634.0.524.63:519.241.2

MEDEDELINGEN LANDBOUWHOGESCHOOL
WAGENINGEN • NEDERLAND • 73-19 (1973)

LINE INTERSECT SAMPLING OVER POPULATIONS OF ARBITRARILY SHAPED ELEMENTS

with special reference to forestry

P. G. DE VRIES
Department of Forest Mensuration

A. C. VAN EIJNSBERGEN
Department of Mathematics

Agricultural University
Wageningen, The Netherlands

(Received 3-X-1973)

CONTENTS

1. INTRODUCTION

A relatively new (WARREN and OLSEN, 1963; VAN WAGNER, 1968) sampling technique in forest inventory, named 'line intersect sampling' appeared to be an application of a theme known as 'BUFFON's needle problem' in probability theory, as was shown by DE VRIES (1973).

In 1777 G. BUFFON in his 'Essai d'arithmétique morale' published the solution to the following problem: 'On a plane on which parallel straight lines at equal mutual distances of W units have been drawn, a straight thin needle of length $\lambda \leqslant W$ is randomly thrown. How large is the probability p that the needle will intersect with a line?' The well-known answer is:

$$p = 2 \cdot \lambda / \pi W \tag{1}$$

KENDALL and MORAN (1963) treat these types of problems in a general fundamental way, listing some 200 references, but they give few solutions to special cases though the latter may be derived from the above authors' general approach.

Line intersect sampling is based on observations made only on those elements of a population, that intersect with a random line through the population; it provides estimates of the total of an observed characteristic (till now generally volume of pieces of cut timber) per unit area, and its variance. However, among the elements that occur in real life situations (for instance the axes of timber logs or branches), many cannot be identified with a straight needle as in the classical BUFFON case. This circumstance induces an element of doubt in the application of line intersect sampling, as questions arise such as: 'How to deal with elements that intersect more than once with the sampling line?', 'Which length measure should be taken from an intersecting arbitrarily curved element in order that the method validly applies?', etc.

The authors herewith present a theoretical basis on which decisions can be made in the above and similar practical situations.

2. ESTIMATION OVER POPULATIONS OF ARBITRARY CURVE SEGMENTS

We assume a population of N elements, numbered $i = 1, \ldots, N$ of lengths λ_i on an area of size A. The element i possesses a characteristic x with value x_i, and we require an estimator $\underset{\sim}{e}(X)$ for the mean quantity of x per unit area, i.e. for:

$$X = (1/A) \sum^N x_i \tag{2}$$

The position of an element i on A can be described by the position $\underline{m}_i$ of a fixed point located on, and a direction $\underline{\theta}_i$ associated with that element. It is further assumed that the probability of $\underline{m}_i$ falling in any subarea of size B is B/A, and that $\underline{\theta}_i$ is uniformly distributed over the range [0, 2π].

The estimator will be based on data obtained by the following sampling design: All N elements of the population will be considered in succession, and by means of N stochastically independent variables $\underline{t}_1, \ldots, \underline{t}_N$ it is decided how many times $\underline{t}_i$ a particular element i will be included in the sample. For convenience, the latter procedure will be referred to as 'the (stochastic) rule $\underline{t}_i$'. Then an unbiased estimator of (2) is:

$$\underline{e}(X) = (1/A)\sum^N x_i \underline{t}_i/\mathscr{E}\underline{t}_i = (1/A)\sum^{\underline{n}} x_j \underline{t}_j/\mathscr{E}\underline{t}_j \tag{3}$$

where $\mathscr{E}\underline{t}_i$ is the expectation of $\underline{t}_i$, and $\underline{n}$ is the number of elements, numbered $j = 1, \ldots, \underline{n}$ that constitute the sample.

From (3) we derive the expression for the variance of $\underline{e}(X)$:

$$\mathrm{Var}\ \underline{e}(X) = (1/A)^2\sum^N(x_i/\mathscr{E}\underline{t}_i)^2\mathrm{Var}\ \underline{t}_i = (1/A)^2\sum^N x_i^2\ [CV(\underline{t}_i)]^2 \tag{4}$$

with

$$\mathrm{Var}\ \underline{t}_i = \mathscr{E}\underline{t}_i^2 - (\mathscr{E}\underline{t}_i)^2$$

so that:

$$\mathrm{Var}\ \underline{e}(X) = (1/A)^2 \sum^N x_i^2\ [(\mathscr{E}\underline{t}_i^2/(\mathscr{E}\underline{t}_i)^2) - 1] \tag{5}$$

The rule $\underline{t}_i$ is based on the intersection of an element i, randomly located as described above, with a straight sampling line of length L, the latter being chosen arbitrarily within the area of size A. If we now consider a rectangular area of size WL, symmetrically around L and within A, the probability that the element i will intersect with L is given by:

$$p_i = (WL/A)\ \tau_i$$

where τ_i is the probability of intersection of element i with L, given that i (i.e. $\underline{m}_i$) is in WL.

The following case was considered by DE VRIES (1973). If the elements are straight line segments of length λ_i, the rule $\underline{t}_i$ is specified as follows: $\underline{t}_i = 1$ in case of intersection, and $\underline{t}_i = 0$ otherwise (Fig. 1A). Then:

$$\mathscr{E}\underline{t}_i = \mathscr{E}\underline{t}_i^2 = (0)p(\underline{t}_i = 0) + (1)p(\underline{t}_i = 1) = p(\underline{t}_i = 1) = p_i = (WL/A)\tau_i$$

or by (1):

$$p_i = (WL/A)\ (2\lambda_i/\pi W) = 2\lambda_i L/\pi A$$

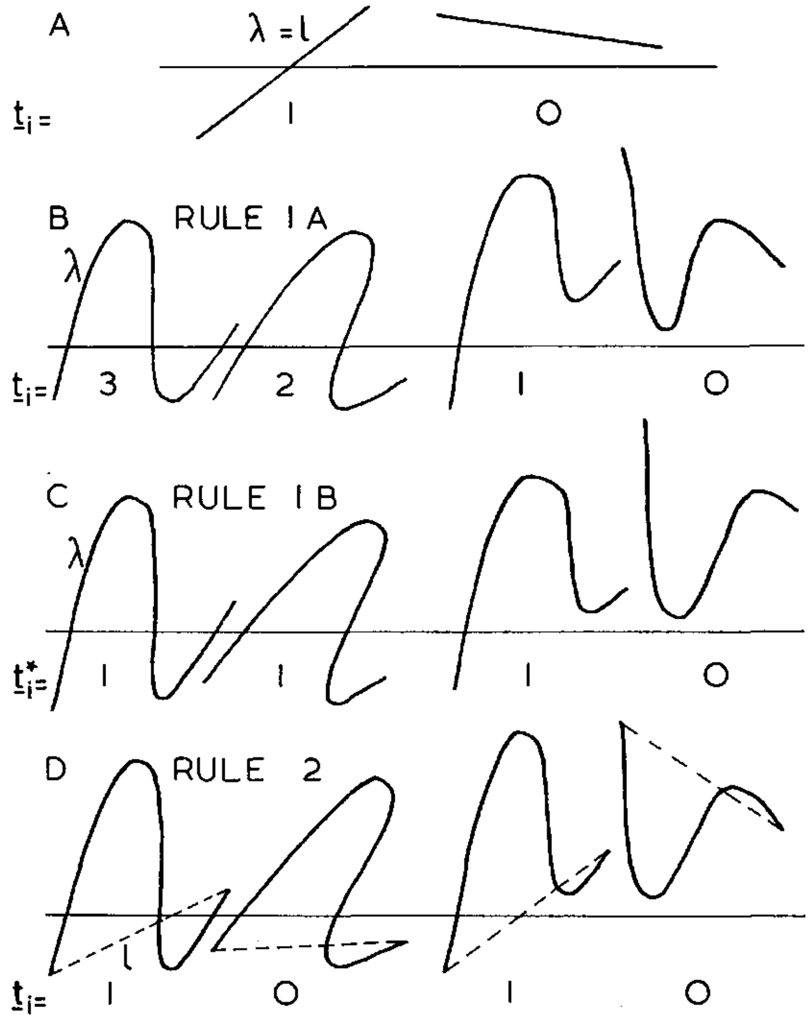


Fig. 1. Stochastic rules for segment of arbitrary curve and line.

It follows that:

$$\text{Var } \underline{t}_i = p_i\,(1-p_i) \simeq p_i \text{ (for } p_i << 1)$$

Consequently we have by (3):

$$\underline{e}(X) = (\pi/2L)\textstyle\sum^n x_j/\lambda_j \tag{6a}$$

and by (5):

$$\text{Var } \underline{e}(X) = (1/A)^2 \textstyle\sum^N x_i^2 \left[(\pi A/2\lambda_i L) - 1\right] \tag{6b}$$

which by (4) for $p_i << 1$ reduces to:

$$\text{Var } \underline{e}(X) \simeq (\pi/2L)\,(1/A) \sum^N x_i^2/\lambda_i \tag{6c}$$

estimated (in a similar way as (2) is estimated by (6a)) by:

$$\text{var } \underline{e}(X) = (\pi/2L)^2 \sum^n (x_j/\lambda_j)^2 \tag{6d}$$

We note that the expressions (6) are based on the classical BUFFON problem, and that (6a, d) are independent of W and A. By proper expansion the latter formulas of course yield estimates for population total. For $p_i << 1$, $\mathscr{E}\underline{t}_i = p_i \simeq \text{Var } \underline{t}_i$, a property reminescent of the POISSON distribution.

We will now consider populations in which the elements are segments of length λ_i of arbitrary plane curves. Then there are various options to define the stochastic rule $\underline{t}_i$, viz.:

Rule 1a: $\underline{t}_i = 0$ in case of non-intersection, and $\underline{t}_i = 1, 2, \ldots, k$ in case the i th element intersects $1, 2, \ldots, k$ times (fig. 1 B).

One would expect that without knowledge of the curve's parameters, the expected values $\mathscr{E}\underline{t}_i$ and $\mathscr{E}\underline{t}_i^2$ could not be calculated. This is true indeed as far as $\mathscr{E}\underline{t}_i^2$ is concerned, but remarkably we have (KENDALL and MORAN, 1963) an expression for the expected number of intersections:

$$\mathscr{E}\underline{t}_i = (WL/A)\,(2\lambda_i/\pi W) = 2\lambda_i(L/\pi A) \tag{7}$$

which obviously only depends on the curve's length, not on its shape. So by (3) we have here the unbiased estimator:

$$\underline{e}_{1a}\,(X) = (\pi/2L)\sum^N (x_i/\lambda_i)\underline{t}_i = (\pi/2L)\sum^n (x_j/\lambda_j)\underline{t}_j \tag{8}$$

For lack of knowledge of $\mathscr{E}\underline{t}_i^2$ we have no expression for the variance of (8) unless the type of curve defining the elements is known. This will be shown in the next section for the simple case of the mother curve being a circle. It is noted that actually (6) is already an illustration.

Rule 1b: $\underline{t}_i^* = 1$ in case of intersection, irrespective whether the latter is single or multiple, and $\underline{t}_i^* = 0$ otherwise (fig. 1C).

We use an asterisk to denote that here the probability that the i th element intersects, is equal to the sum of the probabilities of the i th element intersecting $1, 2, \ldots, k$ times in the preceding case, i.e.:

$$p(\underline{t}_i^* = 1) = \sum_{h=1}^{k} p(\underline{t}_i = h)$$

As (7) is not applicable, we have no expression for $\mathscr{E}\underline{t}_i^*$ or $\mathscr{E}\underline{t}_i^{*2}$ here, and consequently we cannot derive the estimator for total per unit area or its variance. However, if the shape of the curve segments is specified, the latter expressions which will be denoted by

$$\underline{e}_{1b}(X) \text{ and } \operatorname{Var} \underline{e}_{1b}(X)$$

respectively, can be found as will be shown for the simple case of circular arcs in the next section. Actually (6) is already an illustration again.

We note that if $\mathscr{E}\underline{t}_i{}^*$ were known, we would find:

$$\operatorname{Var} \underline{e}_{1b}(X) \leq \operatorname{Var} \underline{e}_{1a}(X)$$

If here we used an estimator of type (8), i.e.:

$$\underline{e}'_{1b}(X) = (\pi/2L)\sum^n (x_j/\lambda_j)\underline{t}^*_j \tag{9}$$

with of course $\underline{t}_j{}^* = 1$, this estimator will be biased, as (8) is not. This negative bias will be investigated for circular arcs in the next section.

Rule 2: A line segment (straight needle) of length l_i is uniquely defined with each element. Then $\underline{t}_i = 1$ in case l_i intersects, and $t_i = 0$ otherwise (fig. 1D).

Though the needle must be uniquely defined with each element, there is no restriction as to its location relative to an element. For instance the needle may be the line segment corresponding to the largest distance between two points on a curve segment. Her · the problem is reduced to the classical BUFFON case (1) with λ_i substituted by l_i, so we have by (6):

$$\underline{e}_2(X) = (\pi/2L)\sum x_j/l_j \tag{10a}$$

$$\operatorname{Var} \underline{e}_2(X) = (1/A)^2 \sum^N x_i^2 \left[(\pi A/2l_iL)-1\right] \tag{10b}$$

which for $p_i << 1$ reduces to:

$$\operatorname{Var} \underline{e}_2(X) \simeq (\pi/2L)(1/A)\sum^N x_i^2/l_i \tag{10c}$$

with its estimate:

$$\operatorname{var} \underline{e}_2(X) = (\pi/2L)^2 \sum^n (x_j/l_j)^2 \tag{10d}$$

We note that formulas (10) change into (6) for the elements being line segments.

3. ESTIMATION OVER POPULATIONS OF CIRCULAR ARCS

The precision of line intersect estimators based on the different definitions of the stochastic rule $\underline{t}_i$, cannot be evaluated for the general case of a random population consisting of segments of arbitrary curves, as in two cases (rules 1a, b) we have no expression for the variance, and in one case (rule 1 b) even an estimator for total per unit area is lacking. However if the shape of the curve segments is known, it is at least theoretically possible to construct estimators and variance expressions for line sampling designs based on all stochastic rules mentioned in section 2.

In order to avoid calculations that may tend to great complexity if assuming other shapes, we will consider the relatively simple case of a population of N circular arcs, the arcs being defined by their radii of curvature r_i and the angles α_i (in radians) they subtend at their centres of curvature.

Starting with rule 2, we define as the straight needle associated with the i th arc, its chord of length:

$l_i = 2r_i \sin \alpha_i/2 = \lambda_i (\sin \alpha_i/2)/(\alpha_i/2) = \lambda_i q_i$

where $\lambda_i = \alpha_i r_i$ and $q_i = l_i/\lambda_i = (\sin \alpha_i/2)/(\alpha_i/2)$

Then by (10a) through (10d) we have, adding a subscript 'o' to refer to a population of circular arcs:

$$\underline{e}_{2o}(X) = (\pi/2L) \sum^{n} x_j/\lambda_j q_j \tag{11a}$$

$$\text{Var } \underline{e}_{2o}(X) = (1/A)^2 \sum^{N} x_i^2 [(\pi A/2L\lambda_i q_i) - 1] \tag{11b}$$

which for $p_i << 1$ reduces to:

$$\text{Var } \underline{e}_{2o}(X) \simeq (\pi/2L)(1/A) \sum^{N} x_i^2/\lambda_i q_i \tag{11c}$$

with its estimate:

$$\text{var } \underline{e}_{2o}(X) = (\pi/2L)^2 \sum^{n} (x_j/\lambda_j q_j)^2 \tag{11d}$$

The construction of the estimator $\underline{e}_{1b}(X)$ and the expressions for Var $\underline{e}_{1b}(X)$ and Var $\underline{e}_{1a}(X)$ for circular arcs is slightly more elaborate. Intersection of the sampling line with a randomly thrown arc occurs if the projection P of that arc onto a line of length W perpendicular to the sampling line, intersects with the latter. For a given orientation φ of the arc, the conditional probability that the arc intersects with the sampling line is:

$$(p|\varphi) = (WL/A)(\tau|\varphi) = (WL/A)(P|\varphi)/W = (L/A)(P|\varphi)$$

and for all possible orientations the expected probability of intersection is:

$$p = \mathscr{E}_\varphi(p|\underline{\varphi}) = (L/A)\mathscr{E}_\varphi(P|\underline{\varphi}) = (L/A)\mathscr{E}P \tag{12}$$

So we have to find $\mathscr{E}\,P$ for a circular arc with parameters r and α. To this end we will indicate the orientation of the arc relative to the sampling line by the angle φ between the arc's chord $l = \lambda q$, and the line segment $W \perp L$. For reasons of symmetry it is sufficient to find $\mathscr{E}\,P$ for angles in the interval $0 \leqslant \varphi < \pi$, over which range φ is uniformly distributed with probability density $1/\pi$. Further, for reasons that will be clear presently, the interval for φ is divided into 4 subranges, viz.:

a. $0 \leqslant \varphi < \pi/2 - \alpha/2$ where the arc is projected 'singly' (13a)
b. $\pi/2 - \alpha/2 \leqslant \varphi < \pi/2$ where part of the arc is projected 'doubly' (Fig. 2)
c. $\pi/2 \leqslant \varphi < \pi/2 + \alpha/2$ where the arc or part of it is projected 'doubly'
d. $\pi/2 + \alpha/2 \leqslant \varphi < \pi$ where the arc is projected 'singly'

The projection P of an arc can always be divided in two parts P_1 and P_2, where P_1 corresponds to 'single' projection, and P_2 to 'double' projection. The values of P_1 and P_2 within the different ranges (13a) are:

a. $P_1 = \lambda q \cos \varphi$ $\quad P_2 = 0$ (13b)
b. $P_1 = \lambda q \cos \varphi$ $\quad P_2 = (\lambda/\alpha)\,(1 - \sin(\varphi + \alpha/2))$
c. $P_1 = -\lambda q \cos \varphi$ $\quad P_2 = (\lambda/\alpha)\,(1 - \sin(\varphi - \alpha/2))$
d. $P_1 = -\lambda q \cos \varphi$ $\quad P_2 = 0$

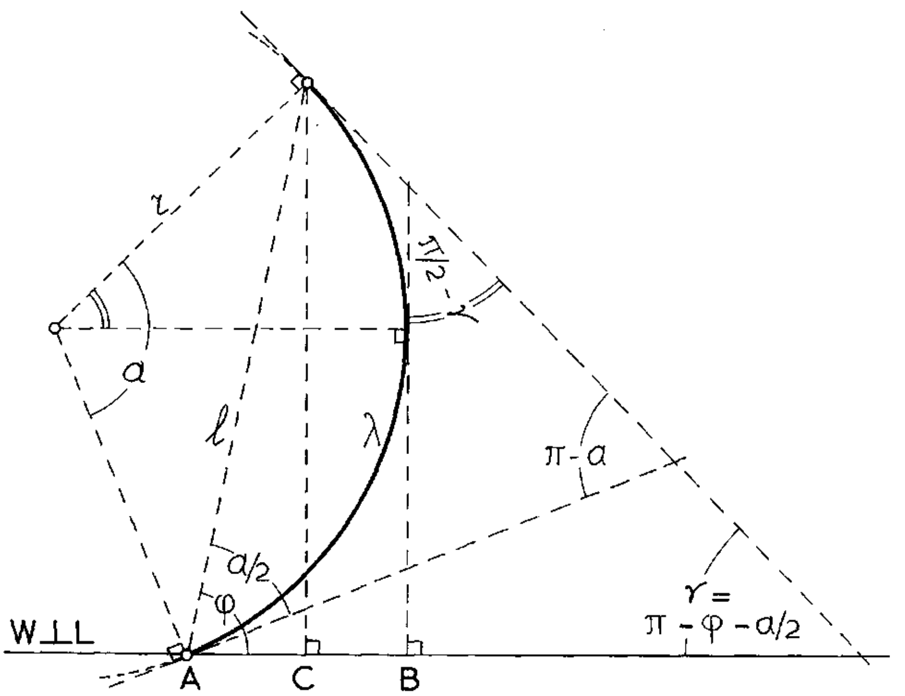


Fig. 2. Projection of circular arc for range $\pi/2 - \alpha/2 < \gamma \leqslant \pi/2$ or $\pi/2 - \alpha/2 \leqslant \varphi < \pi/2$. P = AC + CB = $P_1 + P_2$ (in 13b)

Then:

$$\mathscr{E}\, P = (1/\pi) \int_0^\pi (P_1 + P_2)\, d\, \varphi$$

It follows that:

$$\mathscr{E}\, P = \lambda(1+q)/\pi = 2\, l_m/\pi \tag{14}$$

where

$$l_m = (\lambda + l)/2 = \lambda(1+q)/2$$

So by (12, 14) we have for the probability of intersection (regardless whether single or multiple) of the i th arc (rule 1b, section 2):

$$p_i = (L/A)\,(2\, l_{mi}/\pi) = \lambda_i\,(1+q_i)\,(L/\pi A) \tag{15}$$

and consequently:

$$\mathscr{E} t_i^* = \mathscr{E} t_i^{*2} = p_i \text{ (in (15))}$$

$$\text{Var } t_i^* = p_i(1-p_i) \simeq p_i \text{ (for } p_i << 1)$$

Hence by (3):

$$\underset{\sim}{\varrho}_{1bo}(X) = (\pi/2L) \textstyle\sum^n x_j/l_{mj} \tag{16a}$$

and by (5):

$$\text{Var } \underset{\sim}{\varrho}_{1bo}(X) = (1/A)^2 \textstyle\sum^N x_i^2\, [(\pi A/2Ll_{mi}) - 1] \tag{16b}$$

which for $p_i << 1$ reduces to:

$$\begin{aligned} \text{Var } \underset{\sim}{\varrho}_{1bo}(X) &= (\pi/2L)\,(1/A) \textstyle\sum^N x_i^2/l_{mi} = \\ &= (\pi/2L)\,(1/A) \textstyle\sum^N 2x_i^2/\lambda_i\,(1+q_i) \end{aligned} \tag{16c}$$

with its estimate:

$$\text{var } \underset{\sim}{\varrho}_{1bo}(X) = (\pi/2L)^2 \textstyle\sum^n (x_j/l_{mj})^2 \tag{16d}$$

On comparison of formulas (16) with (6) and (10) it is noted that in (16) the mean l_{mi} of arc and chord lengths plays the same rôle as the length λ_i in case of line segments (6), or as the length l_i of a uniquely defined needle associated with an arbitrary curve (10).

The second terms (P_2) of the projections in (13b) correspond with 'doubly' projected parts of the arc; e.g. segment BC in Fig. 2 corresponds with P_2 in range b. If the sampling line would intersect e.g. BC in Fig. 2, this would imply a double intersection of arc and sampling line. If we then apply the stochastic rule 1a, section 2, we decide that $\underline{t}_i = 2$. The probability of double intersection of the sampling line with a randomly located arc, given its orientation φ, is:

$$(p''|\varphi) = (WL/A)\,(P_2|\varphi)/W = (L/A)\,(P_2|\varphi)$$

and the expected probability of double intersection over all orientations is:

$$p'' = \mathscr{E}_\varphi(p''|\underline{\varphi}) = (L/A)\,\mathscr{E}_\varphi\,(P_2|\underline{\varphi}) = (L/A)\mathscr{E}P_2$$

or

$$p'' = (L/A)\,(1/\pi)\int_0^\pi P_2\,d\varphi = \lambda\,(1-q)\,(L/\pi A)$$

Analogously the probability of single intersection for the arc is found as:

$$p' = \mathscr{E}_\varphi\,(p'|\underline{\varphi}) = (L/A)\,\mathscr{E}P_1 = (L/A)\,(1/\pi)\int_0^\pi P_1\,d\varphi = 2\,\lambda\,q(L/\pi A)$$

The expected number of intersections for the i th arc then is:

$$\mathscr{E}\underline{t}_i = (0)p_i^{o} + (1)p_i' + (2)p_i'' = 2\,\lambda_i\,(L/\pi A) \tag{17}$$

So the general property, already mentioned in (7), has been proved here for circular arcs. We now also can find $\mathscr{E}\underline{t}_i^2$:

$$\mathscr{E}\underline{t}_i^2 = (0)p_i^o + (1)p_i' + (4)p_i'' = 2\lambda_i\,(2-q_i)\,(L/\pi A) \tag{18}$$

By (3, 5, 17, 18) we then have for rule 1a:

$$\underline{\varrho}_{1ao}(X) = (\pi/2L)\sum^n \underline{t}_j x_j/\lambda_j \tag{19a}$$

$$\mathrm{Var}\;\underline{\varrho}_{1ao}(X) = (1/A)^2\sum^N x_i^2\;[(\pi A(2-q_i)/2L\lambda_i)-1] \tag{19b}$$

Assuming both p_i' and p_i'' small relative to one, we may write:

$$\mathrm{Var}\;\underline{t}_i \simeq p_i' + 4p_i'' = \mathscr{E}\underline{t}_i^2 = (18)$$

which by (4) reduces (19b) to:

$$\mathrm{Var}\;\underline{\varrho}_{1ao}(X) \simeq (\pi/2L)\,(1/A)\sum^N (x_i^2/\lambda_i)\,(2-q_i) \tag{19c}$$

which quantity can be estimated by:

$$\text{var } \underline{e}_{1ao}(X) = (\pi/2L)^2 \sum^{\underline{n}} t_j (x_j/\lambda_j)^2 (2-q_j) \tag{19d}$$

where, as before, $\underline{n}$ is the number of intersecting arcs in the sample, and $\underline{t}_j$ is the number of times ($\underline{t}_j = 1$ or 2) that the j th arc in the sample intersects with the sampling line.

It is noted that all formulas derived for the estimator of total per unit area and its variance reduce to those in (6) if the curve segments reduce to line segments. Further, for full circles, we have that $\alpha_i = 2\pi$; $q_i = 0$; $\underline{t}_i = 0$ or 2, and $\lambda_i = \pi d_i$ where d_i is diameter. Hence we obtain from both (16) and (19) for full circles:

$$\underline{e}_{1ab}(X) = (1/L) \sum^{\underline{n}} x_j/d_j \tag{20a}$$

and for $p_i = p_i'' = d_i L/A << 1$:

$$\text{Var } \underline{e}_{1ab}(X) \simeq (1/L)(1/A) \sum^{N} x_i^2/d_i \tag{20b}$$

with its estimate:

$$\text{var } \underline{e}_{1ab}(X) = (1/L)^2 \sum^{\underline{n}} (x_j/d_j)^2 \tag{20c}$$

Summarizing, we have derived the following approximate variance expressions for line intersect sampling in populations of circular arcs:

rule 1b: Var $\underline{e}_{1bo}(X)$ (16c)

rule 1a: Var $\underline{e}_{1ao}(X)$ (19c)

rule 2: Var $\underline{e}_{2o}(X)$ (11c)

As a measure of 'relative precision' (*RP*) of the estimators obtained under the different rules t_i, we will only consider the ratio (for convenience multiplied by 100) of corresponding terms under the summation signs of their variances. As can be seen from (4) this ratio is equal to the squared ratio of the coefficients of variation of the different rules. We then have:

$$RP(1b, 1a) = 100\,(2-q)\,(1+q)/2 \tag{21a}$$

$$RP(1b, 2) = 100\,(1+q)/2q \tag{21b}$$

$$RP(1a, 2) = 100/q\,(2-q) \tag{21c}$$

We now can investigate also which bias is induced if (9) is used for circular arcs instead of (16). Using (9) of course comes to the same as using (6), so we

TABLE 1. Relative precisions of the $e(X)$ in populations of circular arcs for various α.

$\alpha°$	q	RP(1b, 1a)	RP(1b, 2)	RP(1a, 2)
0	1.000	100	100	100
10	.999	100	100	100
20	.995	100	100	100
30	.989	100	101	100
45	.974	101	101	100
60	.995	102	102	100
90	.900	104	105	101
135	.784	108	114	105
180	.637	111	128	115
225	.470	112	156	139
270	.300	110	217	196
315	.140	106	407	384
360	.000	100	∞	∞

will calculate the bias caused by considering a circular arc as a straight needle of length $\lambda = \alpha r = l/q$. Apart from the constant, each term under the summation sign in (6a) then has a negative bias of the type:

$$B = x/\lambda - x/l_m = (x/\lambda)\,(q-1)/(q+1)$$

or expressed as a percentage of a term in (16a):

$$B\,\% = 100\,(q-1)/2 \qquad (22)$$

and the 'relative precision' (here: bias in variance), expressed as 100 times the quotient of corresponding terms in (6c) and (16c) is:

$$RP(16,6) = 100\,(1+q)/2 \qquad (23)$$

TABLE 2. $B\%$ (22) and RP (23), using (6a, c) instead of (16a, c) for circular arcs.

$\alpha°$	$B\%$	RP(16,6)	$\alpha°$	$B\%$	RP(16,6)
10	-0.1	99.9	135	-10.8	89.2
20	-0.2	99.8	180	-18.2	81.8
30	-0.6	99.4	225	-26.5	73.5
45	-1.3	98.7	270	-35.0	65.0
60	-2.2	97.8	315	-43.0	57.0
90	-5.0	95.0	360	-50.0	50.0

As both (22) and (23) only depend on α, specification of the allowable bias percent implies specification of the maximum curvature (table 3) circular arcs of given r may possess, to allow (6) to be applied.

TABLE 3. Maximum lengths of arc (λ), chord (l) and rise ($a = r$ (1–cos $\alpha/2$)) if maximum absolute bias (22) of $B = 5\%$ ($\alpha = 90°$) is tolerated in using (6) for circular arcs instead of (16).

r	$\lambda = \pi r/2$	$l = 2r \sin \pi/4$	$a = r$ (1–cos $\pi/4$)
1	1.57	1.41	0.29
2	3.14	2.83	0.59
3	4.71	4.42	0.88
4	6.28	5.67	1.17
5	7.85	7.07	1.46
6	9.42	8.48	1.76
7	11.00	9.90	2.05
8	12.57	11.31	2.34
9	14.14	12.73	2.64

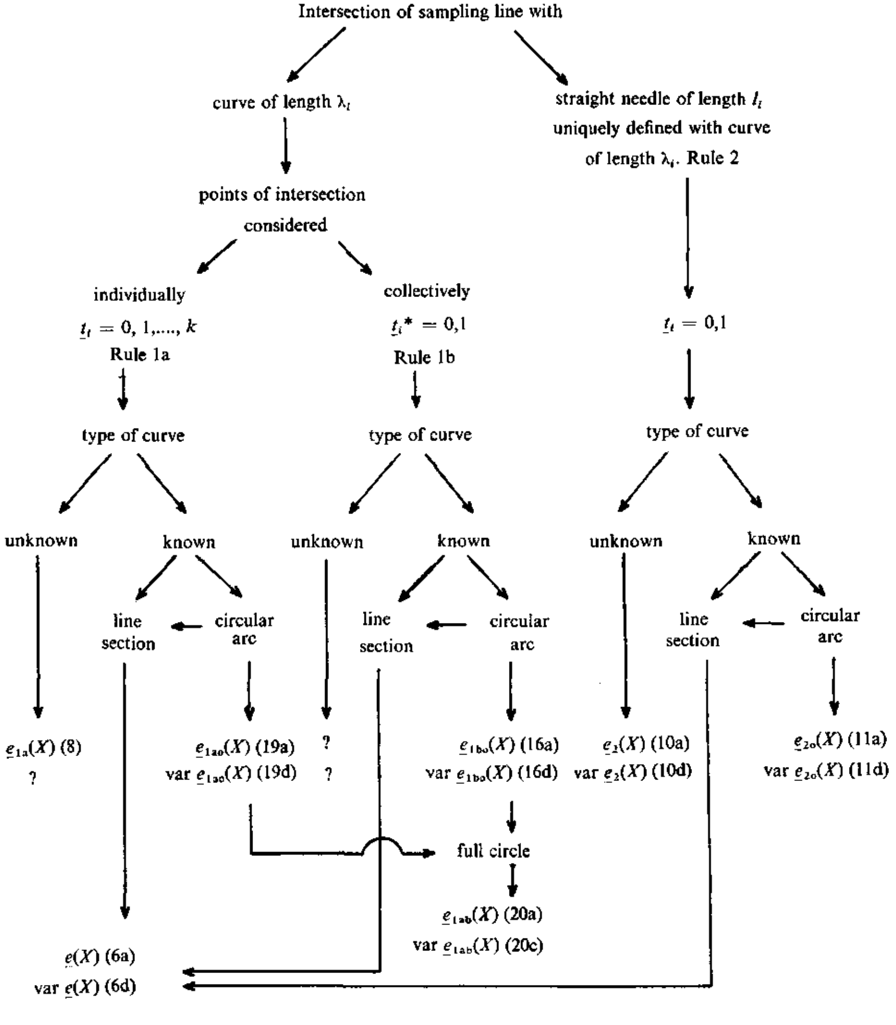
From table 3 it is evident that circular arcs with considerable rise may still be identified with straight needles if underestimations of total per unit area and its variance to the order of 5%, say, are tolerated.

4. DISCUSSION

4.1. Review of estimators in line intersect sampling

In the scheme below the estimators derived in the preceding sections are summarized.

Estimators $\underline{e}(X)$ (unbiased) and var $\underline{e}(X)$ (app. unbiased) in line intersect sampling under different stochastic rules, in populations of elements with known or unspecified shapes



4.2. EVALUATION

4.2.1. *Line intersect sampling under rule 1 b (Fig. 3A)*

For populations of circular arcs we have the unbiased $\underline{e}_{1bo}(X)$ (16a) as the best of the three estimators for total per unit area, as appears from (21a, b). For straight needles (line segments) and full circles this estimator and that of its variance (16d) automatically change into the corresponding expressions in (6) and (20) respectively.

For populations of arcs from other mother curves than circles, we have not tried to derive estimators, though such derivations should be theoretically possible. Consequently, application of the estimators in (16) to populations of other types of arcs leads to unpredictable bias, but no doubt bias will be negligible if only moderate curvature exists. In the latter case, by tables 2 and 3, we even may expect reasonable results from the estimators in (6), where all elements are identified with straight needles.

Under rule 1b all intersecting elements are considered, and any type of intersection is just given $\underline{t}_i^* = 1$. Length measurements always include λ, but (16) also implies measurement of chord length l in order to compute l_m. Chord length of course need not be measured if (6) is used as an approximation.

In branched elements it may sometimes be possible to define an arc-like shape of length λ, as indicated in Fig. 3A, so that the above estimators still can be applied, albeit with some reserve. As rule 1b is based on intersection with this λ, special attention then should be paid to the exclusion from the sample of branched elements that intersect invalidly. This may constitute a psychological drawback of this design.

For populations of arbitrarily shaped, non-arc like elements we have no estimators under rule 1b, and we cannot use (16) as an approximation either, as the measure l that should be taken to compute l_m, is not known. Maybe, if the shape of the elements is relatively simple and shows little curvature, bias will be small if, as in circular arcs here l also is taken as the length of the line connecting the curve's ends. The use of (6) in this case will lead to larger, but likewise unpredictable bias.

Summarizing: For populations of elements shaped like line segments, circular arcs and circles, or for populations in which the elements for practical purposes may be identified with these pure shapes, line intersect sampling under rule 1b provides (16a) as the best unbiased estimator for total per unit area, with estimated variance (16d). Identification of moderately curved circular arcs with line segments, i.e. the use of (6) reduces the number of measurements and will lead to only slightly biased results. For populations of arbitrarily shaped elements, no unbiased estimators exist in sampling under rule 1b.

The treatment of data obtained from more than one sampling line will be considered in section 4.2.4.

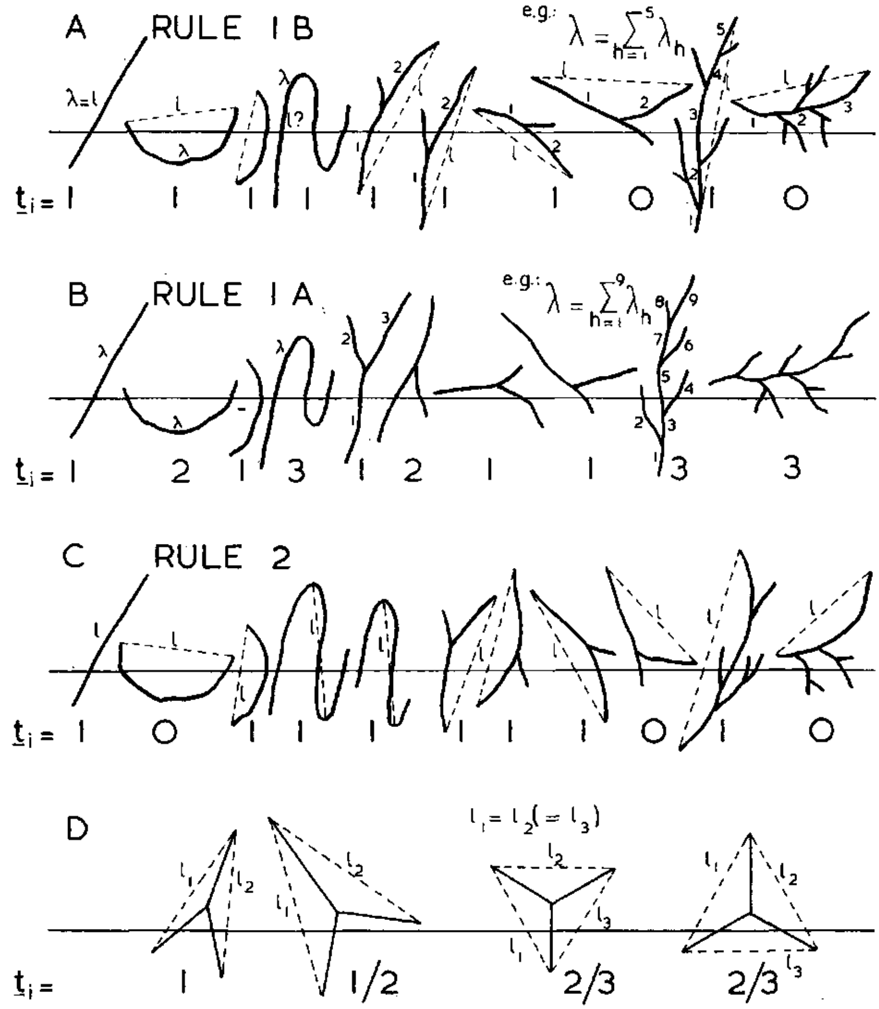


Fig. 3. Values of $\underline{t}_i$ for intersecting elements of various shapes under the rules 1a, 1b and 2.

4.2.2. Line intersect sampling under rule 1 a (Fig. 3 B)

Under rule 1a we always have an unbiased estimator $\underline{e}_{1a}(X)$ (8) for total per unit area from a line sample, i.e. even if the shape of the elements is unspecified. A variance expression however can be derived only if the type of curve is known, as we have shown for circular arcs. As may be expected, in the latter case the variance under rule 1a is equal to or slightly larger than the variance under rule 1b, but may be considerably smaller than that under rule 2 (table 1).

For arbitrarily shaped elements we cannot use (19d) as an approximation of the variance, as a measure l in $q = l/\lambda$ cannot be specified. Maybe under conditions as in the preceding section, bias will be small if (19d) is used with l

as the length of the line connecting the curve's ends.

For populations of arbitrarily shaped elements, an unbiased estimate var $\underline{e}_{1a}(X)$ can only be obtained from more than one sampling line (see section 4.2.4).

Sampling under rule 1a is in general more representative than sampling under rule 1b or rule 2, as more intersecting elements are considered. As, moreover, all intersections per element are counted, this method may be psychologically more acceptable. Measurement of total length λ (i.e. inclusive of branches, if any) involves more work however.

Summarizing: For populations of arbitrarily shaped elements and sampling under rule 1a, we have an unbiased estimator (8) for total per unit area, but no variance estimator. The variance can only be estimated if more than one sampling line is used.

4.2.3. *Line intersect sampling under rule 2 (Fig. 3 C, D)*

Contrary to the preceding cases, in line intersect sampling under rule 2 both an unbiased estimator $\underline{e}_2(X)$ (10a) for total per unit area, and an approximately unbiased one for its variance (10d) always are available. From the results obtained for circular arcs (21b, c) we may expect that in general precision will be smaller than in sampling under rules 1a or 1b (but the latter rules generally do not provide an unbiased variance estimate based on one line!).

It is stressed that the straight needle associated with an element, should be uniquely defined beforehand. In comparison with $\underline{e}_{1a}(X)$ the estimator under rule 2 will be less representative, as only elements with intersecting associated needle are included in the sample. If e.g. the needle is defined as the largest straight distance between two points on an element, intersecting branched elements of which the associated needle does not intersect as well (Fig. 3C) are disregarded. Of course the same holds for unbranched elements. This may constitute a psychological drawback in practical application (e.g. in forestry), especially to unskilled labour, so that due instruction should be given.

Apart from the measurement of other lengths that may be necessary to quantify the characteristic x, only one length measurement is involved, viz. that of the needle l.

For the academic case that there are two or more symmetric branches, the needle cannot be specified uniquely (Fig. 3D). However, for instance in case of 3 symmetric branches, we may imagine that we have defined one of the equal l_1, l_2, l_3 at random as 'the' needle, e.g. l_1. As a case of intersection always implies two intersecting needles, the probability is 2/3 that one of these two is l_1. So we may consider two out of three intersecting elements as intersecting validly, which comes to the same as counting each intersecting element with $\underline{t}_i = 2/3$ instead of 1.

Summarizing: For populations of arbitrarily shaped elements and sampling under rule 2, we have an unbiased estimator (10a) for total per unit area, and an approximately unbiased estimate (10d) for its variance. To improve precision, more than one sampling line can be used (section 4.2.4).

4.2.4. Line intersect estimators used in forestry. Sampling with more than one line

If all elements are considered as straight needles, so that $\lambda = l$, $q = 1$ are assumed (but λ is measured), and if, regardless of the number of intersections per element, $\underline{t}_i$ is put equal to one in case of intersection, all $\underline{e}(X)$ considered reduce to $\underline{e}(X)$ in (6), which then will be biased, unless the assumption is not far from reality. This $\underline{e}(X)$ is indirectly employed by BAILEY (Canada, 1970) in sampling for volume of logging residue in forestry. MATERN (Sweden, 1964), quoted by LOETSCH, ZÖHRER and HALLER (1973) employs an unbiased estimator for total length of roads and waterways per unit area, that can be derived directly from $\underline{e}_{1a}(X)$ (8). VAN WAGNER (Canada (1968) and personal communication (1973)) puts $\underline{t}_i = 0, 1, \ldots, k$ dependent on the number of intersections per element of logging residue, and employs an unbiased estimator for volume per unit area of a type that comes close to (8). In a concise publication on the estimation of total length of hedges from aerial photographs CHEVROU (France, 1973) uses estimators that can be derived from (8) if some extra assumptions are made; derivations or references to literature are not supplied however. As far as we know, the estimator $\underline{e}_2(X)$ (10) has not yet been employed.

Till now the variance of $\underline{e}(X)$ is derived from estimations made by using more than one sampling line. It was shown by DE VRIES (1973) that under the assumption of all elements being straight needles, a variance can be estimated from a one-line sample; this variance of course may be severely biased if the assumption considerably violates reality.

In this paper we introduced sampling under rule 2 which, for elements of arbitrary shapes provides (10) as an approximately unbiased estimate of the variance in one-line sampling. Further we derived this type of variance estimate (16, 19) for populations of circular arcs, including line segments and full circles as special cases.

In order to improve the precision of $\underline{e}(X)$ in the latter situations on the one hand, and on the other to find a variance estimate at all in sampling over populations of arbitrarily shaped elements under rule 1a, more than one random sampling line can be used. If k such lines of equal lengths are employed we have k estimates $\underline{e}(X)$, from which a mean and an estimate of Var $\underline{e}(X)$ can be computed directly; the variance of the mean then is estimated as $1/k$ times the latter quantity. As all variances in one-line sampling so far derived, are inversely proportional to sampling line length L, it seems justified, at least in homogeneously distributed populations, to weight each individual $\underline{e}(X)$ with its corresponding sampling line length if the latter is not the same in all samples. If an unbiased variance estimator exists in one-line sampling, as under rule 2, one might also weight each of the k estimates $\underline{e}(X)$ with the inverse of its estimated variance. In homogeneously distributed populations this will come to about the same as weighting with sampling line length.

Finally, it may also be possible to reduce the variance in a one-line sample from a heterogeneous population by post-stratifying the latter (from field data) in two strata, viz. one containing line segments, arcs and moderately curved other shapes, and the other containing the rest. Then for the first stratum (16a, d) could be used, and for the second (10a, d). Estimates over the entire population then can be obtained from:

$$\underset{\sim}{e}(X) = \underset{\sim}{e}_{1\,bo}(X) + \underset{\sim}{e}_2(X)$$

$$\mathrm{var}\ \underset{\sim}{e}(X) = \mathrm{var}\ \underset{\sim}{e}_{1\,bo}(X) + \mathrm{var}\ \underset{\sim}{e}_2(X)$$

SUMMARY

This paper is a contribution to the theory of line intersect sampling, a relatively new sampling method which finds increasing application in forest inventory.

From a general unbiased estimator for quantity per unit area of a characteristic observed on elements in a sample from a population of arbitrarily shaped elements, distributed randomly over an area, a general expression for the variance of this estimator is derived. The latter two general expressions may, dependent on the interpretation of the event of intersection, yield three different types of estimators in line intersect sampling. The three types of estimators come to the same if the elements reduce to line segments, in which case the sampling method is based directly on BUFFON's needle problem.

As, in line intersect sampling over populations of arbitrarily shaped elements, under one interpretation of intersection an expression for the variance is lacking, and under another interpretation both the expressions for variance and total-estimator do not exist, the precision of the three types of estimators generally cannot be compared. Only if the elements are segments of specified plane curves, this comparison is theoretically possible, though in many cases the derivation of the expressions for estimator for total per unit area, and its variance, will be complicated and tedious. The authors construct the latter expressions for the relatively simple case of the elements being circular arcs (including full circles and line segments as special cases), then compare precisions, and consider biases in case circular arcs are identified with line segments.

A scheme of possible estimators in line intersect sampling is added, showing mutual relationships. The line sampling designs corresponding to the different interpretations of intersection are discussed, and attention is paid to the handling of data obtained in sampling with more than one line. Finally, line intersect estimators used in forestry at present, are linked to the theory developed.

SAMENVATTING

Dit artikel is een bijdrage tot de theorie van 'line intersect sampling', een betrekkelijk nieuwe steekproeftechniek die toenemende belangstelling bij de bosinventarisatie geniet.

Uitgaande van een algemene zuivere schatter voor populatie totaal (per oppervlakte eenheid) van een karakteristiek waargenomen aan elementen in een steekproef uit een populatie van willekeurig gevormde elementen, die aselect over een oppervlak verdeeld zijn, wordt een algemene uitdrukking voor de variantie van deze schatter afgeleid. Uit deze beide algemene uitdrukkingen kunnen, afhankelijk van hoe een snijdingsgeval wordt geïnterpreteerd, drie verschillende stellen schatters (voor totaal per oppervlakte eenheid en zijn variantie) voor 'line intersect sampling' volgen. Deze drie stellen worden identiek, indien de elementen ontaarden in lijnsegmenten; in dat geval is de steekproeftechniek direct gebaseerd op BUFFON's naaldprobleem.

Indien uit een populatie van willekeurig gevormde elementen een steekproef met een lijn wordt getrokken, ontbreekt onder een der interpretaties van snijding een uitdrukking voor de variantie, terwijl onder een tweede interpretatie uitdrukkingen voor zowel totaal-schatter als variantie ontbreken. Bijgevolg kan men de nauwkeurigheid der drie typen van schatters in het algemeen niet vergelijken. Een dergelijke vergelijking is theoretisch slechts dán mogelijk, indien de elementen delen zijn van gespecificeerde vlakke krommen, doch veelal zullen de afleidingen van expressies voor totaal-schatter en zijn variantie gecompliceerd zijn. Laatstgenoemde uitdrukkingen worden afgeleid voor het betrekkelijk eenvoudige geval dat de elementen cirkelbogen zijn; de schatters voor cirkels en lijnsegmenten volgen daaruit dan als speciale gevallen. Voor het geval van cirkelbogen wordt de nauwkeurigheid der schatters vergeleken; tevens wordt de systematische fout nagegaan ingeval cirkelbogen beschouwd worden als lijnsegmenten.

Een schema van mogelijke schatters is opgenomen; daarin is ook hun onderlinge verwantschap aangegeven. De met de diverse interpretaties van snijding corresponderende steekproeftypen worden voorzover mogelijk op hun theoretische en practische merites beschouwd. Ook aan de verwerking van de met verscheidene lijnen verkregen resultaten wordt aandacht besteed. Tenslotte wordt de plaats van de thans in de bosbouw gebruikte schatters in het ontwikkelde systeem aangegeven.

REFERENCES

BAILEY, G. R., A simplified method of sampling logging residue. The For. Chronicle No. 46, 1970 (288–294).

CHEVROU, R., Inventaire des haies. Revue For. Franç. 25(1) 1973 (47–53).

KENDALL, M. G. and MORAN, P. A. P., Geometrical Probability. Griffin and Cy., 1963, London.

LOETSCH, F., F. ZÖHRER and K. E. HALLER, Forest Inventory. Vol. II, 1973, B.L.V., München.

MATERN, B., A method of estimating the total length of roads by means of line survey. Studia For. Suecica, No. 18, 1964 (68–70).

VRIES, P. G. DE, A general theory on line intersect sampling, with application to logging residue inventory. Med. Landbouw Hogeschool 73–11 (1973), Wageningen.

WAGNER, C. E. VAN, The line intersect method in forest fuel sampling. For. Sci. 14(1)1968 (20–26).

WARREN, C. E. and OLSEN, P. F., A line intersect technique for assessing logging waste. For. Sci. 10(3) 1964 (267–276).