SAMPLING DESIGN AND COMPUTER PROCESSING FOR EFFICIENT NURSERY INVENTORIES 1/

Kenneth D. Ware
Gerald Grebasch
David A. Hamilton, Jr. 21

1/ Journal Paper No. J-5791 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 1457.

2/ The authors are, respectively, Associate Professor of Forestry, Iowa State University, Ames, Iowa; Nurseryman, Iowa State Conservation Commission Forest Nursery, Ames; and Research Assistant, Iowa State University. The numerical work and computer programming on which this paper is based was done by Hamilton, under the direction of Ware, on an Honor's Project in the Iowa State University Honor's Program in which Hamilton was a participant as an undergraduate.

INTRODUCTION

How many saleable trees do you have? This is a question often asked of nurserymen. Most of the time, they are hard-pressed for an adequate answer.

That alone may be enough reason to be concerned about nursery inventory procedures. But, we can also get from a well-planned inventory system, good estimates of how many plants we have growing, their quality, and their responses to cultural operations -- and just get a better general understanding of what is going on in the nursery. The inventory must, of course, be tailored to the particular nursery and the information needed for its management.

Consider, for example, the Ames Nursery, where we are currently growing 40 species (some with several age-classes available) and are selling small quantities directly to the public. In addition, we have a rigid size criterion for salability and a heavy soil with resulting cultural difficulties. The compounding of all of this makes it very tough to estimate what will be available. However, situations such as ours are not the only ones in which inventories are of value. Inventories may also be useful where bed-run materials are being distributed and where the interest would be in more directly estimating the numbers of plantable trees per foot of bed.

Previous to 1964 the inventory system was based on Chapman's (1939) suggestions. Data were collected from a systematic sample of plots, and Chapman's graphic method was used to estimate the number of plantable seedlings from the estimate of the total number of seedlings.

Why change from an inventory system that had been reduced to a routine and used for so long? What could we hope to gain from a change?

In considering alternative inventory systems we thought about the following properties, which should be present in any good system:

1. First, the procedures for field-data collection must be simple so that they can be carried out by the personnel ordinarily available. We had a relatively simple procedure for field collection, and all our changes in sampling design have only slightly modified the method of collecting field data.

2. Second, the sampling procedure and analysis should yield objective measures of the reliability of the estimates (e.g., standard errors). This was not so for our old system.

3. Third, the sampling efficiency should be high i.e., the smallest amount of expenditure (data) should be used to obtain the desired precision. This requires that it be possible to calculate the sample size required, and the allocation of the sample, to achieve a specified level of precision. (We believed, for example, that by stratifying our population we could increase the reliability of the estimate over that from the old system.)

4. Fourth, the procedures for computation of estimates should be as efficient as possible. Under our old system, it took most of a month to take the raw data from the field and to translate it to usable facts and figures. With our new system it took 49.5 seconds with the IBM 360 computer, at a cost of $3.66. (With key-punching and other expenses added, our total cost of computation amounted to $22.33.) Does that sound like a good cost saving? Because of the additional amount and accuracy of information that is available sooner, it's even better than that.

This is all fine. But let's not stop now to rest on our green thumbs. Any inventory system, to be efficient in providing the most useful information, must be alive and growing. We are already looking to making changes -- involving such things as permanent plots, subsampling estimates of lifting losses, etc. And each year, we are adding to a large data bank that we can use to completely re-evaluate the system in five years. For the first time, we will have really useful data for evaluating the performance of the inventory.

In short, we do not intend to convince you that we have a package that is the final answer to your specific needs -- though it might go a long step in that direction. We do hope to convince you that you could proceed as we did to improve your inventory system -- even though you might finally choose a somewhat different sampling design. Let us give you more details of what we did, why we did it, and what we found out.

SAMPLING PROCEDURES FOR INVENTORY

Let us begin by considering what has usually been done in the past. We should then briefly mention the results of comparisons of the effectiveness of various alternative sampling procedures, for the Ames Nursery. Finally, it will be useful to discuss some details of the particular sampling design currently used at the Ames Nursery.

History

We do not know how uniform the inventory needs or procedures are -- or have been. We have made no canvass to attempt to discover this . However, literature about inventories of forest nurseries is particularly scarce, which indicates that little attention has been given to the sampling aspects of nursery inventory.

One of the earliest papers about nursery inventory was mainly devoted to a procedure for estimating percentage of plantable seedlings (Chapman, 1939). Then a more extensive, and often cited, study was carried out by Johnson (1941, 1943). He treated the problem in more general terms and gave the results of actual field tests of simple random, systematic, and stratified random sampling plans at the Ames Nursery. That study evidently has been the basis of the inventory procedures at many nurseries. Schumacher and Chapman (1942) also gave some attention to nursery inventories. And, more recently, Canadian researchers evaluated several sampling designs and made recommendations for conditions similar to those represented by the tests (Mullin, Morrison, and Schweitzer, 1955; Mullin, 1964). Barton and Clements (1961) described a design involving systematic sampling of plots with systematic subsampling of individual seedlings to determine percentage of plantable seedlings. Mugford (1962) then made a comparison between the Barton-Clements and Canadian procedures. From the viewpoint of statistical sampling in none of these studies have a sufficiently broad set of alternatives been treated, nor has sufficiently definitive evidence been accumulated about even the restricted set treated. It is clear that further testing and modification is desirable.

Until 1964, the sampling procedure at the Ames Nursery was based on the work of Chapman (1939) and Johnson (1941,1943). The total number of seedlings was estimated for each of the species from systematically selected plots. (Plots were 1-foot-wide strips across the nursery bed, and the sampling interval was usually about 75 feet.) The percentage of total seedlings plantable (salable) was then estimated indirectly by using a separate (independent) small sample and the graphic technique described by Chapman (1939).

For reasons indicated in the introduction (viz. , lack of objective measure of reliability, lack of way to objectively control precision, lack of evidence that sampling procedure was most efficient alternative now, etc.), it seemed desirable to re-evaluate the procedure and to compare it with some alternatives.

Comparison of Alternatives

This is neither the time nor the place to discuss the details of the estimators 3/ we decided to compare, nor to enumerate the reasonable alternatives that we did not compare. The few alternative sampling designs we considered are known to mensurationists and statisticians as:

1. Direct estimation,
    a. Simple random selection
    b. Stratified random selection
2. Indirect estimation,
    a. Double sampling, simple random selection
        1. Regression estimator
        2. Ratio-of-means estimator

3/ The estimator is the relationship, a function of the sample data, that is used to make the desired estimate for the population from the data collected in the sample. A sampling design consists of a sampling rule and an estimator. The sampling rule specifies very distinctly how the sample is to be drawn. The estimator then specifies an appropriate way to make the estimate from the data collected according to the sampling rule. There are often several different appropriate estimators , each with somewhat different statistical properties, that may be used with a given sampling rule.

3. Mean-of-ratios estimator

4. Binomial percentage estimator

b. Double sampling, stratified random selection

1. Regression estimator

2. Mean-of-ratios estimator

Those interested in the details about these sampling designs may refer to the very readable Elementary Forest Sampling (Freese, 1962), the standard reference by Cochran (1963), or the classic by Schumacher and Chapman (1942). For others, perhaps the ensuing discussion will provide an adequate introduction to the general ideas.

In the first phase of each of the several sampling designs involving double (two-phase) sampling, we either drew a simple random or stratified random sample of plots to estimate the total number of seedlings, plantable or not. In the second phase, we either drew a small random subsample of plots from these, or an independent random sample, and used this small sample to estimate the relationship of number of plantable seedlings to total number of seedlings. (These plots were lifted and the seedlings graded.) The two phases together then yielded (via the relationship, such as ratio, percentage, or regression) an indirect estimate of the number of plantable seedlings.

Our objective was to take data collected from the ordinary annual inventory and, from them, to determine which of these procedures seemed the most efficient one for estimating the total number of plantable seedlings (by species) in this particular nursery. The comparisons were based on applying each of the several sampling rules and estimators to the inventory data for 1964, making an estimate of the total number of plantable seedlings (for several of the major species), and then making an estimate of the standard error of this estimate.4/

The results of this single application indicated that stratified sampling with a mean-of-ratios estimator gives the estimate with the smallest standard error. The analogous schemes involving regression and ratio-of-means estimators were also quite effective. Further work would be necessary to learn the complete story of the performance of any of them in a given situation. Though the evidence we have here is by no means definitive, it did provide a place to start in improving the sampling design for this inventory.

In accordance with the outcome, we tentatively adopted a sampling design that involves stratified sampling with mean-of-ratios.5/ estimators based on subsampling from each stratum. The basis of the stratification is a map of the nursery that the nurseryman prepares from a visual inspection by walking through the nursery. Within species, areas that have relatively uniform stands of seedlings and relatively uniform quality are

The standard error of an estimate is an objective measure of the reliability of the estimate. It gives a numerical measure of how the sample-based estimates vary around their average in repeated sampling. Therefore, it permits us to use the statistical theory of probability to make inference to evaluate these standard errors where (where formulas existed) the ordinary ones that appear in sampling textbooks (e.g., Cochran, 1963; Freese, 1962). However, for some estimators it was necessary to extend the theory to make it appropriate for double sampling. Interested persons may find the details in an unpublished paper by Hamilton and Ware ("Sampling Efficiency for Inventory of a Forest Tree Nursery", 28 pp., June, 1966), available on request.

The stratified sampling simply involves arbitrarily dividing the nursery into areas that are relatively uniform within, but different from neighboring areas, and sampling separately from each area. A separate average of the ratios of plantable seedlings to the total number of seedlings is then calculated from the subsample selected in each stratum.

mapped to fall within one stratum. Areas that seem to differ from this, on either a stand or quality basis, are mapped into another stratum. A particular stratum may contain parts of several beds. The number of strata is flexible and can be changed as conditions warrant. During the past two inventories, we have used this method of stratifaction, and the gains in precision due to stratification have been substantial.

We are then able, before sampling, to calculate the number of plots to take in both the first-phase (total number counts) and second-phase (proportion plantable from grading) samples in order that we may achieve tie desired sampling precision. In doing this, we use optimum allocation of plots to the two phases. That is, we mathematically determine the numbers of plots to count in the first phase and the number to lift and grade in the second phase (the latter of which is much more expensive) so that we will achieve the specified precision at the minimum total cost. This feature permits us to reduce the cost of inventory for the specified precision or would, likewise, permit us to increase the precision for the specified expenditure.

By following the sampling procedure outlined here, we are able to calculate an estimate of the standard error so that we have an objective measure of the reliability of the estimate of the total number of plantable seedlings.

It is not possible to go into further details of the sampling design here. And there is not space here to permit contrasting it with other designs described by Mullin (1964) and Barton and Clements (1961). All we hope to have conveyed just now is the general idea and framework of the plan.

COMPUTER PROCESSING

Certainly the complex (even nasty you say?) formulas that apply to the more advanced and efficient sampling designs have caused some nurserymen concern. Because of this, some nurserymen may have avoided using anything but the very simple sampling plans. It can be difficult for any but a person well-versed in sampling surveys to compute the estimates by hand from data collected according to one of the more sophisticated designs. And even the simplest designs require calculations that are a costly drudgery if done by hand. But the electronic computer provides the possibility of emancipation. It can be programmed to perform the most complex and tedious calculations at great savings. The nurseryman is then freed to spend the time improving his nursery management procedures, improving inventory designs, etc., and leaving the tedious routine to the mechanical idiot.

Advantages of Automated Processing

What are the advantages and disadvantages of automated processing of inventory data?

When we speak of processing the inventory data, we mean more than just doing the arithmetic of getting the estimate of the total -- though this is of central importance. The kind of processing we are talking about takes raw field-data lists (transcribed directly to punched cards) and, from them, produces the final reports . These reports include tables giving the estimates of numbers of plantable seedlings according to the desired categories (e.g., species and age class, nursery block) and giving the standard error of these estimates. Intermediate results, such as separate stratum estimates, estimates of percentage plantable, etc., are also possible. The whole procedure, from raw data to the several final reports, complete with table headings and footnotes, is essentially "untouched by human hands".

The most obvious advantage of this is the savings of valuable time of technically trained personnel otherwise required to do the calculations. This saving can be very great. Routine errors of calculation are avoided completely. Reports are available sooner, too. As important an advantage, but a less obvious one, is the ability to do a complete analytical job. Standard errors can be evaluated for each estimate, auxiliary tables prepared, etc. Perhaps the most important advantage of all, however, is that more efficient sampling designs can be adopted because the nurseryman is freed from the more complex statistical calculations required for some such designs.

It is a distinct advantage, too, to have a flexible bank of data, such as is the case when they are stored on punched cards. Data from previous inventories can, if properly analyzed and used, provide very useful information for improving sampling designs for future inventories. One might argue that in no other way could we steadily progress toward the most efficient design for any particular nursery.

What are the disadvantages? The biggest one -- and it can be a very big one -- is the difficulty of obtaining a computer program that exactly fits your needs. If your objectives, sampling design, data format, report format, etc., are exactly the same as those for the Ames Nursery, then our program could be used. Otherwise you have the job of writing a computer program -- and that job requires very specific training, is demanding, and can be costly and frustrating.

How Can it Come About?

Fine, you say? Great! But you don't know any computer programmers and haven't time to become a statistician and computer nursemaid!

Wait. Before you make up your mind that this can't be brought about for you, let's explore the possibilities. We have argued, in general, that you can't afford not to try. Here's what you might try first.

If you have a forestry school (or, for that matter, any college or university) close by, then the chances are good that someone there could help you. The forest mensurationist would be the man to start with. Or if you know exactly that you want done and how to describe it in detail, you could go directly to the computation center and talk with the consultants there. These people are almost always willing to help, but some of them are great specialists in computer jargon and strange computer languages. You may have difficulty getting your message across unless you can spell it out very succinctly in mathematical terms. Actually, a computer program of this kind (for nursery inventory) is very simple -- as computer programs go. So, once you find out what you want to do, it is easy to program, and the program could be written by an undergraduate forestry student who has taken a course in computer programming. (This was the case with Hamilton when he wrote the program for the Ames Nursery.)

If most nurserymen do not have a friendly forest mensurationist nearby then several nurserymen might cooperate in studying alternatives and adopting a relatively standardized procedure (perhaps with several different options for flexibility). (The Division of Cooperative Forest Management of the U.S. Forest Service would seem the logical coordinator and source of technical guidance for any such cooperative effort. We confess that we have not discussed this with representatives of the Forest Service and don't know whether it falls within their responsibility. We may be far afield here.)

The potential gains we have enumerated seem to us to make some investigation worthwhile.

KEEPING INVENTORY PLANS CURRENT

One of the greatest problems we all have is that of keeping up with yesterday. Because of our many responsibilities, we are unable to keep our management procedures up to the level currently thought to be most advantageous. For inventory procedures, it seems particularly easy to fall behind. Once we have what appears on the surface to be all right, we are reluctant to change it. Even when objectives change, environmental and institutional conditions change, knowledge of sampling procedures change, and computational technologies change, we tend to hold fast. Perhaps, as someone has said, we need, rather than the "courage of our convictions ," that greater, and rarer, courage required to regularly scrutinize and modify our convictions.

We cannot propose a recipe for keeping inventory plans current. We assert only that a periodic examination of alternatives would be useful. And we believe that the adoption of the most flexible sampling designs, coupled with a relatively general computer program (i.e., one appropriate to several optional sampling designs) , would make it easier to keep inventory plans current and efficient.

Maybe it would be useful to give a brief indication of how we intend to keep the inventory plan for the Ames Nursery as current and as efficient as we can, practically, make it. Because of the need to know what is actually done, we have prepared a written guide that covers detailed procedures for sampling, collecting the data on the samples, and compiling the results . This means that, for any year, we know exactly what the details about the data are. Therefore, we can extract them from the data bank at any time to examine new alternatives for sampling designs or to examine ways of gaining better control over sample size and allocation -- even with the same design.

Several alternatives are already being considered for improving efficiency and providing more information at the Ames Nursery. One of these, which we think shows great promise is to use permanent plots in conjunction with the principle of partial replacement; i.e., remeasurement of an optimal subsample (see *e.g.* , Ware and Cunia, 1962). The idea of using permanent plots is not new, and several important reasons for doing so were set out by Belcher (1964) and Turner (1964). However, their reasons did not include the most important ones from the standpoint of sampling efficiency. This reason is that the permanent plots may be used in such a way as to increase the sampling efficiency through the use of partial replacement.

We cannot go into detail here. But when we are able to incorporate permanent plots and partial replacement6/ in the way we anticipate, we expect several improvements:

1. Improved efficiency, (i.e., lower cost for specified precision) in estimating the current stand of salable seedlings . This is so because, for example, the seedlings may be 3 years old at time of sale, and the observations made in previous years on the subsample of permanent plots can be used to get a much

The basic principle of sampling with partial replacement is that a subsample of the plots are permanent and are remeasured while the remainder are not remeasured (permanent). Estimates can indirectly be made for the plots that were not remeasured by using the relationship between the number of seedlings present on a permanent plot when it is remeasured and the number that were there at the previous occasion. If there is a strong relationship, then this indirect estimate will have high precision and a sampling design which incorporates sampling with partial replacement will be efficient.

better estimate at the third year -- as well as providing an efficient way of estimating for each of the first 2 years.

2. Better estimates, in any given year, of the anticipated number of seedlings that will be available in the year of harvest.

3. Good estimates of <u>changes</u> in seedling stand and quality over time. This could include estimates of germination and survival percentages, lifting losses, etc. , and other information mentioned by Belcher (1964). However, partial replacement could yield higher sampling efficiency for estimating change than would be so for either total remeasurement of all plots or independent inventories without permanent plots.

4. Greater flexibility possible in selecting future sampling plans to accommodate changes in objectives or management practices. The combination of permanent and temporary plots permits choice from several estimators, wide latitude in planning for sampling intensity, and wide latitude in allocations in various years.

5. Better data for improving future designs and for planning annual inventories . Much useful statistical information -- for example, such as about correlations and changes -- would be obtained at no extra cost in carrying out the regular inventory procedure.

In brief, we believe it is important to look ahead and to continue with attempts to improve inventory procedures. We believe also that permanent plots, used with the principle of partial replacement, may offer important advantages. There are many possibilities. The modern sampling procedures involving unequal probabilities of selection show high promise, but perhaps nurserymen may want to leave that to the mensurationists for the time being.

SUMMARY

We have attempted to point out some of the possibilities for improving procedures for nursery inventories and have discussed some aspects of sampling design. The results of trials with several procedures and the experience gained during the last two years at the Ames Nursery were summarized. Also, some advantages of automated data processing via electronic computer were enumerated.

We attempted to indicate what it might take for others to "become automated" and then stressed the importance of keeping inventory procedures flexible and current. Finally, we sketched some of the potential advantages of using permanent plots with partial replacement.

LITERATURE CITED

Barton, W. W., and C. M. Clements.

1961. A systematic sampling nursery inventory procedure. Tree Planters' Notes 46: 19-25.

Belcher, E. W., Jr.

1964. Nursery seeding and seedling history plots. Proc. Region 8 Forest Nurserymen's Conferences (U.S. For. Serv. Region 8, Atlanta, Ga.): 166-171.

Chapman, R. A.

1939. A method for inventorying planting stock in forest nurseries. U.S. For. Serv., Southern For. Expt. Sta., Occ. Paper 86. 9 pp.

Cochran, W. G.

1963. Sampling techniques, 2nd Ed., John Wiley, N.Y. 413 pp.

Freese, F.

1962. Elementary forest sampling. U.S. Dept. Agric., Handbook No. 232. 91 pp.

Mugford, D. G.

1962. Nursery stock inventory procedures. Proc. Region 9 Forest Nurserymen's Conference (U.S. For. Serv., Region 9, Milwaukee, Wis.): 9-16.

Mullin, R. E.

1964. Comparisons of sampling methods for inventory of nursery stock. Tree Planters' Note 67: 3-8.

Mullin, R. E., L.M. Morrison, and T. T. Schweitzer.

1955. Inventory of nursery stock. Ontario Dept. For.& Lands, Res. Report 33. 64 pp.

Johnson, F. A.

1941. A statistical study of sampling methods for tree nursery inventories. (M.S. thesis) Iowa State Univ., Ames, Iowa. 75 pp.

____

1943. A statistical study of sampling methods for tree nursery inventories. Jour. For. 41: 674-679.

Schumacher, F. X., and R. A. Chapman.

1942. Sampling methods in forestry and range management. Duke Univ., School of For., Bul. 7. 213 pp.

Shiue, C. J.

1960. Systematic sampling with multiple random starts. For. Sci. 6: 42-50.

Shiue, C. J. , and H. H. John.

1962. A proposed sampling design for extensive forest inventory; double systematic sampling for regression with multiple random starts. Jour. For. 60: 607-610.

Turner, E. E.

1964. Seedling inventories. Proc. Region 8 Forest Nurserymen's Conferences (U.S. For. Serv. , Region 8, Atlanta, Ga.): 163-165.

Ware, K. D. , and T. Cunia.

1962. Continuous forest inventory with partial replacement. For. Sci. Monograph 3. 40 pp.