Prepare the map by typewriter even though this means an exaggeration of either the horizontal or vertical scale.

Several of these rules are ones applicable to good map-making practice in general — complete well-made maps are always the most useful. The stipulation that a typewriter be used is a matter of saving approximately 75 per cent of the time involved in hand lettering. It also serves to emphasize the fact that the plantation should be laid out in such a manner that its map can be prepared by typing.

Actual and theoretical contents of a plantation frequently differ. It is necessary to determine the actual number of plots of each seedlot in each replicate before doing any statistical work, and this list can be prepared as easily at the time of planting as later. Why not prepare it then? This list is a third component of the combination plantation map-establishment report so that it is always available when needed.

Duplicate Record Folders for Single Plantations

Preparation and contents. — Immediately after planting is completed duplicate record folders are prepared by means of carbon paper. Each page is numbered and identified by plantation number. Also, the pages are punched (at the top in our folders) and are bound by means of metal fasteners.

One copy of each folder is kept at the experimental forest where the plantation is located. The other is kept at East Lansing. Field data may be recorded in either folder and are not copied into the other one. Instead, the statistical table (plot means arranged by seedlot and replicate, and analysis of variance) is prepared in duplicate in the office. A carbon copy is mailed to the holder of the other folder. In this way both folders are complete

It is frequently necessary to disassemble one of the folders. Therefore, several consecutive pages are glued together at the top so that they may be handled as a unit and rapidly reassembled.

This particular bound-folder system has three major advantages. All previously recorded data are available whenever needed, in either the field or office. Virtually no time needs to be spent in assembling the records or forms for any measurement job. And whatever data have been recorded or analyzed are instantly available.

One possible disadvantage suggests itself, but has not been serious. That is the possibility of loss because all the original data are taken to the field each time measurements are made. That has been met partially by the carbon-paper summaries of statistical data and partially by periodic publiction of the most important results. Care, however, has been the main factor. Between Dr. E. J. Schreiner, with whom I worked for several years, and myself there is a continuous 44-year record without loss.

The principal contents of each folder are as follows.

1. Manila cover, with a brief description typed on the tab.
2. Plantation map glued on the inside back cover, where it is most convenient.
3. Road map (or topographic map) showing location of the plantation.
4. Field data sheets, described later.
5. Note sheets, described later.
6. Statistical summary sheets, described later.
7. Copy of the accession record pertaining to the material in the plantation, correspondence pertaining to the plantation, and extra copies of the plantation map.

Field data sheets. — These are prepared in duplicate on a good grade of ruled paper. Each plot is given one line and each set of identification or measurement data is given one or two columns. We use 8 × 11 inch paper with 36 lines and 42 columns per page, and each page is sufficient for 20 to 30 sets of measurements on 36 plots.

Preparation involves insertion of the following information in the left-hand columns: row number, column letter(s), accession record number, replication number, and an abbreviated description of the material to aid in field interpretation of results *(Figure 1).*

At the time of measurement the right-hand columns are completed by inserting a character-number and abbreviated description of the measurement (Ht.-'67, Dead trees-'67, Lf length-'67, etc.) at the top of the appropriate data column on each page, after which data are recorded in the column. It has proved unwise to determine what measurements will be made far in advance and to print the column headings.

Each set of data sheets is used for several years and previously collected data on a plot are available for comparison with each measurement. To avoid bias it is often desirable to ignore a plot's identity until measurements are made and recorded. When new data are required, they are prepared as described except that data on mortality or the most recent heights are added to aid in identification.

Sequence in which the plots are listed is important. We have found it most convenient to start at one corner of a plantation, walk across the plantation (perpendicular to the plot's long axis) to the other side, and to reverse direc-

Row	Tier	Seedlot No.	Replicate No.	Dead, 11/16/61 1	Lammas, 11/16/61 2	Color, 11/16/61 3	Ht. Ft X 10, 7/24/62 4	Color, 11/8/62 5	Aphids, 11/8/62 6	Winter burn, 5/1/63 7
2	B-E	6	10	0	3	4	32	4	1	1
3		9	10	0	0	3	24	4	0	1
4		24	10	1	0	4	23	4	0	1
5		13	10	0	1	2	32	3	0	1
6		28	10	1	4	4	18	2	0	1
7		29	10	0	1	2	15	2	0	1
8		3	10	1	0	4	35	4	0	1
9		10	10	1	1	3	20	3	0	1
10		1	10	1	1	3	23	4	0	1
11		20	10	0	2	4	18	4	0	1
12		12	10	0	2	3	18	3	0	1
13		21	10	0	2	3	17	3	0	1
14		25	10	1	3	4	27	2	2	1
15		14	10	1	4	5	32	4	0	1
18	F-1	12	9	0	3	4	22	3	0	1
17		28	9	0	3	4	22	1	0	2
16		28	8	1	4	3	25	1	1	1
15		24	8	1	2	5	24	3	0	1
14		3	8	0	1	4	25	4	0	1
13		10	8	0	2	3	23	3	0	1
12		13	8	0	2	4	24	2	0	1
11		19	7	2	2	2	17	1	0	1
10		12	7	0	3	3	25	3	0	1
9		28	7	0	3	2	25	1	0	1
8		6	7	0	3	3	37	2	0	1
7		25	7	1	4	4	19	2	0	1
6		19	6	2	4	2	14	1	0	1
5		1	6	0	1	4	30	4	0	1
4		3	6	1	0	4	31	4	0	3
3		20	6	1	3	3	22	3	0	1
2		24	6	0	1	4	23	2	0	1

Figure 1. — Sample field data sheet, which would ordinarily be prepared with a ball-point pen or hard pencil. The left-hand columns are prepared immediately after planting. The right-hand columns are completed as measurement progresses over a period of years. More complete descriptions of characters are included in another part of a record folder.

tion on the return trip. It is also desirable to devote one line to every plot-space even though the trees are missing. This makes it possible to devote maximum energy to measurement rather than orientation.

Duplication of the data as they are recorded on the field data sheets would be difficult and is rarely attempted. To serve a useful purpose, it would have to be done anew after each set of measurements, the old copy destroyed and the new one put in its place. Insurance against loss would be the main objective and is supplied by duplication of the statistical summary sheets.

Note sheets. — Immediately following the field data sheets a few blank pages of paper are inserted on which to describe the measurements taken. Each set of measurements is identified by a number, a date, and people who did the measurement. Then the method and units used are described in sufficient detail to enable others to repeat the procedure exactly. Such a description may contain a few words or several lines.

Also included are general notes as to the condition of the plantation, special treatment needed, etc. Or possibly such a note may show only that the plantation was visited and that no differences were apparent.

Statistical summaries. — For each trait a statistical summary is prepared in duplicate by transcribing the measurements from the field data sheets. In this summary each seedlot is given one line and each replication one or two columns.

The problem of missing plots is handled immediately in one of two ways. (1) If a seedlot or a replicate contains several missing plots, a line is drawn through the data, which are not included in subsequent statistical calculations. (2) The seedlot mean is entered and underlined (or enclosed in a box) for the few remaining missing plots, always less than 3 per cent of the total. This shortcut, if used when the percentage of missing plots is small, does not measurably affect results.

Seedlot and replicate totals are then entered in a right-hand column and number the replicate columns. Each is calculated to include missing-plot data and is thus as useful as a mean in making comparisons; hence means are usually not calculated. Then the analysis of variance is calculated and the results entered immediately below the seedlot × replication transcription. Least significant differences and variance components are usually recorded also.

The transcriptions are complete but concise (often three traits per page) records for each trait. As each one is made (preferably within a few days of measurement) the carbon copy is labelled and sent to the holder of the duplicate folder to be glued in the proper place. Thus the two distant folders are maintained complete and up to date.

The same plant material is often included in each of several different plantations. In such cases the transcription pages are partially completed in advance for all plantations, with the seedlots in exactly the same order for all. This facilitates later summarization work.

Special formats, which need not be described here, are used for other types of statistical work (usually regression or correlation) done on data from a single plantation. Again, such work is done in duplicate and bound in the folders together with the more routine analyses.

The Computer's Role in Record Keeping

This record system is fully compatible with mechanical data processing. It is possible to proceed directly from the field to a card-punch machine, prepare a deck of data cards from the field data sheets, and submit the deck for analysis. The only precaution is to write neatly and in columns.

The use of "mark-sense" and "port-a-punch" cards in the field is impractical. A good card-punch operator can punch 15,000 to 25,000 data entries per day from neatly written field records. More than one extra day per 15,000 data entries is needed to handle and record on punch cards in the field. Therefore their use involves a probable loss rather than a saving of time.

Computers have obvious advantages but too much reliance on them can result in reduced efficiency. I prefer to do most simple analyses (analysis of variance, rank correlation, Chi-square but not product-moment correlation or regression) by pencil and paper. To do this, it is necessary to focus attention on all phases of data gathering and interpretation. That in itself suggests many shortcuts which would not be obvious if all the data were mechanically processed.

Capabilities of the Measuring and Recording System

No account of this system would be complete without reference to its capabilities. In a well laid out plantation mortality counts can be made and analyzed in a few seconds per plot. Measurement of foliage color, always done in terms of live-tree standards, often proceeds at the rate of 300 to 500 plots per hour whether one or two men are used. Height measurements of 3- to 15-foot trees usually proceeds at the rate of 60 to 75 four-tree plots per hour. Up to 2 or 3 minutes per plot may be needed when detailed counts are made or leaves must be collected and measured.

When mostly zeros are involved (data on mortality, fruiting of young trees, insect attacks, etc.) transcription and statistical analysis is simple and may require only ½ hour for a 1000-plot experiment. But when a different entry must be made for each plot (most measurement data), time tends to increase geometrically with size of plantation. We have several 20-origin 10-replicated experiments for which hand transcription usually proceeds at the rate of 10 to 20 minutes per trait. But 5 to 6 hours per trait were needed for a 2000-plot experiment.

The formal nature and brevity of the records has facilitated later information recovery and summarization for publication purposes. Except in the very largest experiments, one to two days usually suffice to combine several years' data from the nursery and from several plantations in such a manner as to permit calculation of the various relationships between times of measurement, plantations and tree characteristics.

Abstract

Record keeping is as important as and offers more chance for improvement than other types of research activity. The system described uses a series of bound folders prepared at the time of planting, with simple but appropriate forms for recording all data and statistical analyses. It also provides a quick method of maintaining duplicate records for storage and interpretation at different places. The system's full advantages accrue only if records are carefully made. Several shortcuts in measurement and analysis are described.