Development of Sampling Guidelines for Estimating the Proportion of Weeviled Trees on a Plantation

Forest Service Internal Report PM-PB-18

by

V.E. Fletcher
Research Branch
British Columbia Ministry of Forests
Victoria, B.C.

January, 1986

NOT FOR PUBLICATION

(Contents of this report may not be cited in whole or in part without written approval of the Director, Protection Branch, 1450 Government St., Victoria, B.C. V8W 3E7)

634.909711
BCMF RES
1986
MR 10
c. 1 ma

CFXP
634.909711
BCMF
RES
1986
MR 10

ERRATA

Development of Sampling Guidelines for Estimating the Proportion of Weeviled Trees on a Plantation

634.909711/BCMF RES/1986/M
FLETCHER, VAL.
DEVELOPMENT OF SAMPLING
GUIDELINES FOR ESTIMATING
CFXP c. 1 ma Main........

Forest Service Internal Report PM-PB-18

Page 12, Table 6

No. of Transects required (sample size) for 80% confidence limits should read:

Bound on Error of Estimation	No. of Transects Required (sample size)
	80% Confidence Limits
0.10	3
0.10	3
0.10	3
0.05	12
0.05	12
0.05	12

Page 19, APPENDIX II

(ii) Estimated variance of p:

$$\widehat{Var}\,(\hat{p}) = \left(\frac{N-n}{Nn\bar{M}^2}\right) \frac{\sum_{i=1}^{n}(a_i - \hat{p}m_i)^2}{(n-1)}$$

Should be "n" NOT "m" as in original manuscript.

TABLE OF CONTENTS

INTRODUCTION

The intensity of leader weevil (<u>Pissodes strobi</u>) attack in Coastal Sitka spruce plantations is on the increase. The potential losses to this pest in the interior, with its extensive spruce plantations, is also cause for concern.

Information on the intensity (or level) of attack in a plantation is crucial to the pest manager for determining the timing and perhaps type of control measures to taken. A sampling methodology developed to estimate attack intensity, as measured by the proportion of attacked trees in a plantation, must provide accurate estimates while being operationally efficient.

Several sampling methods have been proposed in the rather limited literature available for estimating the proportion of diseased or insect-attacked trees on a plantation. Yandle and Roth (1971) investigated the properties of ratio estimators to determine the incidence of fusiform rust in southern pine plantations. They used individual rows of trees as the basic sampling unit. A simple random sample of rows was chosen from the plantation, with the proportion of diseased trees on the sampled rows providing the estimate for the plantation. Talerico and Wilson (1973) described the use of 0.10 acre (0.04 ha) fixed area square plots to obtain estimates of the proportion of never-weeviled trees in an effort to evaluate control of <u>P. Strobi</u> in plantations.

Gross <u>et al</u>. (1980) tested a number of sample rules for performance on estimating the proportion of jack pine affected by sweetfern rust. These rules included fixed-area strips and plots, prism (point) samples, point to tree distances and "m-tree" strips in which a constant (m) number of trees were tallied along a narrow row transect. They recommended the latter method as most efficient. In a study dealing with the assessment of spruce budworm impact, Karpinski and Witter (1982) investigated the size and number of fixed area plots needed to precisely estimate tree mortality. Concentric circular plots of various sizes together with one rectangular plot were established at 2 to 5 randomly located points on a stand for comparison.

Finally, MacLean and Ostaff (1983) investigated sample size-precision relationships for prism point samples and circular fixed area plots (0.01 and 0.05 ha in size) for estimating levels of tree mortality caused by Spruce budworm.

Of the papers cited above, only two (Yandle and Roth, 1971; Talerico and Wilson, 1973) dealt with proportion estimates alone. The remaining studies were also concerned with estimating stand characteristics such as density, basal area and volume. Traditionally, these latter characteristics have been measured using fixed area, point (prism) and point-to-tree methods. Thus, it appears the choice of these methods reflects more on the need for information on stand characteristics rather than proportion estimates.

Sequential sampling is another technique that has potential application in this instance. While it does not provide population parameter estimates (proportion of infested trees), it can classify the level of infestation as above or below a specified (critical) value (Waters, 1974). In this case, two classes would be appropriate, control versus no control for example. Also, sample size is not fixed but varies with level of infestation providing more information per unit cost at low or high infestation levels than fixed sample size schemes. A major disadvantage to sequential sampling is that it requires random selection of sampling units (trees) in the field for valid results. This requirement, from an operational standpoint, precludes it from further consideration.

In light of the above, the random row sampling technique described by Yandle and Roth (1971) was chosen for further study. The objectives of this investigation were to: (i) evaluate the performance (accuracy of estimation) of the candidate sampling methodology through computer simulation and (ii) provide preliminary sampling guidelines based on results from (i) coupled with limited field data currently available.

MATERIALS AND METHODS

(i) Simulation Methods

Computer programs were written in the BASIC language as implemented on a WANG P.C. to simulate row sampling methodology (Yandle and Roth, 1971). A hypothetical plantation was generated consisting of a rectangular array of 149 rows of 99 trees each for a total of 14,751 "planted" trees. Next, various combinations of infestation level (about 1%, 5% and 50%) and spatial patterns (highly clumped and completely random) were simulated on the plantation. Currently, there is little information available on the initial spatial pattern and dynamics of weevil spread. Subjective observations indicate clumping is evident at low population levels perhaps grading onto a random pattern at higher levels, probably due to merging of clumps. Thus, both random and clumped infestation dispersion pattens were simulated on the "plantation". Furthermore, as a true test of the reliability of the sampling estiamtors, small, tightly packed centers of infestation were modelled for the clumped situtaion rather than large, more dispersed centers.

Both random and systematic row sampling were simulated with estimates and their variances calculated according to the formulas listed in Appendix II. Under random sampling, 100 independent random samples of 5, 10, 15, 20, 25 and 30 rows were chosen for each combination of infestation level/spatial pattern. Systematic row sampling was also simulated on the same infestation configurations. Here 6 systematic samples were taken, choosing every 5th, 10th, 15th, 20th, 25th and 30th row giving corresponding sample sizes of 30, 25, 20, 15, 10 and 5 rows. For each group of 6 samples the starting row was varied from 1 to 10.

Accuracy (or closeness of an estimate to its actual value) of each sampling scheme was evaluated through measure and analysis of average bias (the difference between estimated and actual values) and average sample variance estimates.

(ii) Field Methods

Field surveys for weevil attack were conducted on selected Sitka spruce plantations on Vancouver Island and Coastal areas of the mainland in 1981-84. These surveys were not directly related to the present study as the intent was for 100% coverage of a plantation. However, the data gathered appeared useful for establishing preliminary sampling guidelines in the context of this study.

The survey technique consisted of 4 to 6 observers lining up abreast and within site of one another and "sweeping" through the plantation along visually maintained, adjacent, transects. The number of trees observed versus the number with weeviled leaders were recorded with hand-held counters. The width of transects depended on visibility within a particular plantation but, in any case, sweeps were continued back and forth across the plantation until 100% coverage was obtained.

Sample estimates of proportions of trees attacked and variances were calculated treating each recorded transect as a "row" in a plantation.

RESULTS AND DISCUSSION

Simulation:

(i) Random Row Sampling

Table 1 shows that the actual proportions of infected trees do not differ much from their corresponding estimates for sample sizes of n = 5 to 30 rows and randomly dispersed infestation. Generally, bias and variance of the estimates stabilize in the n = 20 to 25 range. The average bias expressed as a percentage of the true proportion is 2 to 3% in all but one case (p=0.00068, n=10, bias = 4%). As expected, increased sample sizes result in decreased sample variance (or increased precision). Thus, random row sampling appears to perform quite well over the various levels of randomly dispersed infestation and corresponding sample sizes. If precision of estimates can be sacrificed somewhat even small sample sizes have biases within tolerable limits. Yandle and Roth (1971) achieved similar results in their study of actual plantations in which fusiform rust attack exhibited a random pattern. They felt that even with sample sizes as small as n=3 rows bias was negligible in using the ratio estimate for proportion of infected trees.

Random row sampling also performed well under clumped infestation patterns (Table 2). Again, average bias was under 3% for all sample sizes.

(ii) Systematic Row Sampling

Systematic sampling has several advantages over random sampling including: ease of locating sampling units, better representation of samples over the entire population and greater potential for mapping pockets of infestation throughout the plantation. One disadvantage of systematic sampling is that results can be biased if the spatial distribution of weevil attack has some underlying regularity throughout the plantation - an unlikely situation.

TABLE 1. Comparison of actual proportion of infested trees (p) to estimates ($\hat{p}$) based on randomly chosen rows. Infested trees randomly located on plantation. Average values for 100 samples presented for each case.

No. of Rows/Sample (n)	Average Estimated Proportion ($\hat{p}$)	Average Bias ($p-\hat{p}$)	Average Variance $\hat{V}(\hat{p})$
	p = 0.00068		
5	0.0006707	-0.000072	0.000013
10	0.0007061	0.000281	0.000006
15	0.0006721	-0.000059	0.000004
20	0.0006702	-0.000077	0.000003
25	0.0006739	-0.000040	0.000002
30	0.0006828	0.000049	0.000002
	p = 0.0652		
5	0.065374	0.000158	0.000095
10	0.066848	0.001633	0.000050
15	0.064970	-0.000246	0.000032
20	0.066035	0.000819	0.000025
25	0.065402	0.000186	0.000020
30	0.065495	0.000279	0.000017
	p = 0.4923		
5	0.501859	0.009553	0.000544
10	0.497758	0.005452	0.000305
15	0.496115	0.003809	0.000213
20	0.497005	0.004699	0.000163
25	0.497108	0.004713	0.000130
30	0.497340	0.005030	0.000102

TABLE 2. Comparison of actual proportion of infested trees ($p=0.0319$) to estimates ($\hat{p}$) based on randomly chosen rows. Infested trees randomly located in 20 tight clusters of 25 trees each, spaced randomly over plantation. Average values for 100 samples presented for each case.

No. of Rows/Sample (n)	Average Estimated Proportion ($\hat{p}$)	Average Bias ($p-\hat{p}$)	Average Variance $\hat{V}(\hat{p})$
5	0.031515	-0.000347	0.000288
10	0.031929	0.000067	0.000142
15	0.031798	-0.000064	0.000089
20	0.031884	0.000022	0.000070
25	0.031885	-0.000008	0.000054
30	0.030892	-0.000970	0.000046

Scheaffer *et al*. (1979) state that systematic sampling is equivalent to random sampling if the population being sampled is random. Results presented in Table 3 confirm this, with proportion estimates and their variances comparable to values presented in Table 1. Again, sample variance estimates stabilize in the n = 20 to 25 range and sample estimates appear reasonably accurate for sample sizes as low as n = 5.

As seen for random sampling, systematic sampling does not perform as well in the case were infestation is tightly clumped (Table 4). However, sample estimates are still within reasonable limits.

In summary, it appears the advantages of systematic sampling can be gained with no significant loss in accuracy of estimates.

(iii) Field Results

Table 5 presents results obtained from analysis of field data. 95% confidence limits for most of the estimates are relatively narrow. However, those for Campbell River - Sayward Blk. A are quite broad, mostly due to the small sample size (n=4). Again, because of the way the data was collected (a "100% sample" estimates of between transect variance are low resulting in confidence intervals that are too short.

In order to calculate required sample sizes given a required bound on the error of estimation (e.g., $\hat{p} \pm 0.10$ or 0.05) and confidence interval (e.g., 95%, 80%) the between transect variation (S_c^2), average number of trees in a transect (M) and total number of transects in the plantation (N) must be specified. Table 6 shows sample sizes required for various combinations of values for the above variables. The value for between transect variation was set a $S_c^2 = 200$, based on a conservative estimate from the field data. Formulas used for calculating sample sizes appear in Appendix II.

TABLE 3. Comparison of actual proportion of infested trees (p) to estimates ($\hat{p}$) based on systematically chosen rows. Infested trees randomly located on plantation. Average values for 10 samples presented for each case - rows 1 through 10 successively chosen as starting point.

Row Interval	No. of Rows/Sample (n)	Average Estimated Proportion ($\hat{p}$)	Average Bias ($p-\hat{p}$)	Average Variance $\hat{V}(\hat{p})$
		p = 0.00068		
30	5	0.0061	0.0014	0.000013
15	10	0.0065	-0.0003	0.000006
10	15	0.0068	0.0000	0.000004
8	19	0.0067	-0.0001	0.000004
6	25	0.0070	0.0004	0.000003
5	30	0.0069	0.0001	0.000002
		p = 0.0652		
30	5	0.0622	-0.0030	0.000099
15	10	0.0622	-0.0003	0.000046
10	15	0.0658	0.0000	0.000034
8	19	0.0661	0.0009	0.000028
6	25	0.0649	-0.0005	0.000021
5	30	0.0650	-0.0002	0.000017
		p = 0.4923		
30	5	0.5046	0.0123	0.000730
15	10	0.4974	0.0051	0.000279
10	15	0.4922	0.0036	0.000285
8	19	0.4937	0.0014	0.000201
6	25	0.4927	-0.0004	0.000135
5	30	0.4913	-0.0010	0.000133

TABLE 4. Comparison of actual proportion of infested trees (P=0.0319) to estimates ($\hat{p}$) based on systematically chosen rows. Infested trees located in 20 tight clusters of 25 trees each, spaced randomly over plantation. Average values for 10 samples presented for each case.

Row Interval	No. of Rows/Sample (n)	Average Estimated Proportion ($\hat{p}$)	Average Bias ($p-\hat{p}$)	Average Variance $\hat{V}(\hat{p})$
30	5	0.0283	-0.0036	0.000214
15	10	0.0364	0.0045	0.000166
10	15	0.0319	0.0000	0.000099
8	19	0.0335	0.0016	0.000079
6	25	0.0328	0.0011	0.000059
5	30	0.0324	0.0006	0.000044

TABLE 5. Estimated proportion ($\hat{p}$) of weeviled trees and 95% confidence limits for field data collected at various Sitka spruce plantations through out Coastal British Columbia

District	Location	No. of Transects	$\hat{p}$	95% Confidence Limits	
				Lower Limit	Upper Limit
Campbell River	Sayward Blk. A	4	0.27	0	0.56
Campbell River	Sayward Blk. B	12	0.39	0.35	0.41
Midcoast	Noosgulch	16	0.32	0.27	0.38
Midcoast	Sheemahat Blk. A	16	0.021	0.0095	0.033
Midcoast	Sheemahat Blk. B				
Kalum	Onion Lake	32	0.037	0.021	0.053
Kalum	Pontoon Creek	19	0.17	0.14	0.21

Table 6. Sample size calculations for estimating proportion of infested trees. Between transect variation (S_b^2=200) estimated from previous field data.

Bound on Error of Estimation	No. of Transects in Population	Average Transect Size(no. of trees)	No. of Transects Required (sample size)
		80% Confidence Limits	
0.10	100	100	5
0.10	150	100	5
0.10	200	100	5
0.05	100	100	18
0.05	150	100	20
0.05	200	100	20
		95% Confidence Limits	
0.10	100	100	7
0.10	150	100	8
0.10	200	100	8
0.05	100	100	24
0.05	150	100	26
0.05	200	100	28

SUMMARY AND RECOMMENDATIONS

Based on results obtained from computer simulation studies and operational considerations a systematic sample of rows (or groups of adjacent rows) throughout a plantation is the method of choice. The sampling method provided accurate estimates of proportion of infested trees for both randomly and tightly clustered spatial patterns of attack. Systematic sampling has the further advantages of uniform coverage of the plantation, easy implementation in the field and potential for mapping infestation pockets. At this time a conservative sampe size estimate of n=20 is recommended to gather pilot data.

Appendix I outlines preliminary sampling guidelines for collection of pilot data. Subsequent analysis will allow adjustment of sample sizes to reflect actual field conditions. Depending on between plantation variability a general guideline for sample size may be developed related to plantation size. Future study could also include an analysis of the cost relationships between the component parts of the sampling scheme with a view to increasing sampling efficiency as well.

LITERATURE CITED

Gross, H.L., A.R. Ek and R.F. Patton. 1980. Comparison of sample rules for estimating jack pine density, basal area, and percentage with sweetfern rust cankers. Can. J. For. Res. 10:190-198.

Karpinski, C. and J.A. Witter. 1982. Plot size and number of plots needed to assess spruce budworm impact. Can. J. For. Res. 12:332-336.

MacLean, D.A. and D.P. Ostaff. 1983. Sample-size precision relationships for use in estimating stand characteristics and spruce budworm caused tree mortality. Can. J. For. Res. 13:548-555.

Scheaffer, R.L., W. Mendenhall and L. Ott. 1979. Elementary survey sampling. 2nd ed. Duxbury Press, North Scituate, MA.

Talerico, R.L. and R.W. Wilson Jr. 1973. Sampling plantations to determine white-pine weevil injury. USDA Forest Serv. Res. Note NE-173. 4 p.

Waters, W.E. 1974. Sequential sampling applied to forest insect surveys. Cunia, T., ed. Monitoring forest environment through successive sampling: proceedings of the conference; June 1974; Syracuse, New York. State University of New York: pp. 290-311.

Yandle, D.O. and E.R. Roth. 1971. Ratio estimation of fusiform rust incidence in southern pine plantations. For. Sci. 17:135-142.

APPENDIX I

Preliminary Sampling Guidelines for Estimating the Proportion of Weeviled[1] Trees in a Plantation

1. Objective: To determine the proportion of weeviled leaders in a plantation within .10 at the 95% level of confidence.

2. Target Population:

The target population is all the living trees on a plantation; missing or dead trees are not considered.

3. Sampling Procedure:

3.1 Sampling Unit

The basic sampling unit is an individual tree. For convenience, however, they will be examined in groups - in this case the obvious choice is a row (or rows) of trees in the plantation, which here will be called a transect. Each tree observed will be tallied as attacked or not attacked. Usually, only current attack will be recorded.

3.2 Size of Transects

The number of plantation rows observed per transect will vary from one to four or more depending upon growth of brush, size of trees and perhaps terrain. Based on 3 m spacing it would appear feasible to walk between two rows and tally both healthy and weeviled trees on both sides of the observer. In young plantations with little brush problem, perhaps 4 or more rows could be tallied per transect with the observer walking down the central row(s). Naturally, the observer should attempt to tally as many trees (rows) as possible per transect, in any case the minimum number of trees observed per transect should be in the order of 100, whereas the maximum will be determined by row length. While not crucial to the sampling scheme, an attempt should be made to keep the number of trees observed per transect roughly equal.

[1] e.g., *Pissodes strobi*, *P. terminalis*.

3.3 Orientation of Transects

Transects will be run parallel with plantation rows.

3.4 Number of Transects

The number of transects required for any particular plantation is dependent on an number of factors, namely:

(i) the required bound on the error of estimate (i.e. ± .10, ± .05)

(ii) the required level of confidence in the estimate (i.e.,95%, 80%, etc.)

(iii) the total number of transects (whether defined as 1 to 4 or more plantation rows) in the entire plantation

(iv) the average number of trees in each transect

(v) the transect to transect variation in numbers of weeviled trees

Of these only (i), (ii) and probably (iii) can be defined before sampling begins.

Determination of (iv) and (v) depends on previous (pilot) studies or analysis of preliminary field data (using values from the first 10 transects taken in a plantation for example).

Based on the above, a general guideline of 20 transects per plantation is proposed. This will provide preliminary information that will allow fine-tuning of the sampling scheme.

3.5 Placement of Transects

Transects should be placed systematically (i.e., spaced at equal distances) throughout the plantation. Divide the total number of rows in the plantation by the number of transects required and place transects at fixed distances apart using this "transect interval" with a random start.

For example, suppose we wish to tally 20 transects in a plantation of 150 rows, the transect interval would be 150/20 = 8 (round up). Now for our random start choose a random number between 1 and 8, say 5, and our transects would then occur at (or center around) rows 5, 13, 21, 29, 37, ... etc. Thus, the transects will be spread uniformly over the entire plantation.

3.6 Tally Sheet

Attached is a suggested format for data collection. It includes sections for raw data entry and observers notes.

In order to develop sample size estimates, it is necessary to know the total number of transects (N) in the plantation from which the systematic sample is taken. Thus, if a plantation has 500 rows of trees and a transect consists of 2 rows then N = 500/2 = 250 transects, of which n = 20 will be systematically sampled as recommended in this guideline. This value for N need not be exact but could be estimated from an air photo or plantation layout map.

Note that these are proposed guidelines; please contact Val Fletcher, Research Branch (ph. 387-3025) if you have any immediate questions, comments or concerns.

Tally Sheet

Spruce Leader Weevil Survey

Plantation I.D. (Admin. No.)__________________ Date_______________

Region _______________ District _____________ Location ___________

Transect No.	No. of Trees Observed	No. of Trees Attacked	Remarks

Notes:____

Total no. of transects in plantation (N) = ____ (see paragraph 3.6, pg. 17)

Sketchmap:

APPENDIX II

(i) Estimated population proportion p:

$$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

(ii) Estimated variance of $\hat{p}$:

$$\hat{V}ar\ (\hat{p}) = \left(\frac{N-n}{Nm\bar{M}^2}\right) \frac{\sum_{i=1}^{n}(a_i - \hat{p}m_i)^2}{(n-1)}$$

where: N = total number of rows in the plantation

n = number of rows selected at random without replacement

m_i = number of trees observed in the i^{th} row

a_i = number of weevil infested trees in the i^{th} row

$\bar{M}$ = the average number of trees per row for all rows

($\bar{M} \approx \bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$ = the average number of trees per row for the sample)

(iii) Confidence intervals for p:

95% confidence interval = $\hat{p} \pm 1.96\sqrt{\hat{V}ar(\hat{p})}$

80% confidence interval = $\hat{p} \pm 1.68\sqrt{\hat{V}ar(\hat{p})}$

(iv) Required sample size for a given bound (B) on p at 95% confidence:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$$

where: $D = \frac{B^2\bar{M}^2}{4}$ and

σ_c^2 estimated by: $\frac{\sum_{i=1}^{n}(a_i - p\,m_i)^2}{n-1}$

$$\bar{M} \approx \bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$$

All formulas are after Scheaffer _et_ _al_. (1979).