Use of Photography in Sampling for Number of Fruit per Tree

By H.F. Huddleston

The costs of obtaining a given level of accuracy in estimating yields of tree fruit might be lowered by the use of photographs or supplemental information on fruiting potential. A small-scale experiment of this sort is described for peaches and apples. A procedure for evaluating the relative cost and efficiency of different methods of estimation is given and some provisional judgments of results are offered.

Key words: Double sampling; remote sensing; fruit yields; photography; sampling frame.

The development of new methods of estimating yields of crops typically begins with investigations based on small samples to explore alternative techniques and characteristics to be measured. One such investigation, carried out by the Department of Agriculture, concerns the potential use of photography for estimating the number of apples or peaches per tree. This paper describes the study and presents some preliminary results.

Among the tree crops to which objective sampling techniques have been applied to estimate numbers of fruit in the United States are oranges, peaches, cherries, apples, almonds, pecans, walnuts, filberts, grapefruit, and lemons.[1] It is characteristic of most tree crops that the number of fruit per tree varies widely among trees classified by age and variety in commercial orchards. Normally, the number of fruit per tree contributes much more to the variability of yield per tree than size or weight per fruit. The counting of fruit on a large sample of trees is indicated unless auxiliary data on the yield potential are available to provide more efficient estimates of production per tree. Accurate visual counting of fruit on sample limbs by field crews is costly and difficult to achieve in large-scale surveys. In addition, there may be a problem of getting fieldmen trained and disciplined to carry out objectively the sampling and fruit counting procedures.

[1] Studies of objective sampling techniques as applied to some of these crops include: R.J. Jessen, Determining the fruit count on a tree by randomized branch sampling, *Biometrics*, Vol. 11, pp. 99-109, 1955; R.P. Small, Research report on tart cherry objective yield surveys, U.S. Dept. Agr., Statis. Rptg. Serv., 1964 (unnumbered); R.R. Sturdevant, Research report on Virginia apple objective count surveys, U.S. Dept. Agr., Statis. Rptg. Serv., 1967 (unnumbered).

The investigation of the use of photography has two specific purposes:

(1) To obtain pictures of bare trees which can be used as a frame for rigorously defining sampling units for small portions of a tree, and which can provide a visual record designating sample limbs that field crews can find for making counts of fruit.

(2) To create auxiliary information on fruit set for individual trees—information that can be utilized either with counts of fruit made by field crews for a small fraction of a tree in the sense of double sampling, or as a variable which would be useful in ratio estimation. The work has progressed to a point where results for small samples of trees are available for several kinds of fruits. USDA plans to collect data for somewhat larger samples to evaluate these findings before making recommendations for operational surveys.

Constructing a Frame Using Photography

Pictures are taken early in the spring before leaves appear and may be used for several years. For each of two sides of a tree, approximately 180 degrees apart, a stereo transparency is obtained. In the office, a copy of one member of the stereo pair is reproduced for identifying the sampling units. Normally there are three branching stages for sampling a tree: (1) Primary limbs corresponding to the main scaffolds off the trunk, (2) secondary limbs originating from the primary limbs, and (3) terminal limbs branching from the secondary limbs. Terminal limbs correspond to the ultimate sampling units which are small enough to be counted by field crews in 1 hour or less. Generally the cross-sectional area of a terminal limb is 1 to 3 square inches. The total number of these units on a tree is a function of age, which is normally reflected by tree trunk size.

The photographs provide a complete identification of the limbs for the sample trees. This introduces the possibility of optimizing the sample design which otherwise would not exist.

For four alternative methods of selecting limbs, relative variances within trees of the number of fruit on terminal limbs are shown in table 1. The random-path method[3] with equal probability of selection at each stage of branching (EPS) requires only a count of the number of limbs at each branching stage; whereas random path with probability of selection proportional to size of limb (PPS) requires measurement of limb sizes at each branching stage. Single-stage method refers to direct selection of terminal limbs either by EPS or PPS after all terminal limbs on the tree have been defined. The random-path method can be used either in the field or in the office from photographs, whereas the single-stage method is possible only from a photograph (or mapping) for large trees.

Table 1.—Analyses of number of fruit per terminal limb for alternative sampling schemes

Items	Virginia peaches	Virginia apples	California peaches
Number of trees = K	9	6	16
Total number of terminal units ($\Sigma^k N_i$)	125	134	320
Variances relative to the mean squared within tree:			
Single-stage EPS	.519	.502	.341
Single-stage PPS	.293	.238	.493
Random-path EPS	.561	1.260	.924
Random-path PPS	.317	.240	.397

The within-tree variances were computed as follows:

Single-stage EPS:

$$\sigma_1^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{N_i} (N_i X_{ij} - X_{i.})^2}{\sum_{i=1}^{K} (N_i - 1)}$$

Single-stage PPS:

$$\sigma_2^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{N_i} P_{ij} \left(\frac{X_{ij}}{P_{ij}} - X_{i.} \right)^2}{\sum_{i=1}^{K} (N_i - 1)}$$

[3] See Jessen (cited in footnote 1).

Random-path EPS:

$$\sigma_3^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{N_i} E_{i,j} \left(\frac{X_{ij}}{E_{i,j}} - X_{i.} \right)^2}{\sum_{i=1}^{K} (N_i - 1)}$$

Random-path PPS:

$$\sigma_4^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{N_i} Z_{i,j} \left(\frac{X_{ij}}{Z_{i,j}} - X_{i.} \right)^2}{\sum_{i=1}^{K} (N_i - 1)}$$

where

X_{ij} = number of fruit on the jth terminal on the ith tree

P_{ij} = probability of selecting the jth terminal limb on the ith tree

N_i = number of terminal limbs on the ith tree

K = number of trees

$X_{i.}$ = number of fruit on the ith tree

$E_{i,j}$ = limbs at each of the j stages on the ith tree were selected with equal probability

$Z_{i,j}$ = limbs at each of the j stages on the ith tree were selected with probability proportional to size

$$\sum_{j=1}^{N_i} P_{ij} = 1$$

$$\sum_{j}^{N_i} E_{i,j} = 1$$

$$\sum_{j}^{N_i} Z_{i,j} = 1$$

The notation (i,j) in the subscripts to E and Z indicates that a variable number of stages, j, was required to reach the individual terminal or elementary units on the ith tree. The value of j is commonly 2 or 3, but may be 1 or occasionally 5 for a few limbs on each tree.

Conceptually, one might expect that the random-path PPS and the single-stage PPS would have approximately the same sampling error. The difference in sampling error for single-stage EPS and single-stage PPS is appreciable, but reverses the magnitude somewhat unexpectedly for California peaches (table 1). It is suspected this may be the result of the thinning of fruit to meet specific marketing order requirements to control production, which alters the correlation between limb size and number of fruit. Three alternative schemes of sampling limbs are under consideration: (1) Estimate the limb size from the photograph, (2) try to define on the photographs terminal limbs that are as close to equal size as possible, or (3) use two-stage selection, choosing primary units with PPS and terminals within primary limbs with EPS.

It is feasible and desirable to use a two-stage procedure which is a slight variation from the random-path method and which will materially reduce the amount of time required to select terminal units whether the selection is made from photographs or completed in the field. This procedure will give quite efficient estimates of fruit per tree since the size of the primary limb is highly correlated with the total fruit on its terminal limbs. The closeness of the relative variances for single-stage PPS and random-path PPS is largely the result of this relationship.

Fruit Counts From Photographs

Eight 35-mm. slides were obtained of each tree when the field crews made counts of fruit by limbs. Four slides were obtained from each of two sides of the tree, 180° apart. Each side of the tree was divided into quarters by using vertical and horizontal aluminum poles which formed a "plus" sign. One slide covered each quarter with some overlap with the adjoining quarter to insure complete coverage. Four pictures of a side were more satisfactory for counting or interpretation purposes than one picture taken with a wide-angle lens.

Some individual fruit near the edge of a tree may be seen on pictures from both sides and hence counted twice. More importantly, some fruit cannot be seen at all. However, the problem under consideration is the possible use of fruit counts from photos in the context of double sampling or eventually in lieu of physical counts, if relationships between photo and physical counts can be found which do not vary among years.

The count of fruit from slides is highly correlated with the total fruit (last line of table 2); consequently, efforts to develop a practical statistical scheme of using this information are justified. For double sampling there may be better covariates than photo counts, in the sense of minimum variance per dollar, but finding qualified people for sampling work in the field may be an equally important consideration. The task of recruiting, training, and supervising a large field crew may be more difficult than hiring a very small field crew and a group of photo interpreters for counting fruit in a double sampling scheme. Some of the results to date indicate that stable relationships between photo and physical counts can be found.

Table 2.—Correlations between number of fruit per tree and different measures of tree sizes and fruit counted on photographs

Item	Virginia peaches	Virginia apples	California peaches
Trunk size	0.12	0.89	Not obtained
Sum of sizes for primary limbs	.33	.87	0.50
Sum of sizes for terminal limbs	.42	.90	.56
Number of terminal limbs	.26	.73	.52
Fruit counted on photos	.85	.98	.85

For the six apple trees reported in table 1, the linear regression coefficients (relating fruit counts on the trees to counts on photos) and the fraction of fruit visible were computed from 1967 data. These statistics were used to estimate the number of fruit on the same trees in 1968, based on a single random selection of one terminal per tree. The estimates are shown in table 3. These fragmentary results are encouraging inasmuch as the estimates utilizing the supplementary information (first three columns of table 3) are closer to the actual count than the single-stage EPS estimator. There may be reason to hope to eliminate the need for physical counts except for verification or occasional updating of the relationships. The estimators used for table 3 were:

Ratio estimator:

$$\hat{Y}_1 = X_i \div \hat{R}$$

where X_i = fruit counted on 8 slides in 1968

$\hat{R}$ = average fruit counted on 8 slides divided by actual fruit per tree in 1967.

Regression estimator:

$$\hat{Y}_2 = \hat{a} + \hat{b} X_i$$

where $\hat{a}$ = intercept derived from 1967 data

$\hat{b}$ = slope derived from 1967 data.

Composite estimator:

$$\hat{Y}_4 = X_i + (1 - \hat{R})\, y_i$$

3

where $\hat{y}_i$ = 1968 estimated fruit per tree = $N_i X_{ij}$; based on a random sample of one limb (N_i and X_{ij} are defined on page 64).

Single-stage EPS:

$$\hat{Y}_3 = \hat{y}_i.$$

4

Table 3.—Fruit numbers estimated by four methods, 1968, and actual counts, 1967 and 1968

Tree number	Estimated 1968				Actual count	
	Ratio estimator	Regression estimator	Composite estimator	Single-stage EPS estimator[1]	1968	1967
1	434	254	415	396	602	403
2	563	413	377	196	399	214
3	820	730	651	487	758	1,658
4	596	454	630	663	746	1,575
5	789	691	944	1,094	1,075	1,901
6	1,645	1,745	949	272	2,181	1,448
Total fruit	4,847	4,287	3,966	3,108	5,761	7,199

[1] Estimates in this column correspond to the results that would have been obtained using counts by field crews.

In table 4, "sides" refer to the two sides on a tree 180° apart; "diagonal" refers to a combination of the counts from two quarters of a side, either upper left and lower right or upper right and lower left. For diagonals there is one degree of freedom per side for each tree.

Use of Auxiliary Variables

The investigation of alternative measures of fruit set per tree is in the formative stages in the research. Table 2 shows correlation between each of four variables determined prior to fruiting with fruit set per tree. The photo counts are considered in a different context since a count of the visible fruit is obtained only after fruiting has occurred.

The best measure of potential fruit set prior to actual fruiting appears to be the sum of the sizes of all the terminal or elementary sampling units on the tree. The two principal factors which seem to influence the various measures of fruit set are (1) kind of fruit and (2) age of tree. However, the number and size of terminal units appear to be a more effective way of expressing tree age. The results in table 2 show that the sum of the primary limb sizes contains almost as much information on fruit set as the terminal limb sizes. Consequently, this variable is to be preferred because of the relative ease of securing the information. Normally, a bearing fruit tree will have about five primary (scaffold) limbs whose sizes can be measured easily. While this phase of the research is just getting underway, the use of auxiliary variables as a basis for tree selection or double sampling appears promising.

Table 4.—Nested analyses of variance of number of fruit counted from photographs[1]

Source variation	Virginia peaches		California peaches		apples Virginia peaches	
	degrees of freedom	mean squares	degrees of freedom	mean squares	degrees of freedom	mean squares
Total	71	320	55	211	23	5,028
Trees	8	1,561	6	540	5	21,462
Sides	9	142	7	490	6	548
Diagonals ..	18	98	14	73	12	421
Quarters ...	36	200	28	139		

[1] Nested analysis is described in: G.W. Snedecor and W.G. Cochran, *Statistical Methods*, 6th ed., Iowa State Univ. Press, 1967.

Cost and Efficiency Implications

The development of techniques using photography or other supplementary information to reduce either the variance of the estimator or the costs of acquiring data is the principal consideration in evaluating new tools. While the exploratory studies reported in this paper do not provide a satisfactory basis for judging these factors, some hypothetical relative costs and variances indicate what conditions are required to result in greater overall efficiency.

Three schemes of double sampling are considered, based on the use of three sources of additional or supplementary information: (1) Counting the fruit on the tree from colored slides, (2) measuring tree cross-sectional area based on either trunks or primary scaffolds, or (3) constructing a sampling frame from bare tree photography to define sampling units for use by field crews for each of n trees in a block and counts of fruit on n' trees by field crews using conventional limb selections and fruit counting procedures where $n \geq n'$. For (1) and (2) we can approximate the effect on the sampling error by:

$$S^2_{\bar{x}_n} \doteq \frac{S_x^2}{n'} \left[1 - \rho^2 \left(1 - \frac{n'}{n}\right)\right]$$

where S_x^2 is the variance per tree of the conventional estimate of fruit per tree, ρ^2 is the correlation coefficient between the new information and the conventional estimate X, and n' is a random subsample from the n trees in the block.[1] For (3) the reduction in variance is expressed as the ratio of two variances such as those given in table 1. Table 5 illustrates the gains or losses that may be achieved with double sampling using regression estimation and a linear cost function ($c = c_1 n + c_2 n'$) for the same total costs. The values in the body of the table are given by

$$\left[\rho\sqrt{c_1/c_2} + \sqrt{1 - \rho^2}\right]^2$$

where c_1 = cost per unit for the information for the n first-stage units and c_2 = cost per unit for the information for the n' units subsampled.

The use of supplementary information results in gains in efficiency for the same costs for large values of ρ and small values of the cost ratio c_1/c_2. Based on the correlation values in table 2, apples are the most promising of the fruits studied. However, the counting of fruit on photos seems likely to merit further study for all fruits.

Based on the present methods of data collection, the most favorable cost ratio of the variables in table 2 is "trunk size" and the least favorable ratio "fruit counted on photos." The attainable cost ratios for these variables are probably of the order of 1/9 and 1/2. The consideration of factors other than costs and variances, such as control of nonsampling errors, and survey training and supervision, may provide additional gains using supplementary information.

[1] M. Hansen, W. Hurwitz, and G. Madow, *Sample survey methods and theory*. Vol. 2, Theory, John Wiley & Sons, Inc., 1953.

Table 5.—Ratio of variance with optimum double sampling design to the single sample for various values of ρ and c_1/c_2

ρ	Cost ratio: c_1/c_2						Gain or loss
	1	1/2	1/3	1/5	1/7	1/9	
.4	1.73	1.44	1.32	1.20	1.14	1.10	Loss
.5	1.87	1.49	1.33	1.19	1.11	1.07	
.6	1.96	1.50	1.31	1.14	1.05	1.00	
.7	2.00	1.46	1.25	1.05	.96	.90	
.8	1.96	1.36	1.13	.92	.81	.75	
.9	1.78	1.15	.91	.70	.60	.54	Gain
.97	1.47	.86	.64	.46	.37	.32	
1.0	1.00	.50	.33	.20	.14	.11	