A Ground Survey Method For Estimating Loss Caused By *Phellinus weirii* Root Rot

I. Development of Survey Design

W. J. Bloomberg, P. M. Cumberbirch and G. W. Wallis
Canadian Forestry Service / Pacific Forest Research Centre
Victoria, B.C., BC-R-3, March, 1980

SUMMARY

A transect line sampling system comprising sets (grids) of lines was developed to estimate the incidence, distribution and area of root rot caused by *Phellinus weirii* (Murr.) Gilbertson in Douglas-fir stands. Estimates of total area of root rot were derived from intersection length and probability of occurrence; the latter was also used to estimate area and number of infection centers by size class. In simulated stands, two methods of measuring infection centers, radial and linear, gave accuracies of 96-109 and 94-114% of true area, respectively. Double sampling with regression was used to estimate areas of centers measured by an approximate (rectangular) method from those measured by linear or radial methods. Estimates of center area by regression ranged from 97-101% of true area. In simulated stands, estimates of total root rot area by intersection length and probability of occurrence methods ranged from 94-112 and 95-105% of true area, respectively. Systematic location of transect lines within grids resulted in smaller variation than randomly located lines.

Field tests of the methods in 17 stands, 3 to 119 ha in area, gave estimates of 0.8 to 41.4% of stand area ir infection centers. Average estimate of root rot area by intersection length method was 110% of that estimated by probability of selection and variance of estimate was greater. Estimates of root rot area remained relatively stable as the number of grids used was increased from 3 to 5. Regressions of accurately measured infection center area on approximately measured area were highly significant in all stands.

RÉSUMÉ

Un système d'échantillonnage par transect linéaire a été mis au point pour évaluer l'incidence, la répartition et l'aire de *Phellinus weirii* dans les peuplements de Douglas taxifolié. Les estimations de l'aire totale du Pourridié ont été dérivées à partir de la longueur d'intersection et de la probabilité de sélection; on a utilisé cette dernière aussi pour évaluer l'aire et le nombre de centres par catégories de grandeur. Un double échantillonnage avec régression a été utilisé pour évaluer la superficie des centres mesurés approximativement à partir de ceux mesurés avec précision. Dans des peuplements simulés, la localisation systématique de lignes de transect à l'intérieur des grilles a produit de plus petites erreurs d'échantillonnage que les lignes aléatoires.

Des essais de la méthode sur le terrain dans 17 peuplements d'une superficie de 3 à 119 ha ont fourni des estimations de 0,8 à 41,4% de la surface des peuplements dans les centres où sévit le Pourridié. Dans deux peuplements où l'aire du Pourridié avait été mesurée par d'autres méthodes, les estimations concordaient. L'estimation moyenne de l'aire du Pourridié par longueur d'intersection a été de 110% de celle évaluée par probabilité de sélection. La régression des centres mesurés approximativement sur les centres mesurés avec précision était grandement significative dans tous les peuplements (b = ,710 - ,829). Cinq organismes et compagnies utilisant cette méthode l'ont favorisée.

INTRODUCTION

Root rot caused by *Phellinus weirii* (Murr.) Gilbertson is responsible for large volume losses in second-growth stands and imposes severe limitations on the number of intensive management options available to managers for infected stands. Although there is a need to obtain a better estimate of the overall losses to this disease, the first priority is to develop a method which provides estimates of this root rot in stands designated for intensive silvicultural treatment. To choose optimal stand treatments, managers must have a knowledge of the extent and distribution of *P. weirii*. Infection of young trees by *P. weirii* occurs initially when roots contact the fungus surviving in the roots and stumps of the previous stand. Subsequent tree-to-tree spread occurs where diseased and healthy roots are in contact. This mode of spread results in aggregation of infected trees in centers. The distribution of infection centers in a stand may be random or clumped, partly depending on the distribution of infected stumps in the previous stand.

Infection centers are most often recognizable by the presence of stand openings containing standing dead trees or wind-thrown trees with characteristic root rot symptoms, *i.e.,* root-balls and laminated pitted rot, or living trees with crown symptoms, *i.e.,* shortening of terminal shoot growth, chlorosis and distress cone crops (Wallis 1976). Trees in the earlier stages of infection may show no above-ground symptoms, thus the disease almost invariably extends beyond the boundaries of a visible infection center.

Detection and delineation of *P. weirii* centers using various remote sensing media has been only partially successful. Weber and Wear (1970) found little thermal radiance differences among the various classes of root rot-infected trees they examined. Williams and Leaphart (1978) identified well-established *P. weirii* centers in Douglas-fir stands in northern Idaho to a 92% accuracy, using Ektachrome film at a scale of 1:4000. Johnson and Wear (1975) reliably rated large openings in stands in the Cascades of central Oregon; however, in coastal stands where ground cover obliterated many details of infection centers, errors in interpretation ran to 50%. Reliable estimates of the occurrence and distribution of small openings in young stands using remote sensing techniques appear to be somewhere in the future. Aerial sketch mapping of root rot is unsuited to quantitative estimates because of these difficulties in recognition and mapping of root rot centers (Harris,

pers. comm.).

Various ground sampling systems have been investigated. Sequential sampling of root rot (caused primarily by *Armillaria mellea* (Fr.) Kummer) in 15-to-17-year-old Douglas-fir plantations provided only a broad classification of incidence (Foster and Johnson 1963). Waters (1978) concluded that without information on the mathematical distribution of infected trees, it was impossible to construct decision boundaries for sequential sampling. Random samples using 20-tree quadrats (Foster and Johnson 1963) gave large coefficients of variation (C.V.), 52-230%, and required sampling intensities of 4.3 to 46.8% to give 10% precision of estimate. Stratified random samples reduced this variation only slightly. Waters (1978) found that prism plot sampling in *P. weirii*-infected Douglas-fir stands gave large sampling errors. Sampling by belt transect, 33 feet wide, was more accurate than prism plots in estimating the proportion of stand area affected by root rot and provided estimates of number and average size of infection centers. However, biased estimates resulted when centers partially intersected by the transect were included. Including only those centers wholly within the transect greatly decreased the number sampled and resulted in a large variation. Waters inferred that estimates of center size would be greatly improved if the survey methodology included a factor for "probability of encountering" various sized centers.

The highly aggregated distribution of root rot in stands (Foster and Johnson 1963) is the major factor responsible for the large variation in sampling systems which are based on number of diseased trees in plots. A further contributing factor is the deterioration and disappearance of root rot-killed trees over time.

This report describes the derivation and testing of a design for a ground-survey for estimating incidence, distribution and area of *P. weirii* infection centers, taking account of the complexities described above. Application of the method is described in a separate report (Bloomberg et al. 1980).

DEVELOPMENT OF SURVEY METHODS

Choice of Survey Method

A ground-survey system for *Phellinus weirii* root rot, acceptable to operational managers and providing the necessary information for treatment decisions in infected stands, should meet the fol-

lowing criteria.

a) The system should be flexible in allowing some choice of accuracy relative to the time and funds available for surveying stands.

b) The system should be applicable over the range in stand conditions in which the disease is found.

c) Following a short training period, field procedures should be understandable by field crews with a knowledge of volume cruising.

d) The system should be compatible with stand volume cruise procedures.

e) Data recording, processing and analysis should be designed to minimize clerical input and maximize computer application.

Because infection centers continue to enlarge over many years, during which many of the early-killed trees deteriorate and are lost to the stand, accurate estimates of disease incidence cannot be based on number or size of infected trees. The occurrence of the disease in discrete infection centers suggested that the frequency and total area of these centers relative to the area of the stand would be appropriate estimators of disease incidence and could be converted to number of stems or volume using stand tables. Sampling methods based on transect lines have been successfully applied to estimating area of vegetation types independently of their distribution (Canfield 1941; Hasel 1941) and are compatible with methods of volume cruising in forest stands.
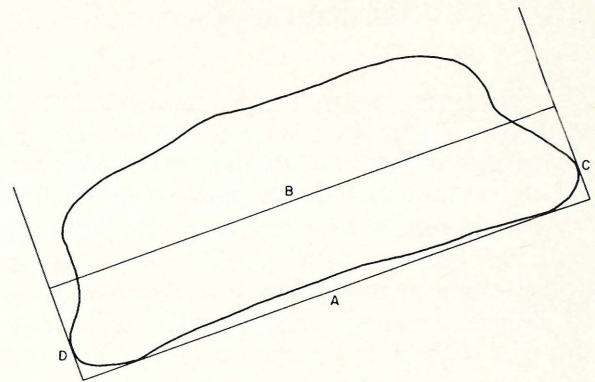

Fig. 1. Location of baseline may be outside (A) or inside stand (B), but must extend full length of stand; *i.e.*, between lines at right angles to baseline and passing through opposite extreme edges of the stand (C,D).

Intersection Length Method

Theoretical justification for the following method of estimating root rot area is given in Appendix 1 (Cumberbirch 1976). The total area of root rot centers within a stand can be estimated by running a straight transect line through the entire stand, summing the lengths of the segments of the line which
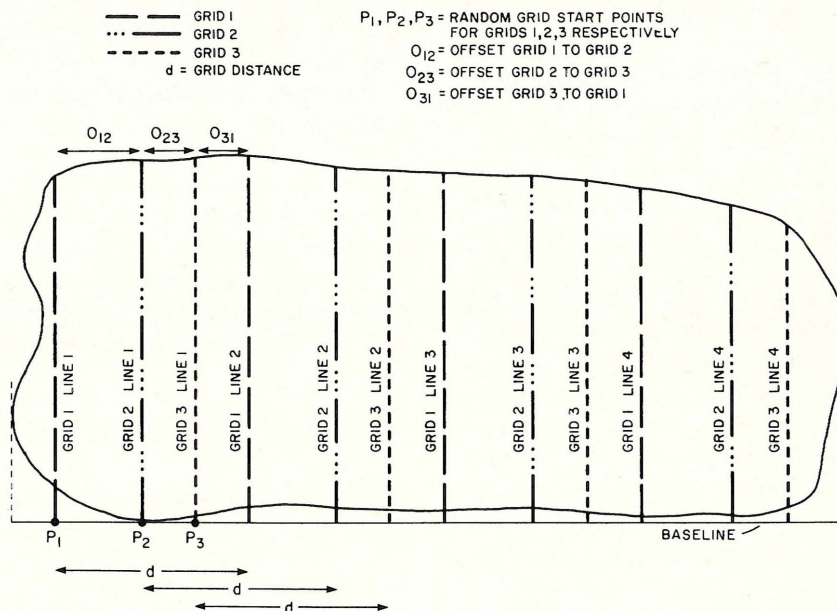

Fig. 2. Sampling plan consisting of three grids each containing four transect lines.

fall within centers (intersection lengths) and multiplying the total by a constant. This estimate is unbiased if the transect line passes through a randomly selected point on a baseline located so that the entire stand is exactly contained between perpendicular lines drawn at each end of the baseline (Fig. 1). The constant in this case is the length of the baseline. Increase in accuracy of an estimate of total root rot area is obtained by increasing the number of transect lines and averaging the estimates. A series of equally spaced lines perpendicular to the baseline is termed a grid (Fig. 2). The fixed distance between lines is called the grid interval.

Unbiased estimates of the total root rot area from a number of grids can be averaged to obtain a combined estimate. Random location of grid lines in a stand results in independent estimates given by each grid, hence the variance of their means (sample variance) can be estimated (Cochran 1963). Ideally, a baseline should be located to minimize differences in the lengths of transect lines, but this choice will be constrained by practical considerations.

Sampling variation will be small when root rot centers are relatively uniform in size and in distribution within a stand and will be increased by disparate size or aggregated distribution of centers because, in the former case, total intersection lengths of each transect line will tend to equality and, in the latter case, tend to be unequal (Appendix 1). Choice of stand boundaries to reduce sampling variation is desirable, but limited in practice. Experience indicates that forest cover types, or stand types used in the B.C. Ministry of Forests cover maps, or their equivalent, are generally satisfactory for root rot sampling purposes. Combination of markedly different cover types in the same sample probably increases sampling variation and is not recommended.

Probability of Occurrence Method

Accuracy of transect line estimation of root rot area can be improved by including probabilities of root rot centers being encountered along a line (Waters 1978). Probability of encounter is determined by the ratio of projection length of a center to the grid interval (Appendix 2). Projection length is defined as the distance parallel to the baseline between lines passing through the extremes of a center boundary at right angles to the baseline (Appendix 2, Fig. 1). Probability of occurrence of centers in the entire stand is estimated from their probability of
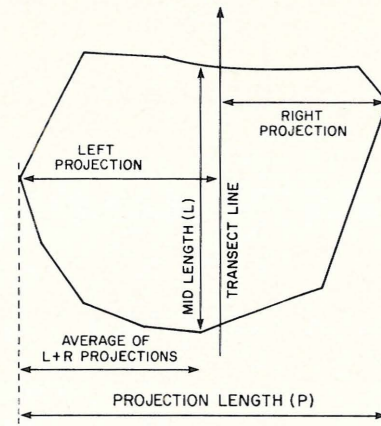

Fig. 3. Rectangular measurement of infection center area is product of mid-length (L) and projection length (P); (P = sum of left and right projections).

encounter in the sample. Estimates of centers with specific characteristics; e.g., size class, are given by the sum of the areas of such centers multiplied by their probability of occurrence (Appendix 2). Probability of encounter can be increased by adding width to transect lines.

Rapid, approximate estimates areas of centers can be derived by measuring two axes at right angles to each other and multiplying them (rectangular method, Fig. 3) More accurate estimates can be made from measuring several parallel intersection lengths at fixed intervals within the center (linear method, Fig. 4), or by measuring several radii of a center subtended at a fixed angle (radial method, Fig. 5). Mathematical derivation of the linear and radial methods are given in Appendices 2 and 3, respectively. Double sampling with regression can be used to improve the accuracy of a rectangular estimate of area of a center by using regression of linear- or radial-measured on rectangularly measured area (regression method).

The regression relationship of rectangularly measured and true area was tested in 60 simulated infection centers randomly located in a stand. Their true area was calculated from center coordinates. The rectangular area of each center was calculated as the product of the projection length and i) one intersection length passing at right angles through the midpoint of the projection length, or ii) the average of two intersection lengths passing through points approximately one-third and two-thirds the
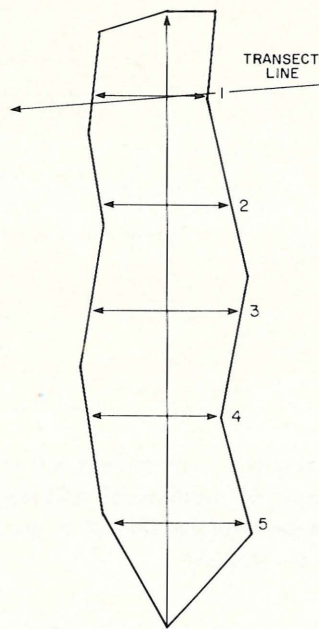
Fig. 4. Linear measurement of root rot centers is based on a line running the full length of the center and a number of intersection lengths at right angles to and at equal intervals on the line.
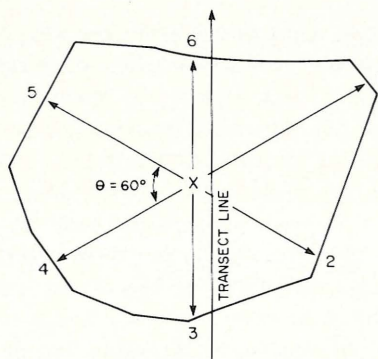

Fig. 5. Radial measurement of root rot centers is based on a number of radial measurements from the approximate mid-point of the center to the margin at equal angular intervals (θ).

distance along the projection length. The regression equations were $A_1 = .731X_1 + 2.94$ and $A_2 = .491X_2 + .59$, where A_1 and A_2 represent estimated true area (m^2) based on one or two intersection lengths, respectively, and X_1 and X_2 are the corresponding

regression coefficients. The regression equations were highly significant ($p = .01$). They were then applied to 30 simulated infection centers having different sizes, shapes and distributions from those used to develop the regression equations. The results (Table 1) show that both equations gave accurate estimates of area and measurement, with two intersection lengths being slightly more accurate.

The accuracy of linear or radial measurement of infection centers was tested in 13 to 15 simulated centers of various sizes. The linear method was used on centers which were generally elongated and the radial method on centers which approximated rectangular or circular shapes. The results (Table 2) show a slight trend for accuracy of the linear method to increase, and variation to decrease, with increasing number of intersection lengths measured. Accuracy of the radial method was not affected practically with increased number of radii measured (Table 3).

Amount of variation incurred in measuring real root rot centers was investigated by measuring centers by the radial, linear, or both methods, in a 30-year-old Douglas-fir stand. Two estimates were made for each method, each using 6 radii or intersection lengths and randomly selected starting points. Results show coefficient of variation (C.V.) was practically insignificant in small centers, but increased in larger centers (Table 4). Differences between two measurements of one center were attributed to inaccuracy in measurement and irregularity of center boundaries. Coefficents of variation for center measurement were of the same general magnitude of those estimated from simulated centers.

TESTING THE SAMPLING METHOD IN SIMULATED STANDS

Total Root Rot Area Estimates by Intersection Length

Accuracy of total root rot area estimates by the intersection length method were tested on approximately 50 simulated 1 ha stands generated by computer, each containing up to 25 infection centers, 1 to 500 sq m area. Distribution of centers in the simulated stands was varied from random to highly aggregated by allocating centers to one or more subdivisions of the stand. Shape of centers was varied from almost circular to almost linear

Table 1

Accuracy of estimation of simulated infection center area using regression of true area on rectangular measurement area based on 30 centers.

Simulated stand no.	True area (sq units)	No. of intersection lengths per center	
		1	2
		% of true area estimated by regression equation	
1	10.65	97.9	99.5
2	17.34	97.9	101.4
3	22.70	98.1	96.5
Avg.		98.0	99.1

Table 2

Effect on accuracy of estimates of simulated infection center area by linear method[a] based on different numbers of intersection lengths.

No. of intersection lengths	Center size class (m^2)				C.V. %
	10-25 (3)[b]	26-50 (4)[b]	50-100 (3)[b]	100+ (3)[b]	
	% of true area				
2	102	114	106	120	10.9
3	104	109	107	110	5.4
4	101	106	94	107	5.0

[a] See text for explanation.
[b] Number of infection centers.

Table 3

Effect on accuracy of estimates of simulated infection center area by radial method[a] based on different number of radial measurements.

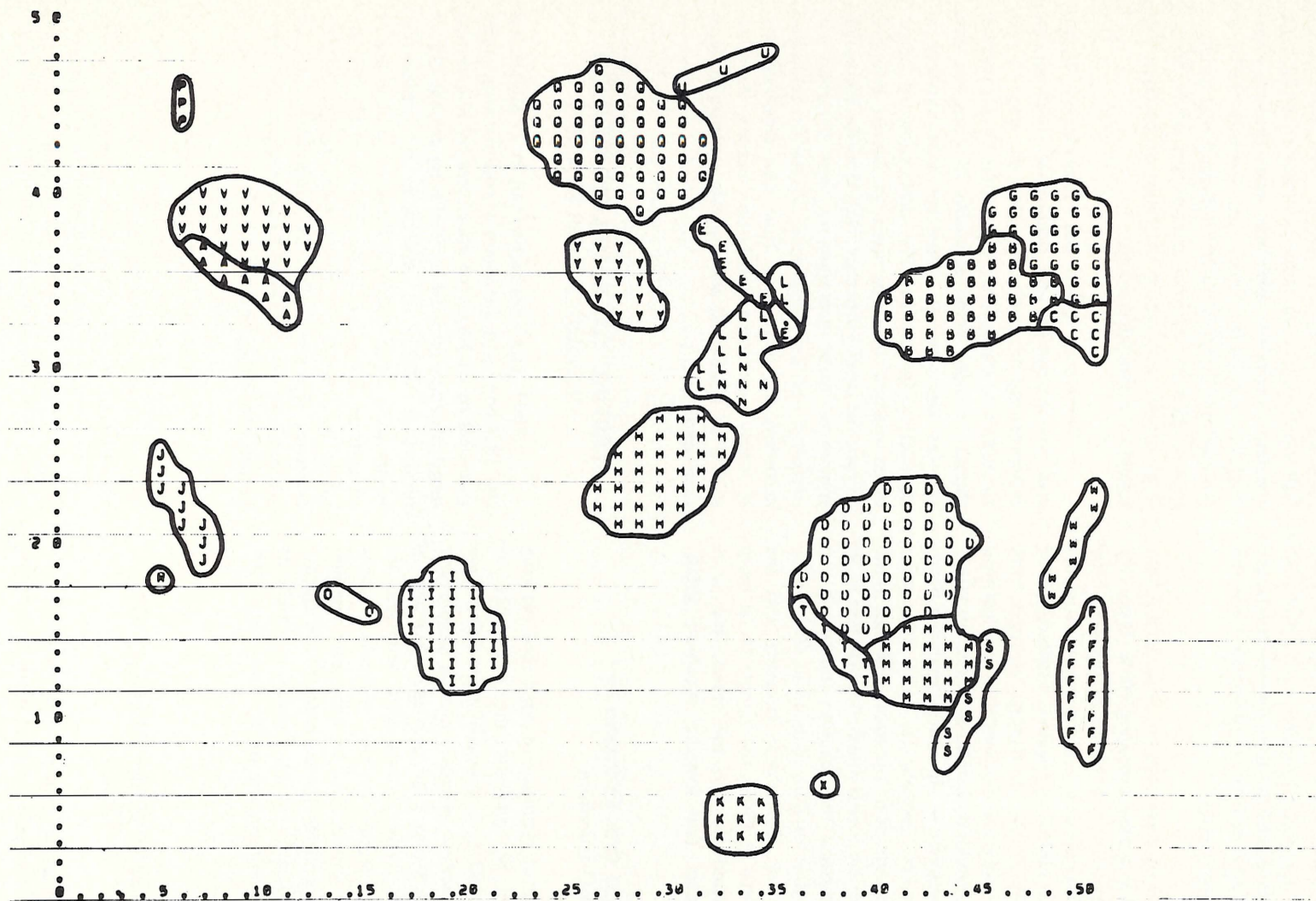| No. of radial lengths measured	Center size class (m^2)				C.V. %
	12-25 (5)[b]	26-75 (5)[b]	76-150 (3)[b]	151+ (2)[b]	
	% of true area				
4	97	101	109	102	16.0
6	96	98	100	102	8.1
8	99	98	104	101	15.1

[a] See text for explanation.

[b] Number of infection centers.

Table 4

Estimation of sampling error in linear and radial measurements of infection centers in a 30-year-old Douglas-fir stand.

Center No.	Measurement method	Measurement no.	Estimated area (m^2)
1	Radial	1	325
		2	334
		C.V.(%)	1.9
2	Linear	1	4535
		2	3838
		C.V.(%)	11.9
3	Radial	1	1403
		2	1143
		C.V.(%)	14.5
	Linear	1	1364
		2	1074
		C.V.(%)	16.9

Fig. 6. Simulated root rot centers used to test various sampling designs. Each center is represented by a letter and may merge with other centers. Shape, size and orientation of centers was varied for each test.

(Fig. 6) by using different ratios of major to minor area, and orientation of the center axes was varied with respect to the baseline. Twenty parallel transect lines per stand, randomly or systematically spaced, were simulated for each computer run. Total root rot area was calculated from total intersection length and compared with true area calculated from the coordinates of each center.

Estimates of total root rot area ranged from 53 to 153% of true area, averaging 101%. Sixty-two percent of estimates were between 90 to 110% of the actual area. Divergence of estimates from true area was greatest in highly aggregated distributions of infection centers (Table 5). Average estimate from systematically positioned transect lines was identical with that from randomly positioned lines. As expected, C.V. was greatest in the most aggregated distribution of infection centers; *i.e.,* with all centers located in one quadrant of the stand, and least in the least aggregated; *i.e.,* equal numbers of centers in all quadrants. Systematic line location resulted in less sampling error than random location in all distributions of root rot. Systematic sampling also has advantages in that it forces the sample to be more equally distributed through the stand, and it is simpler to design and execute (Cochran 1963).

Total Root Rot Area Estimates by Probability of Occurrence

Accuracy of estimates of total root rot area was determined by applying the probability of occurrence method to a simulated stand using three grids with randomly selected starting points and transect lines 25 m apart. Areas of individual centers encountered on transect lines were estimated by one or two intersection lengths. From the results (Table 6), the probability of encounter method was considered accurate for purposes of the survey. Accuracy of estimate was not reduced for practical purposes by measuring one rather than two intersection lengths.

CONCLUSIONS FROM SIMULATION TESTS

The intersection length and probability of occurrence methods gave accurate average estimates of total root rot area. Estimates of root rot area by intersection length were more variable in accuracy than those using probability of occurrence, and most variable in highly aggregated distributions. Estimates of individual infection center size by double sampling with regression were considered accurate for purposes of the survey. Average area estimates of individual centers by radial or linear methods were accurate, provided sufficient measurements were made in relation to center size. Limited testing of the methods in real situations indicated that variation in area estimates of small centers was small, and in large centers was within the range of precision obtained from simulated results.

Therefore, both the intersection length and the probability of occurrence methods are recommended for estimates of total area of root rot within a stand. The intersection length method should be used when estimates of root rot area by size class are less important than rapidity in conducting the survey. The probability of occurrence method must be used to obtain estimates of root rot area by size class. Distributions aggregated enough to cause unacceptably large variation will probably not be encountered in stands delineated according to forest cover type map classifications. If highly aggregated distribution occurs, the stand should be subdivided and separate samples made in each subdivision.

TESTING THE SAMPLING METHOD IN DOUGLAS-FIR STANDS

Field tests were carried out in 17 stands, 3.2 to 118.8 ha in size. Estimates of root rot area ranged from 0.8 to 41.4% of the stand area by intersection length method and 0.4 to 21.1% by probability of occurrence (Table 7). Estimates by intersection length were 77 to 228% of those by probability of occurrence, averaging 110%. Eleven of the 17 estimates by intersection length were within 25% of estimates by probability of occurrence, and all estimates by the former included those of the latter in their confidence limits.

Coefficient of variation of estimates of root rot area by the intersection length method ranged from 5 to 55%, averaging 23.9% (Table 8). There appeared to be no relationship between variation and per cent root rot. Stands with the highest variation had highly aggregated distribution of infection centers. Confidence limits at the 95% level ranged from ±26 to ±129% and ±10 to ±130% of the estimated area by intersection length and probability of occurrence methods, respectively, averaging

Table 5

Accuracy of root rot area estimates by intersection length method based on 50 simulated forest stands, with 20 transect lines per stand.

		Distribution of root rot centers in stand[a]				
Type of sample	Estimated statistic	x | x ‾‾ x | x	‾‾ x | x	x | ‾‾ x |	‾‾ x |	Avg
Systematic location of transect lines	Area[b]	99	100	112	94	101
	C.V.[c]	11	11	12	21	14
Random location of transect lines	Area[b]	98	103	96	105	101
	C.V.[c]	13	21	30	33	24

[a] x indicates simulated infection centers were located in quadrant of stand; quadrants without an x were void of root rot.

[b] As average per cent of actual area.

[c] Based on area estimates.

Table 6

Accuracy of total root rot area estimates by probability of occurrence method using two methods of individual center measurement and three grids in a simulated stand.

Center measurement method[a]	Grid no.			Avg	C.V. %
	1	2	3		
	% of true area				
1	102	99	105	102	6.5
2	102	95	95	97	5.4

[a] Methods 1 and 2 used 1 and 2 intersection lengths per center, respectively.

Table 7

Estimated root rot area and number of centers using intersection length and probability of occurrence.

Stand area (ha)	Estimated root rot area by		IL as % of PO	Est. no. centers
	Intersection length (IL) (% of Stand area)	Prob. of occurrence (PO)		
3.2	22.6	16.2	139	15
3.6	19.4	17.6	111	3
7.8	0.8	0.4	200	5
7.8	1.3	0.8	162	3
7.8	41.4	18.1	228	1
9.0	10.4	7.9	115	72
10.4	0.6	0.6	100	5
11.7	5.4	4.5	120	1
12.0	12.9	10.7	121	45
14.6	14.4	13.8	104	12
17.8	1.0	1.3	77	12
38.8	3.4	1.9	179	19
40.5	8.3	6.3	138	67
50.8	18.0	21.1	85	439
53.0	11.9	10.5	113	241
89.1	4.9	5.0	98	38
118.8	19.2	16.8	114	635
Avg.			110	

±54.8 and ±41.3%. Limits tended to be large in sampling plans using only two (the minimum number) grids, since the appropriate Student's-t value was based on only one degree of freedom. Plans with three or more grids resulted in smaller confidence limits than those with two grids.

Comparisons of the effects of using different numbers of grids on estimates of total area and variation of root rot were made by recalculating the estimates based on two, three or four arbitrarily selected grids from surveys which had been conducted using five grids with four transect lines in two stands, 67 and 140 ha in size. The results (Table 9) showed that estimates of total area using intersection length and probability of occurrence were relatively stable in both surveys using 4-5 and 3-5 grids, respectively. Estimates of numbers of centers were less stable. Except for estimates of number of centers using two grids in the Manson Creek survey, confidence limits decreased with increasing numbers of grids.

In all surveys, regression of center area measured by radial and linear measurement on rectangularly measured area were highly significant ($p = .01$). Except in one stand, the regression line intercepts were not significantly different from zero; *i.e.*, it can be assumed that the regression lines passed through the origin. Regression coefficients fell within a narrow range (.710 to .829).

Problems were encountered in measuring centers in stands with heavy brush layers and difficulty in recognizing center boundaries was reported in stands where individual centers had coalesced into large meandering infection areas. Measurement by intersection length method only is recommended for such stands. For field tests, most centers were measured by rectangular and radial or linear methods. However, substantial saving in time can be achieved by selectively measuring centers to obtain a range of center sizes on which to base the regression equation.

Table 8

Sampling error of estimates of root rot area by intersection length method.

Estimated root rot area (% of Stand area)	Confidence limit (95% level)	C.V.[a] %
	% of estimated root rot area	
0.6	33	5
1.0	100	46
3.4	129	55
4.9	90	42
5.4	26	4
8.3	65	28
10.4	47	21
11.9	48	22
12.9	49	12
14.4	94	47
18.0	33	15
19.2	24	11
19.4	27	5
22.6	40	28
41.4	76	18
Avg.	54.8	23.9

[a] Based on area estimates.

Some differences occurred among the field crews in diagnosis of infected trees, with resultant variation in definition of root rot center boundaries. Only above-ground symptoms should be used to classify healthy and infected trees (Wallis 1976). Examination of the root collar or roots for presence of decay or mycelium of *Phellinus weirii* not only greatly slows down survey progress but, for reasons explained in "Introduction", introduces much more variation to boundary definition than crown examination. The latter, therefore, gives more relatively consistent, though conservative, results.

FURTHER DEVELOPMENT OF SURVEY DESIGN

Investigations will be made into correlation of infection center area measured from crown symptoms with actual area of infection determined from examination of root systems. Correlation factors, if significant, will be included in the calculations to give estimates of true area of root rot. Also, the possibility of incorporating root rot survey procedures into volume cruising procedures will be examined. The possibility of using large-scale aerial photos to broadly classify stands by root rot incidence is being investigated, with the objective of deciding the priority of stands for ground survey.

ACKNOWLEDGMENTS

Table 9

Comparison of total root rot area estimates by intersection length and probability of occurrence using different numbers of grids of 4 lines in two stands.

No. grids	Intersection length			Probability of occurrence						Area estimate by intersection length as % of estimate by probability of occurrence
	Area (ha)	Confidence limit (ha)	$(\pm)^a$ (%)	Area (ha)	Confidence limit (ha)	$(\pm)^a$ (%)	No. centers	Confidence limit (no.)	$(\pm)^a$ (%)	
Boulder Creek (67 ha)										
2	8.18	5.88	71.9	4.93	3.63	73.6	105	53	50.7	+65.9
3	8.55	4.31	50.4	5.55	2.99	54.0	105	45	43.0	+54.3
4	6.95	3.49	50.2	4.81	2.42	50.3	100	34	44.0	+34.0
5	7.25	3.09	42.6	5.12	2.24	43.8	114	39	34.4	+41.6
Manson Creek (140 ha)										
2	16.3	19.6	120.2	13.1	12.6	96.2	108	17	15.7	+24.4
3	13.0	12.7	97.7	9.7	8.4	86.6	123	57	46.3	+34.0
4	13.8	9.9	71.7	10.1	6.6	65.3	153	80	52.3	+36.6
5	14.0	8.5	60.7	10.0	5.7	57.0	140	68	48.6	+40.0

[a] 95% confidence level.

U.S. Forest Service, Forest Insect and Disease Management, Region 6, Portland, Ore.; W.G. Thies, U.S. Forest Service, Forest Sciences Laboratory, Corvallis, Ore.; E.M. Hansen, Department of Botany and Plant Pathology, Oregon State University, Corvallis, Ore.; A.J. Waters, MacMillan Bloedel Ltd., Nanaimo, B.C.

LITERATURE CITED

Bloomberg, W.J., P.M. Cumberbirch and G.W. Wallis. 1980. A ground-survey method for estimating loss caused by Phellinus weirii root rot. II. Survey procedures and data analysis. Can. For. Serv. Rept. BC-R-4.

Canfield, R.H. 1941. Application of the line interception method in sampling range vegetation. J. For. 39: 388-394.

Cochran, W.G. 1963. Sampling techniques. John Wiley and Sons, New York. 411p.

Cumberbirch, P.M. 1976. Development of specifications for disease survey and analysis of root rot infected stands. I. Statistical considerations. Mimeo report prepared under Contract No. 08876-02165 Dept. of Environment, Pacific Forest Research Centre, Victoria, B.C.

Foster, R.E. and A.L.S. Johnson. 1963. Studies in forest pathology XXV. Assessment of pattern, frequency distribution, and sampling of forest disease in Douglas-fir plantations. Canada Dept. For. Publ. 1011. 52p.

Hasel, A.A. 1941. Estimation of vegetation-type areas by linear measurement. J. For. 39: 34-40.

Johnson, D.W. and J.W. Wear. 1975. Detection of Poria weirii root rot centers in the Pacific Northwest with aerial photography. Pl. Dis. Reptr. 59: 77-81.

Wallis, G.W. 1976. Phellinus (Poria) weirii root rot. Detection and management proposals in Douglas-fir stands. Can. For. Serv. For. Tech. Rept. 12. 16p.

Waters, A.J. 1978. A problem analysis of root rot sampling. B.S.F. thesis, Univ. Brit. Columbia. 82p.

Weber, F.P. and J.F. Wear. 1970. The development of spectro-signature indicators of root disease impacts on forest stands. NASA Ann. Prog. Rept.

Williams, R.E. and C.D. Leapheart. 1978. A system using aerial photography to estimate area of root disease centers in forests. Can. J. For. Res. 8: 214-219.

APPENDIX I

1. Estimating Area - a line intersection method.

A. Theory

Consider the problem of estimating the area of the region R. (Fig. 1) Note that R can be defined as the area contained between two curves $f_1(x)$ and $f_2(x)$. i.e. can we write

$$A = \int f_1(x)dx - \int f_2(x)dx, \qquad (1)$$

where A is the area of R and the limits of integration are from o to a. That is, the area of R is just the area under the curve $f_1(x)$ minus the area under the curve $f_2(x)$. Defining $h(x) = f_1(x) - f_2(x)$, expression (1) above becomes

$$A = \int h(x)dx.$$
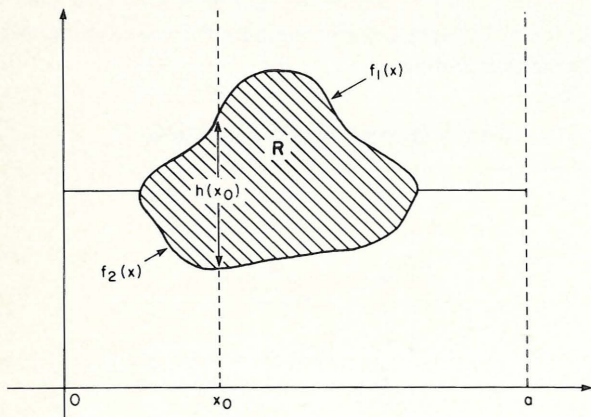
Fig. 1

To estimate A, let x be a random variable such that x ~ Uniform (0,a). Then x has density

$$g(x) = \begin{cases} 1/a \text{ for } 0 < x < a \\ 0 \text{ otherwise.} \end{cases}$$

An unbiased estimate of A is then given by

$$\hat{A} = ah(x)$$

To see that $\hat{A}$ is unbiased, note that

$$E(\hat{A}) = \int ah(x)g(x)dx$$

$$= \int h(x)dx$$

$$= A.$$

In practice one would obtain a sample of observations from the density g(x) and then use the sample mean $a\bar{h}(x)$. That is, one would use

$$\hat{A} = a \Sigma h(x_i)/n,$$

where $x_1, x_2, \ldots, x_n$ is the sample from g(x).
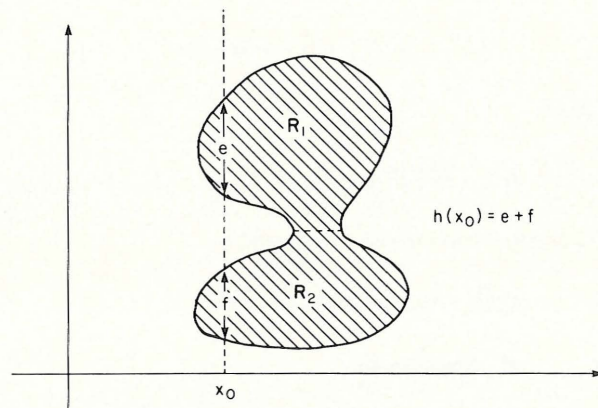
Fig. 2

For more irregular shaped regions it will generally be possible to divide these regions into subregions each of which can be represented as the area between two continuous functions (Fig. 2). In this case, if $h(x_0)$ is the sum of the lengths of intersections of the line $x = x_0$ with the subregions, the estimate $\hat{A}$ remains valid.

Note that in the above there was no reason for the region R to be connected. That is, the same estimate could also be applied to a number of disjoint regions as illustrated in Fig. 3. In general we have the following.

If $X_1,\ldots,X_n$ are independently distributed Uniform (0,a) random variables and $h(X_0)$ is the length of intersection of the line $X=X_0$ with the region R, then the statistic

$$\hat{A} = a \Sigma h(X_i)/n$$

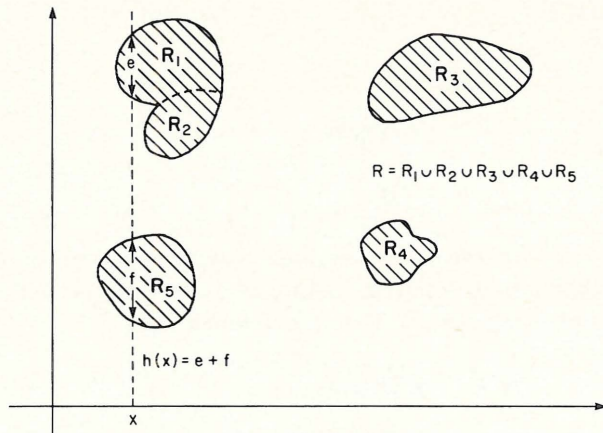is an unbiased estimate of the area of the region R.

Fig. 3

An unbiased estimate of the variance of $\hat{A}$ is

$$\hat{V} = a^2 \, \Sigma \, (h(X_i) - \bar{h})^2/n(n-1),$$

where

$$\bar{h} = \Sigma \, h(X_i)/n.$$

The actual variance of $\hat{A}$ is given by

$$E(\hat{V}) = (a \int h^2(x)dx - A^2)/n.$$

B. Application

The method described in Part A of estimating the area of an irregularly shaped region is easily applied in practice.

The following is a step by step procedure for taking the sample and obtaining the estimate $\hat{A}$ and an estimate of its variance. In Section C to follow a method applying $\hat{A}$ to a systematic sample is introduced.
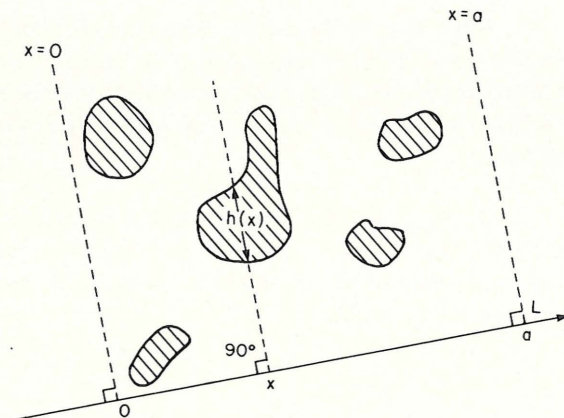
Fig. 4

Sampling Procedure

(1) Construct a base line L corresponding to the X-axis of Part A. (Fig. 4) The orientation of the base line will affect the variance of the area estimate. Obviously, the optimum orientation is that which minimizes the variation in h(x) as one moves along the base line.

(2) Pick a point of origin '0' and a point 'A' on the base line so that the entire area of interest is contained between the parallel lines through these points and at right angles to the line L. The points 0 and A on the line L should be as close together as possible while still including the entire area of interest between the lines X=0 and X=A.

(3) Having established the base line and the points 0 and A, the next step is to select a sample of points along the line L between 0 and A. This can be done as follows.

Let D be the distance between 0 and A. Then select n random numbers $\{r_1,...,r_n\}$ between 0 and 1, either using a random number generator on a computer or by selecting the numbers from random number tables. Define
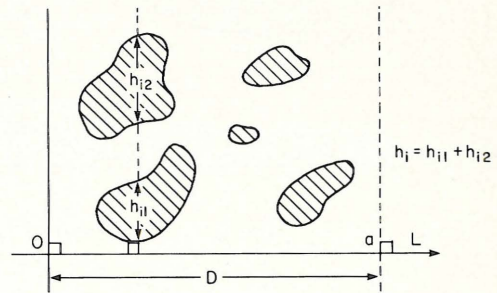
$$X_1 = Dr_1, X_2 = Dr_2, ..., X_n = Dr_n.$$

Fig. 5

(4) Locate the points $X_1,...,X_n$ between 0 and A on L. For each point X_i measure the length of intersection h_i of the line $X=X_i$ with the region R. Note that h_i may be the sum of more than one intersection i.e. the intersection with more than one subregion as shown in Fig. 5.

(5) Calculate the estimate

$$\hat{A} = D\bar{h}$$

and the variance estimate

$$\hat{V} = D^2 \Sigma (h_i - \bar{h})^2/n(n-1),$$

where

$$\bar{h} = \Sigma h_i/n .$$

NOTE: In the field, as one travels the line perpendicular to the base line through the point X_i each length of intersection h_{i1}, h_{i2}, etc. will be recorded and combined to give $h_i = h_{i1} + h_{i2} + ...$ (See Fig. 5)
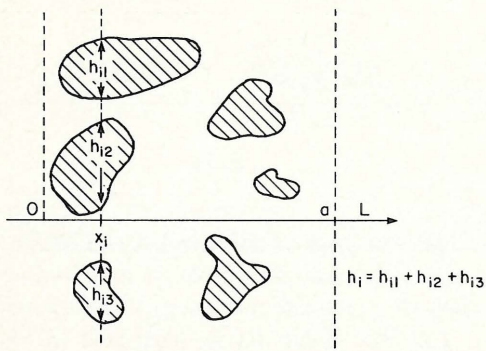

Fig. 6

Another point worth noting is that some of the subregions may lie on either side of the base line. In this case we define h_i to be the sum of the lengths of intersections obtained by travelling the line through the point X_i and perpendicular to L in both directions, as indicated by Fig. 6.
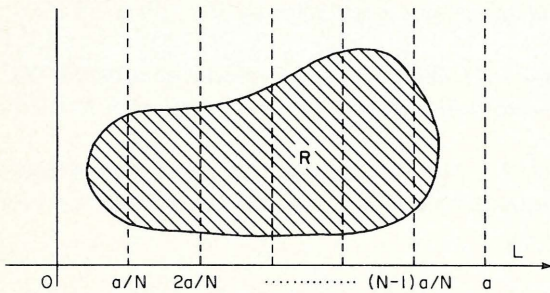
Fig. 7

C. Estimating the area of R with a systematic line sample.

(1) Theory: We shall only consider the simplest case of estimating the area of the region R of Fig. 7. The generalization to more complex shaped regions follows as for the previous estimator based on a random sample.
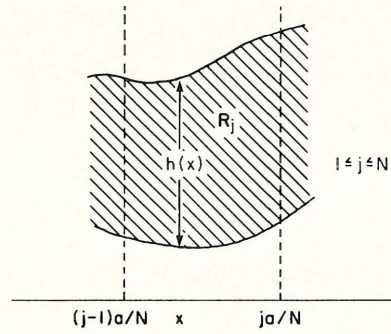

Fig. 8

Referring to Fig. 7, divide the base line (0,a) into N intervals of equal length and consider an arbitrary interval of length a/N, as in Fig. 8. Denote the area of R_j by A_j. Then if X is a random variable from the uniform distribution with parameter $\theta = a/N$, we have from Section A,

$$A_j = \int_{(j-1)a/N}^{ja/N} h(x)dx$$

and $\hat{A}_j$ defined by

$$\hat{A}_j = (a/N)h(X)$$

is an unbiased estimate of A_j. To see this note that

$$E(\hat{A}_j) = (a/N) \int (N/a)h(x)dx$$
$$= A_j$$

where the limits of integration are from $(j-1)a/N$ to ja/N.

Now, let $X_1 \sim$ Uniform $(0,a/N)$ and define $X_i = X_{i-1} + a/N$ for $i = 2,...,N$. Then the statistics $\hat{A}$ defined by

$$\hat{A} = (a/N) \Sigma h(X_i)$$

is an unbiased estimate of the area of the entire region R. That this is true is demonstrated below.

$$E(\hat{A}) = \Sigma \int_0^{a/N} h(x_i)dx_i$$

$$= \int_0^{a/N} h(x_1)dx_1 + \int_0^{a/N} h(x_1 + a/N)dx_1 + ...$$

$$+ \int_0^{a/N} h(x_1 + a(N-1)/N)dx_1$$

$$= \int_0^{a/N} h(x_1)dx_1 + ... + \int_{(j-1)a/N}^{ja/N} h(x_1)dx_1$$

$$+ ... + \int_{(N-1)a/N}^{a} (x_1)dx_1$$

$$= A_1 + ... + A_j + ... + A_n$$

$$= A.$$

Note that $\hat{A} = \Sigma \hat{A}_j$ and that we will not have Cov $(\hat{A}_i, \hat{A}_j) = 0$ in general. The variance of $\hat{A}$ is $Var(\hat{A}) = \Sigma Var(\hat{A}_j) + 2 \underset{i<j}{\Sigma} Cov(\hat{A}_i, \hat{A}_j)$. To obtain an estimate of the variance of $\hat{A}$ we can obtain two or more independent estimates of the area based on independent random starts. Let $\hat{T}_1,...,\hat{T}_m$ be m such estimates obtained by calculating $\hat{A}$ for each of m independently chosen Uniform $(0,a/N)$ random observations. Then the combined estimate $\hat{T} = \Sigma \hat{T}_j/m$ is unbiased and has variance estimate S^2 where

$$S^2 = \Sigma (\hat{T}_j - \hat{T})^2/m(m-1).$$

The variance of statistics based on systematic samples is often estimated by assuming certain distributional properties of the statistic or sampled population. See Cochran 1963, "Sampling Techniques", p. 224-227.

D. Application of the Systematic Sampling Method.

(1) Construct the base line L and locate the points 0 and a as in the steps 1 and 2 of Part B.
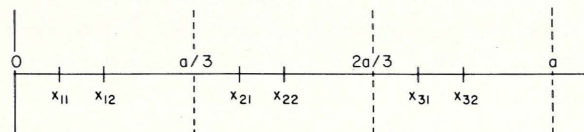
(2) Define d = (distance between 0 and a)/N.

Fig. 9

(3) Select m random numbers $\{r_1, r_2,...,r_m\}$ between 0 and 1. Define

$$x_{11} = r_1d, \; x_{12} = r_2d, \; ..., \; x_{1i} = r_1d, \; ..., \; x_{1m}=r_md$$

$$x_{21} = x_{11} + d, \; x_{22} = x_{12} + d, \; ..., \; x_{2m} = x_{lm} + d$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$x_{N1} = x_{(N-1)1} + d, \; ..., \; x_{Nm} = x_{(N-1)m} + d.$$

We have thus defined m sets of equally spaced points $\{x_{1i}, x_{2i},...,x_{Ni}\}_{i=1}^{m}$ along the base line. (Fig. 9).
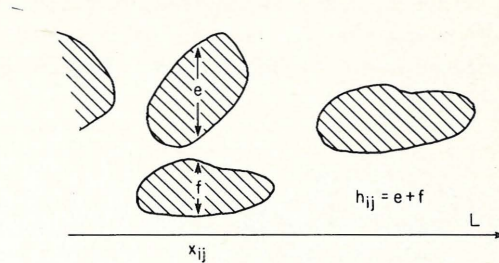

Fig. 10

(4) For each of the points x_{ij} occurring along the base line obtain the length of intersection with the region R of the line through x_{ij} and perpendicular to L. Call this length h_{ij} as illustrated in Fig. 10.

(5) We can now calculate m estimates $\hat{T}_1,...,\hat{T}_m$ of the area of the region R. That is, define

$$\hat{T}_1 = d\sum_{j=1}^{N} h_{j1}, \; \hat{T}_2 = d\sum_{j=1}^{N} h_{j2},..., \; \hat{T}_m = d\sum_{j=1}^{N} h_{jm}.$$

(6) The final estimate is given by $\hat{T} = \sum_{i=1}^{m} T_i/m$ and its variance is estimated by $\hat{V} = \sum_{i=1}^{m} (\hat{T}_i - \hat{T})^2/$ $m(m-1)$. Confidence intervals can be based on the Student t distribution with m−1 degrees of freedom.

NOTE: The first method (i.e. the random selection method) is a special case of the above when N = 1 and m = n.

In practice, we will generally have N much larger than m (m is usually 1, 2 or 3). If m=1 we obviously can not estimate the variance of the area estimate by the above method. In fact, if m = 1 no unbiased estimate of the variance can be obtained without making some additional assumptions. (Cochran 1963)

The systematic sampling method has the obvious advantage of ease of sampling (ie. locating sampling lines) and forces the sample to be repre-

sentative of the whole area. In practice one would expect a slightly smaller variance by the systematic procedure but this is not always the case. A disadvantage of the systematic method is that the variance estimate will usually have a small number of degrees of freedom. For example, if $m = 2$ the variance estimate will have 1 degree of freedom and consequently be fairly unstable. As a result, the confidence intervals constructed using the Student t distribution with $m-1$ d. f. will generally be long.

APPENDIX 2

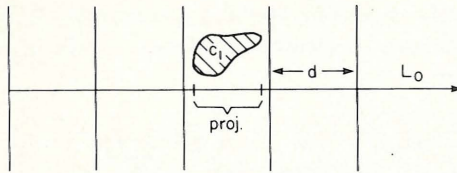## Investigating the Distribution of the Sizes of the Individual Infection Centres

Fig. 1

Theory

Consider a set, L, of parallel lines spaced a distance d apart and completely covering an area of interest. The probability that one of the lines of L will intersect with an infection centre is given by $p = \text{proj}/d$, where proj = the projection of the centre onto a line running perpendicular to the lines of L. (Fig. 1).

Prob (that a line will intersect C_1) = proj/d; proj $\geq$ d, prob = 1.0. In the above it was assumed that the centre C_1 was randomly located, or equivalently, the system of lines L was randomly located. By randomly located we mean randomly located with the restriction that the orientation of the centre to L_0 be unchanged. i.e. the projection onto L_0 remains constant.

Given an area containing a number of infection centres, construct a base line L_0. Then randomly locate the system of parallel lines L, and subject to the constraint that L be perpendicular to L_0. Let w_i be the projection of the ith centre onto L_0. The probability, p_i, of intersection of the ith centre and L is given by $p_i = w_i/d$.

Let x_0 be an arbitrary point on L_0 and let x be the distance from x_0 to the nearest line of L to the right of x_0. Now suppose x ~ uniform (0,d).

Then if we define

$$\xi_i(x) = \begin{cases} 1 & \text{L intersects } C_i \\ 0 & \text{otherwise} \end{cases}$$

we have the expected value of $\xi_i(x)$ given by

$$E_x(\xi_i(x)) = \int_0^d \frac{\xi_i(x)}{d} dx$$

$$= \int 1/d \, dx$$

$$\{x(\xi_i(x)) = 1\}$$

$$= w_i/d$$

$$= P_i$$

Estimating Totals

Let $f(C_i)$ be some characteristic of the centre C_i. We wish to estimate $\sum_{i=1}^{N} f(C_i)$, the total over all centres. To accomplish this let x_i be an observation from Uniform (0,d), then $\hat{T} = \sum_{i=1}^{N} \xi_i(x_i) \frac{f(C_i)}{P_i}$ is an unbiased estimate of the total $T = \sum_{i=1}^{N} f(C_i)$. Note that $\hat{T}$ can also be written as $\hat{T} = \sum^n f(C_i)/P_i$, where $\sum^n$ indicates the sum is over all centres which intersect L.

To show that $\hat{T}$ is an unbiased estimate of T we take the expectation with respect to X i.e.

$$E_x(\hat{T}) = E_x\left[\sum_{i=1}^{N} \xi_i(x)f(C_i)/P_i\right]$$

$$= \sum_{i=1}^{N} \frac{f(C_i)}{P_i} E_x[\xi_i(x)]$$

$$= \sum_{i=1}^{N} \frac{f(C_i)}{P_i} P_i$$

$$= \sum_{i=1}^{N} f(C_i).$$

We shall now obtain the variance of $\hat{T}$. First note that

$$Var_x(\xi_i(x)) = E_x[\xi_i(x)] - E_x[\xi_i(x)]^2$$

$$= P_i - P_i^2$$

$$= P_i(i - P_i)$$

and $Cov_x(\xi_i(x), \xi_j(x)) = E_x[\xi_i(x)\xi_j(x)] - P_iP_j$,

which can be written as

$Cov_x(\xi_i(x), \xi_j(x)) = P_{ij} - P_iP_j$, where P_{ij} is the probability that L intersects both C_i and C_j. Therefore,

$$Var_x(\hat{T}) = Var_x[\sum_{i=1}^{N} \xi_i(x)f(C_i)]$$

$$= \sum_{i=1}^{N} \frac{f^2(C_i)}{P_i^2} Var_x(\xi_i(x)) +$$

$$2 \sum_{i<j}^{N} \frac{f(C_i)f(C_j)}{P_iP_j} Cov_x(\xi_i(x),\xi_j(x))$$

$$= \sum_{i=1}^{N} \frac{f^2(C_i)(1-P_i)}{P_i} +$$

$$2 \sum_{i<j}^{N} \frac{f(C_i)f(C_j)}{P_iP_j} (P_{ij} - P_iP_j)$$

For notational convenience define $f(C_i) = f_i$ so that

$$Var_x(\hat{T}) = \sum_{i=1}^{N} f_i^2 \frac{(1-P_i)}{P_i} +$$

$$2 \sum_{i<j} f_i f_j \frac{(P_{ij}-P_iP_j)}{P_iP_j} \qquad (1)$$

Assuming that $f_i \geq 0$, the maximum variance of $\hat{T}$ occurs when P_{ij} is maximum i.e. when $P_{ij} = \min P_i, P_j$.

In this case we have

$$\text{Max of } Var_x(\hat{T}) = \sum_{i=1}^{N} f_i^2(1-P_i) +$$

$$2 \sum_i \sum_j f_i f_j \frac{(1-\max(P_i,P_j)}{\max(P_i,P_j)}. \qquad (2)$$

The minimum variance will occur when $P_{ij} = 0$ in which case

$$\text{Min. of } Var_x(\hat{T}) = \sum_{i=1}^{N} f_i^2 \frac{(1-P_i)}{P_i} - 2 \sum_{i<j} f_i f_j. \qquad (3)$$

If the population of centers is assumed to be randomly scattered about an area then it is not unreasonable to assume that $Cov_x(\xi_i(x), \xi_j(x))=0$ i.e. that

$$Var_x(\hat{T}) \underset{\text{approx.}}{\thickapprox} \sum_{i=1}^{N} f_i^2 \frac{(1-P_i)}{P_i}. \qquad (4)$$

Note that (4) can be estimated by replacing $f(C_i)$ in the formula for $\hat{T}$ by the expression $\frac{f_i^2(1-P_i)}{P_i}$.

An Application to Size Class Estimation.

Consider the problem of estimating the number of infection centres which fall into each of a number of size classes. Let C_i, C_m be the size classes and define $f_k(C_i) = 1$ if C_i belongs to size class C_k
$= 0$ otherwise.

Then we can estimate the total number of centres falling in size class C_k by

$$\hat{N}_k = \sum_{i=1}^{N} \frac{\xi_i(x) f_k(C_i)}{P_j} \qquad (5)$$

$$= \sum^{n'} 1/P_i$$

where $\sum^{n'}$ indicates the sum is over all centres which intersect L and belong to the size class C_k. (x is chosen randomly between 0 and d.)

The variance of $\hat{N}_k$ can be obtained from (1). If $K \neq J$, then

$$Cov_x(\hat{N}_k, \hat{N}_j) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{f_k(C_i) f_j(C_j)}{P_iP_j} .$$

$$Cov(\xi_i(x), \xi_j(x)).$$

Note that if we assume that the centres are randomly located then Cov ($\xi_i(x)$, $\xi_j(x)$) = 0 which yields Cov($\hat{N}_k$, $\hat{N}_j$) = 0.

If we are interested in the total area of infection for each size class, we merely define our functions $f_k(C_i)$ by

$$f_k(C_i) = \begin{cases} a_i - \text{the area of the } i^{th} \text{ centre if it belongs} \\ \qquad \text{to } C_k \\ 0 - \text{Otherwise} \end{cases}$$

Then the estimate of the total area for class C_k is

$$\hat{A}_k = \sum_{i=1}^{N} \frac{f_k(C_i) \; \xi_i(x)}{P_i}$$

$$= \Sigma^{n'} a_i/P_i$$

where Σ^n indicates the sum is over all centres which intersect L and belong to size class C_k.

$$\text{Cov}_x (\hat{A}_k, \hat{A}_j) = \sum_{i,j} \frac{a_i a_j}{P_i P_j} \text{Cov}(\xi_i(x), \xi_j(x)),$$

where the sum is over all pairs (i,j) of centres such that one centre is chosen from each of the classes C_k and C_j.

An estimate of the average size of the centres within a class C_k, say, can be obtained by using the ratio

$$\hat{R}_k = \hat{A}_k/\hat{N}_k.$$

This estimate, however, is not unbiased.

Note that when we have a single size class C_1, the estimates $\hat{N}_1$ and $\hat{A}_1$ are estimates of the total number of centres and their total area, respectively.

Estimating Variance

The case when $f(C_i)$ must be estimated.

In practice the measuring of the area of an infection centre is itself an estimation problem. For notational convenience let $F_i = f(C_i)$ be the value of the characteristic of interest associated with the ith centre (i.e. area).

Let $\hat{F}_i$ be an unbiased estimate of F_i and suppose Cov($\hat{F}_i$, $\hat{F}_j$) = 0 for i ≠ j. Then

$$\hat{T} = \sum_{i=1}^{N} \xi_i \frac{(x)}{P_i} \hat{F}_i$$

is an unbiased estimate of $T = \Sigma F_i$.

To find the variance of $\hat{T}$ we use the identity

$$\text{Var}(f(x,y) = \text{Var}[E(f(x,y) \, / \, y)] + E[\text{Var}(f(x,y)/y)]$$

where x and y are random vectors.

Now $E(\hat{T} \; / \; \hat{F}) = \sum_{i=1}^{N} \hat{F}_i$ \hfill (6)

and

$$\text{Var}(\hat{T}/\hat{F}) = \sum_{i=1}^{N} \hat{F}_i^2 \frac{(1-P_i)}{P_i} + \qquad (7)$$

$$2\Sigma \sum_{i<j} \frac{\hat{F}_i \hat{F}_j}{P_i P_j}(P_{ij} - P_i P_j) \qquad (8)$$

The variance of (6) is $\text{Var}(\sum_{i=1}^{N} \hat{F}_i) = \sum_{i=1}^{N} \sigma_{fi}^2$ where $\sigma_{fi}^2 = \text{Var}(\hat{F}_i)$.

The expected value of (7) is given by

$$E(\text{Var}(\hat{T} \mid \hat{F})) = \sum_{i=1}^{N} (\sigma_{fi}^2 + F_i^2)\frac{(1-P_i)}{P_i} +$$

$$2 \sum_{i<j} \sum \frac{F_i F_j}{P_i P_j} (P_{ij} - P_i P_j).$$

Therefore Var ($\hat{T}$) is

$$\text{Var}(\hat{\sigma}) = \sum_{i=1}^{N} \sigma^2_{fi}/P_i + \sum_{i=1} \frac{F_i^2(1-P_i)}{P_i} +$$

$$2\Sigma \sum_{i<j} \frac{F_i F_j}{P_i P_j} (P_{ij} - P_i P_j).$$

Note that the increased variability which results from having to estimate the function $f(C_i)$ is the first term on the right hand side of equation (8).

It may be reasonable to assume that $\sigma_{fi}^2/F_i^2 = K^2$, that is, the coefficient of variation is the

same for each centre. In this case (8) becomes

$$Var(\hat{T}) = \sum_{i=1}^{N} \frac{F_i^2}{P_i}(1 + K^2 - P_i) + \qquad (9)$$

$$2 \sum_{i<j}^{N} \frac{F_i F_j (P_{ij} - P_i P_j)}{P_i P_j}.$$

It is apparent from (9) that if the coefficient of variation K is reasonably small ($\leq 20\%$ say) the Var $(\hat{T})$ is inflated very little. This is because $1 + K^2 - P_i = (1 - P_i) + K^2$ where $1 - P_i \approx 1$ i.e.

$$\frac{K^2}{(1 - P_i)} \approx K^2 \ (\leq .04)).$$

An Alternative Variance Estimate

If two or more independent estimates $\hat{T}_i,\text{--},$ $\hat{T}_n$ of the total can be obtained then the variance can be estimated by the usual sample variance

$$S^2 = \sum_{i=1}^{n} (\hat{T}_i - \hat{T})^2/(n-1),$$

where $\hat{T} = \sum_{i=1}^{n} \hat{T}_i/n$ is the sample mean.

$$Var\ (T) = S^2/n.$$

The independent estimates $\hat{T}_1, \text{--}, \hat{T}_n$ can be obtained by randomly selecting n observations $x_i, \text{--}, x_n$ from the uniform (0,d) distribution and calculating $\hat{T}_i = \hat{T}(x_i)$, i.e. by calculating the estimates which result from each of n random starts.

Increasing the Probability of Selection

We have seen that the probability of selecting a centre C_i is given by $P_i = W_i/d$, where W_i is the projection of C_i onto the base line L_0. An obvious method of increasing P_i is to decrease d, the distance between the lines L. An alternative method is to consider a set of strips L_r where the width of the strip is r and the distance from the centre of a strip to that of its nearest neighbour is d.

Then the probability of selecting C_i (i.e. the probability that C_i intersects L) is $P_i = (W_i + r)/d$. ($W_i \leq d - r$)

Note that we can increase the probability of selection substantially by the method, and leave travelling distance practically unchanged.
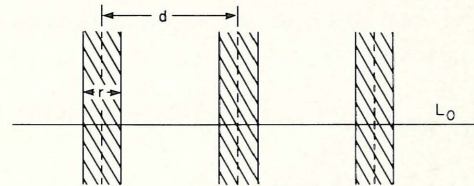

Fig. 2

Consider the diagram of an infection centre given in Fig. 2.

Define

x_i to be the distance of the left-most point of the infection centre from the centre line of the intersecting strip with the convention that distances to the right of the centre line be recorded as negative and those on the left be recorded as positive. In the diagram x_i will be negative.

y_i to be the distance of the right-most point of the infection centre from the centre line of the intersecting strip with the convention that distances to the left of the centre line be recorded as negative and those on the right be recorded as positive. In the diagram y_i is positive.

NOTE: The length of the projection of the centre onto the base line L_0 is given by $W_i = x_i + y_i$

r to be the width of the strip

d to be the distance between centres of the strips.

APPENDIX 3

Estimating Area - a radial sampling method

Fig. 1

Fig. 2

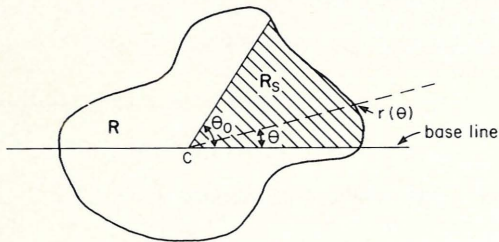### A. Theory:

Consider the problem of estimating the area of a region R contained within a sector $(0 < \theta < \theta_o)$, (Fig. 1)

If we define $r(\theta)$ to be the distance from C to the boundary of R along the line through C at an angle of θ to the base line, the the area, A_s, of R_s is given by

$$A_s = \iint_{R_s} r \, dr \, d\theta = \int_{\theta=0}^{\theta=\theta_o} \left[\int_{r=0}^{r=r(\theta)} r \, dr \right] d\theta$$

$$= \int_{\theta=0}^{\theta=\theta_o} \frac{r^2(\theta)}{2} \, d\theta.$$

Now, suppose θ is a Uniform $(0, \theta_o)$ random variable. Then the function $h(\theta) = \theta_o r^2(\theta)/2$ has expected value A_s. That is $h(\theta)$ provides an unbiased estimate of A_s. To see that this is true we write

$$E(h(\theta)) = \int h(\theta) \times f(\theta) \, d\theta, \text{ where } f(\theta) \text{ is the density function of the uniform distribution on } (0, \theta_o).$$

$$= \int_0^{\theta_o} \frac{\theta_o r^2(\theta)}{2} \times \frac{1}{\theta_o} \, d\theta$$

$$= \int_0^{\theta_o} \frac{r^2(\theta)}{2} \, d\theta$$
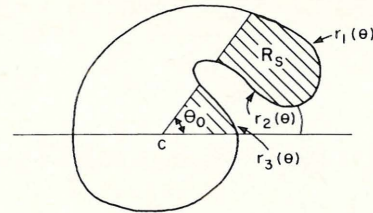
$$= A_s$$

Consider the following situation (Fig. 2)

Assuming $\theta \sim$ Uniform $(0, \theta_o)$ we can calculate A_s by

$$A_s = \int_0^{\theta_o} \frac{r_1^2(\theta)}{2} \, d\theta - \int_0^{\theta_o} \frac{r_2^2(\theta)}{2} \, d\theta +$$

$$\int_0^{\theta_o} \frac{r_3^2(\theta)}{2} \, d\theta$$

$$= \frac{1}{2} \int_0^{\theta_o} (r_1^2(\theta) + r_3^2(\theta) - r_2^2(\theta)) \, d\theta.$$

Define the exterior border of the region R_s as that portion of the border for which the interior of R_s lies on the same side of the border as the point C. The interior border can then be defined as that portion of the border that is not an exterior border (Fig. 3).
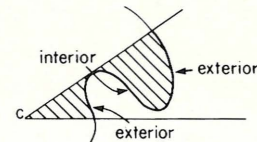

Fig. 3

We can now express the area A_s of R_s for the general case as

$$A_s = \frac{1}{2} \int_0^{\theta_o} \left(\Sigma^+ r_i^2(\theta) - \Sigma^- r_i^2(\theta) \right) d\theta,$$

where Σ^+ indicates that the sum is over all functions $r_i(\theta)$ which comprise the exterior border and Σ^- indicates that the sum is over all functions $r_i(\theta)$ which comprise the interior border.

If we define $h_i(\theta) = \frac{\theta_o r_i^2(\theta)}{2}$ and

$h(\theta) = \Sigma^+ h_i(\theta) - \Sigma^- h_i(\theta)$, then the expected

value of $h(\theta)$ is

$$E(h(\Theta)) = \int_0^{\Theta_o} [\frac{\Theta_o}{2} \Sigma^+ r_i^2(\Theta) - \frac{\Theta_o}{2} \Sigma^- r_i^2(\Theta)] \frac{1}{\Theta_o} \, d\Theta$$

$$= A_s$$

Estimation of A_s by systematic sampling

Let $\alpha = \Theta_o/R$ and let $X \sim$ Uniform $(0, \alpha)$. Then let x_1 be an observation from the distribution of X and define

$$x_2 = x_1 + \alpha \quad x_3 = x_2 + \alpha , \cdots , x_n = x_{n-1} + \alpha$$

Define $h_j(x_i) = \alpha \, r_j^2(x_i)/2$

$$h(x_i) = \sum_j^+ h(x_i) - \sum_j^- h_j(x_i)$$

Then

$$E(h(x_i)) = \int \frac{\alpha}{2} [\sum_j^+ r_j^2(x_i) - \sum_j^- r_j^2(x_i)] \frac{1}{\alpha} \, dx_i$$

$$= A_i \text{ (the area of the region R, contained in the } i^{th} \text{ sector.)}$$

From the above, it is easily verified that

$$\hat{A}_s = \sum_{i=1}^n h(x_i)$$

is an unbiased estimate of A_s ($= \Sigma A_i$).

We can write $\hat{A}_s$ in an alternative form as

$$\hat{A}_s = \alpha/2 \sum_{i=1}^n \sum_{j=1}^k \xi_j(x_i) \, r_j^2(x_i);$$

where $\xi_j(x_i) = 1$ if r_j comprises part of the exterior boundary

$\qquad = -1$ if r_j comprises part of the interior boundary

and $K =$ the number of functions r_j.

NOTE: K will usually be equal to 1 — always true for convex regions.

Comments

The radial method is most easily applied to regions of fairly regular shape (circle). With the point of origin, C, located roughly in the centre of the region and $\Theta_o = 360^o$ the resulting estimate A_s is an estimate of the area of the whole region R.

By choosing two or more random starts an estimate of variance can be obtained as for the systematic line sampling method. (Fig. 4)
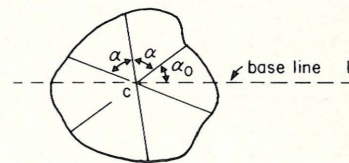

Fig. 4

Canada