

# What is a diameter distribution?

Oscar García  
Forest Research Institute  
Rotorua, New Zealand\*

## Summary

Spatial correlation of tree diameters causes diameter distributions to vary with the extent of land considered. In particular, distributions derived from plot data may not be representative of the distribution for a whole stand or compartment. A review of the literature on spatial correlations in forests suggests that positive spatial correlations, attributable to slow changes in microsite (soil/topography) with distance, are common. Often the effect of microsite is overridden by competition, causing negative correlations at shorter distances. Predominantly positive correlations between the trees in a plot would cause the variance within a plot to be smaller than the variance for the stand, while negative correlations would have the opposite effect. It is shown that the stand variance can be expressed as a sum of two terms, one involving the within-plots variance, and the other the variance of the plot means. Numerous studies of the relationships between plot size and variance of plot means in forest inventories, have found that the variance decreases more slowly with size than what would be the case in a random sample. This indicates positive correlations over distances comparable to the size of the plots tested. Variance partitioning is used to derive a method for estimating stand variances from inventory data. Using this method in three radiata pine plantations, it was found that the plots diameter variance underestimated the variance of the stand distribution by 3, 15, and 20%. Mapped data from an unthinned radiata pine stand, remeasured over a number of years, is used to illustrate short-range spatial correlations. Correlations are initially positive, indicating a microsite effect, but change to negative as competition intensifies. The implications of the area-dependency of tree size distributions for forest information systems and for growth model development are discussed. Although the practical impact on forest management decisions may not be large, at least the highly detailed data produced by distribution-based systems should be viewed with caution. Ignoring microsite correlations in individual-tree models can produce unrealistic and misleading results.

## Introduction

In the spirit of the title of this Symposium, it seems appropriate to discuss an issue that lies at the interface between growth modelling, forest inventory, and

---

\*Presented at the IUFRO Symposium on Integrated Forest Management Information Systems, Tsukuba, Japan, October 13–18, 1991

forest management systems. Stand tables or diameter distributions have always played an important role in forestry. With the increased availability of computer power and sophisticated statistical techniques, software packages based on distributions are being used to integrate data acquisition, growth forecasting, product utilization planning, and management information. It is claimed that individual-tree growth models produce satisfactory predictions of tree diameter distributions, and that these, and/or distributions from stand inventories or stand-level models, can be manipulated to produce detailed information on product mixes (Hyink and Moser 1983, Depta 1984, Borders 1990, for example).

It is interesting that the concept of diameter distribution is taken as self-evident, with no need for definitions. It is commonly said that a stand diameter distribution is “characterized by a probability density function”, or determined by a list of tree diameters. Forest stands or sample plots, however, are not random collections of trees. The diameters are associated to the locations of the trees on the ground, and are affected by the spatial interactions of the trees and the environment. It follows that a satisfactory definition of diameter distribution is not obvious, and must be related to the extent of the piece of land considered. García (1984, 1988) warned that, at least in theory, the characteristics of plot distributions, on which models are based, can be different from those for whole stands or compartments, which are the ones usually required in the applications.

In this paper I examine further the likely differences between plot and stand distributions, and try to determine their significance. The next section considers possible interpretations of the concept of diameter distribution, and some of their properties. Then, the literature on spatial correlation in forest stands, the feature responsible for the changes in distribution with area, is reviewed. This is followed by an analysis of the connection between the differences in plot and stand variances, and the well-known relationships linking plot size and sampling error in forest inventory and cluster sampling. A method for estimating plot and stand variances from inventory data is presented next, and results from three stands are shown. Some examples of the evolution of spatial relationships in a radiata pine stand are given. Finally, the potential consequences of spatial pattern, and implications for growth modelling and forest management are discussed.

## Stand and plot distributions

A forest stand consists of a set of trees, each characterised by its location and size. Obviously, once the trees have been felled and extracted the locations are irrelevant for utilization purposes. A size distribution for the stand provides the necessary information about sizes, ignoring the tree locations. For simplicity, I will talk about diameter distributions, but most of the discussion is applicable to other tree variables, possibly vector-valued.

A diameter distribution, either for a whole stand or for a sample of trees, may be specified by listing the diameter measurements, a *tree list*. If instead of repeating values we assign to each unique measurement value the number of

times it appears in the list, we have the equivalent of a discrete frequency function (Figure 1). The frequencies may be given as absolute numbers, as numbers per hectare, or as proportions. This is the classical *stand table*, also known as *diameter-distribution series* or *assortment by diameters* (Spurr 1952). Usually the diameters are measured in a coarse scale, rounded, or grouped into diameter classes. The grouping into classes may be seen as a histogram approximation or smoothing of the distribution (Silverman 1986). Other commonly-used approximate/smoothed representations are probability density functions and lists of quantiles or percentiles. We shall ignore the issue of measurement errors. The measurement precision just affects the likely number of trees having the same diameter.

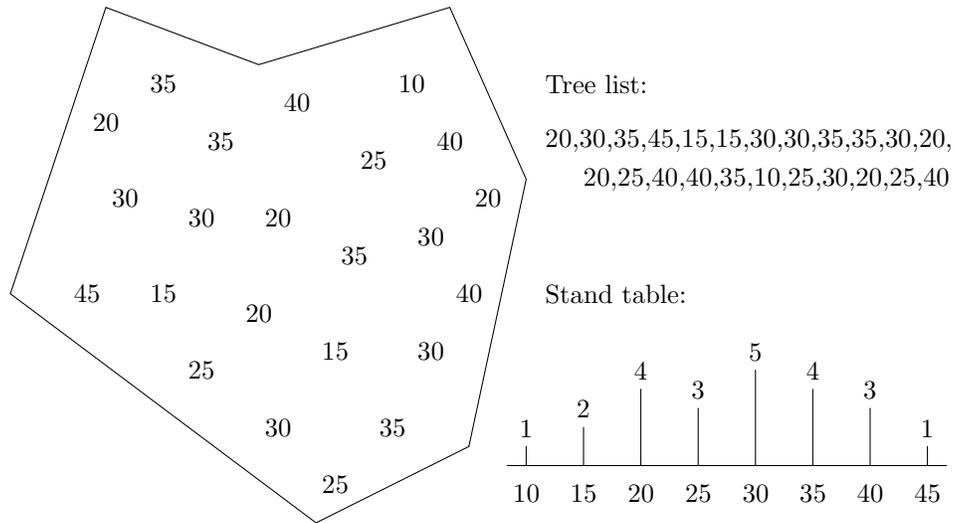


Figure 1: Tree lists and stand tables.

Unless stated otherwise, I shall consider the stand diameter distribution (and tree locations) as a fixed population characteristic, without an obvious probabilistic interpretation (although the stand table, as proportions, is also the probability density for the diameter of a tree selected at random). An alternative is to think of the actual locations and diameters as a realization of a marked point stochastic process (Matérn 1960, Stoyan et al. 1987, Daley and Vere-Jones 1988, König et al. 1991), or of the tree list as a sample from some continuous probability distribution (Smith 1976). This “superpopulation” view presents some conceptual and philosophical difficulties, although it might be interpreted in terms of prior knowledge in a Bayesian setting.

In practice, one rarely knows the diameters of all the trees in the stand, but only those for the trees in one or more sample plots. In the literature, usually no distinction is made between stand and plot distributions. There is an implicit assumption that the plot distributions are somehow representative of the stand. In growth modelling applications, distribution functions or other approximations derived from plots are said to “characterize” the stand distribution.

In what sense do plot distributions represent or characterize a stand? Obviously, a plot tree list is not directly comparable to the stand tree list, being much shorter. A stand table built from a sample is more irregular than the stand table for the whole population. The relationships between plot and stand distributions must, therefore, be examined in a sampling context.

The trees in a random plot, or in a simple random sample of plots, are *not* a simple random sample from the trees in the stand, they have been selected through a form of cluster sampling (Cochran 1963) or, more specifically, area sampling (Sukhatme and Sukhatme 1970). Consider first just one plot located at random (uniformly) anywhere in the stand, and ignore edge effects. Then, all trees have the same probability of appearing in the plot. Pairs of trees, however, do not; pairs of trees closer together are more likely to be selected than pairs further apart (Fig 2). Clearly, this is also true for a random sample of any number of plots. In forest inventory, a stand is usually partitioned into plots of uniform size, and a sample of them is selected at random, with equal probabilities and without replacement. Again, all trees have the same selection probability, but only pairs from a same partition element can occur in a sample plot.

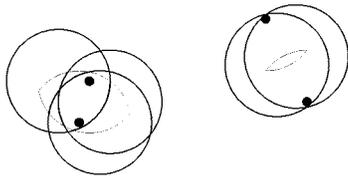


Figure 2: All circular plots with centres inside the shaded area include the pair of trees. Pair selection probabilities decrease with increasing distance, becoming zero for distances greater than the plot diameter.

Because of the equal individual selection probabilities, the expected values of the frequencies (as trees per hectare) in a plot stand table are equal to the frequencies for the whole stand. This applies also to any linear combination of the frequencies (expectation is a linear operator), so that the value calculated on a plot is an unbiased estimate of the population value. These linear combinations include the per hectare totals of functions of diameter (e.g. basal area per hectare), as well as amounts between fixed diameter limits.

Many quantities of interest, however, are nonlinear functions of the stand table frequencies. Examples of these are variances, percentiles, relative diameters, and many of the parameters calculated when fitting probability density functions. Their expectations involve joint selection probabilities for two or more trees. Therefore, unless pairs of diameters and the tree locations are unrelated, expected plot values vary with plot size and, in particular, are not representative of the stand values. In other words, considering all the diameters

in a random plot as a multivariate random variable, the marginal distributions are all equal to the distribution for a random tree in the stand, but the joint distributions change with the size of the plot.

Thus, the existence of spatial diameter patterns implies that, at least in principle, scale and shape characteristics of diameter distributions at the plot level are different from those at the stand level. I will focus mainly on the variance, and explore now the question of the sign and magnitude of these differences.

It is interesting to note that there is also a different source of bias that can affect some variables. Although plot totals such as basal area per hectare are unbiased estimators of the stand values, averages such as mean diameter are not, being analogous to ratio estimators (Cochran 1963). The source of bias here is the variability in the denominator, the number of trees per plot. This effect is likely to be small in plantations, but it might be important in highly irregular natural forests.

## Spatial correlation

We saw that the variance and other properties of plot diameter distributions differ from the stand values if the relative sizes of pairs of trees tend to vary with the distance between them. There are two main sources for this kind of spatial correlation: competition, and microsite variability.

Spatial correlation is modelled and measured in a number of ways. Usually, it is postulated that the observations are a realization of a stochastic process with certain properties (stationarity, isotropy, ergodicity). Then, a spatial covariance or correlation function is defined as the expected value of the product of the values (diameters) at two points, possibly centred and standardized using the mean and variance (Daley and Vere-Jones 1988, Matérn 1960, Ripley 1981, Stoyan et al. 1987). The terms *autocovariance* and *autocorrelation* are also used. A closely related alternative popularized by the “geostatistics” school (Matheron 1971) is the variogram, based on the variance of the differences between points. Incidentally, Matérn (1960, §4.1) mentions that the variogram was used by Langsaeter in a study of systematic sampling for forest inventory published in 1926. Most of the work has dealt with continuous functions of the spatial coordinates, or with values at all the points of a rectangular grid (e. g., a regularly spaced plantation with no mortality). These concepts have been extended to randomly distributed points through a *mark-covariance function* (Stoyan et al. 1987, König et al. 1991).

In these approaches, the covariance/correlation functions and the variogram are defined in terms of expectations of the stochastic process. They are averages over “all the possible stands” for a fixed pair of points on the ground. In the applications these are estimated by spatial averages, where a pair of points with a given relative displacement move over the fixed observed realization. The calculated functions (functions of the displacement, or just of distance if isotropy is assumed) are often called *correlograms*, or empirical covariances, correlations and variograms. A number of issues are involved in producing

good estimates of the “real” functions, including necessary properties of the stochastic process, smoothing, and corrections for edge effects.

It would be possible also to define finite population and sample correlation functions as spatial averages over the actual tree diameters, in the spirit of the *intracluster correlations* of survey sampling (Cochran 1963, Sukhatme and Sukhatme 1970). I will avoid here a precise definition, using mainly the intuitive meaning of spatial correlation as a measure of the similarity of “typical” pairs of trees at any given distance (which may vary with direction if there is anisotropy).

The likely effects of competition and microsite on the correlogram (or on the variogram, which is essentially a mirror-image of it) are discussed by Matérn (1960), Bachacou and Decourt (1976), and Bouchon (1979) (see also García 1984, 1988). Other things being equal, at distances close to the average tree spacing one could expect a negative correlation due to the competition between neighbours (large trees would tend to have small neighbours; note that this is a basic assumption in distance-dependent growth models). At about twice the spacing the correlation would be positive (we have then competitors of the competitors), and so on, causing the correlogram to oscillate around zero. The oscillations would dampen quickly, due to variation in neighbour distances, genetic and other non-spatial causes of diameter variation, and to “three-at-a-time” effects (consider the competition among three trees located on the vertices of a triangle). Superimposed on this is the site-induced correlation. Many soil properties vary slowly over a stand, so that the fertility at two nearby points tends to be more similar than at points far apart. The result is a positive correlation component, monotonically decreasing with distance.

Matérn (1960, §4.5) states that, with a few exceptions, there is little evidence of negative competition-induced correlations in published observations on the variation of the yield of agricultural fields and forests. He attributes this to masking by strong correlations in soil properties, to the lack of observations at very short distances, or to the fact that most observations refer to plots of finite area and not to points. The existence of positive correlations decreasing with distance appears to be well established, though.

Millier, Poissonnet and Serra (1972) calculated variograms for number of trees and basal area in two stands. The reported values are totals for plots of  $2 \times 2$  m, located at 1 m intervals. In terms of the correlogram, they found relatively small correlations, declining at a roughly linear rate over the whole range of distances. In one of the stands the spatial pattern disappeared after a thinning. There is no evidence of competition effects, although this could be expected with plot data, as suggested by Matérn. In the discussion, however, Decourt mentioned that Poissonnet had found a sinusoidal variogram in a 30 year-old spruce stand.

A more extensive study was carried out by Bachacou and Decourt (1976). They computed diameter variograms along the rows in stands planted on regular square grids. The few missing or dead trees were assigned a diameter equal to the stand minimum. A large number of stands of various species, ages and spacings were analyzed. In general, the results are in good agreement with the theoretical expectations. The short-range oscillatory behaviour of the

variogram/correlogram associated with competition was often clearly visible, although varying with species, age and spacing. At longer distances there is frequently a positive, decreasing correlation, unrelated to spacing, as would be expected from microsite effects. They noted the very different behaviours of clonal and non-clonal plantations. The clone variograms showed little or no evidence of competition-induced fluctuations, but large and sustained long-range trends, while the opposite was true for populations of different genotypes. This could be attributed to a lower size differentiation from competition in genetically homogenous populations and a higher site sensitivity of the clones, especially poplars, as well as to obvious environmental gradients caused by drainage in some of the clonal stands.

A different approach was taken by Reed and Burkhart (1985) to study short-range correlations. Instead of expressing spatial correlation as a function of distance, they use an index based on the correlations among neighbouring trees. Neighbours were defined as those whose cells in the Dirichlet tessellation (Brown's APA (Brown 1965) or Voronoi polygons) shared common boundaries. It was found that the neighbour basal area correlations tended to be positive for low levels of competition, negative at intermediate levels, and positive again at high levels of competition (perhaps because of death of the initial competitors?). Schoonderwoerd and Mohren (1988) used the same index, but defining as neighbours of a tree all those trees within a fixed distance from it (distances from 2 to 10 m were tried). They found mostly negative diameter correlations at some neighbourhood distance after canopy closure, although varying with age and soil heterogeneity. Thinning appeared to temporarily destroy the correlations.

We may conclude that there is often a positive diameter correlation, decreasing with distance, associated with microsite heterogeneity. At shorter ranges the correlation may be reduced or become negative due to competition (see also the radiata pine examples, below).

It is not difficult to understand the effect of spatial correlation on plot variance. A predominantly negative correlation among the trees in a plot would cause the variance within plots to be higher than that for the whole stand, since it indicates that trees included in a plot tend to be more dissimilar than trees located anywhere in the stand. The opposite is true with a positive correlation. In fact, the difference between the stand and plot variances is essentially a weighted average of the covariance over all pairs of points, with weights equal to their inclusion probabilities (see next section and Whittle (1956) or Matérn (1960, §2.5)). Therefore, for plots of intermediate sizes where positive microsite-induced correlations dominate we can expect the plot variance to be lower than the stand variance, and to decrease with increasing plot size. Strong competition might reduce or reverse this effect in small plots.

## Plot size in forest surveys

From the properties of conditional expectation, it is known that

$$V(Y) = E_T[V(Y|T)] + V_T[E(Y|T)],$$

where the subscript  $T$  indicates expectation or variance over a conditioning random variable  $T$  (Rao 1973). If we take  $Y$  as the size of a tree selected at random from a stand, and  $T$  as indicator of the event of the tree belonging to a particular sample plot, we can then write the stand variance as

$$\sigma^2 = \overline{\sigma_w^2} + V(\bar{y}). \quad (1)$$

Here  $\overline{\sigma_w^2}$  is the average within-plot variance, and  $V(\bar{y})$  is the variance of the plot means. For simplicity, in this section I shall assume that the stand is large relative to the plot size, and ignore the variation in the number  $n$  of trees per plot. An alternative formulation, in a finite-population setting, is presented in the next section.

If the trees in the plot were a random sample from the trees in the stand,

$$V(\bar{y}) = \frac{\sigma^2}{n}.$$

The usual sample variance

$$s_w^2 = \frac{\sum(y - \bar{y})^2}{n - 1} = \frac{n}{n - 1} \sigma_w^2 \quad (2)$$

would then be an unbiased estimator of the stand variance:

$$\overline{s_w^2} = \frac{n}{n - 1} [\sigma^2 - V(\bar{y})] = \sigma^2.$$

In general, however, this is not the case.

The relationship between the variance of the plot means and the plot size has been extensively studied in Forest Inventory. It is generally found that the decrease of  $V(\bar{y})$  with  $n$  (or with the plot area  $A$ ) is slower than that for a random sample. Over the range of plot sizes used in practice, a formula proposed by Fairfield Smith in 1938, based mainly in agricultural uniformity trials, is a good approximation (Kinashi 1954):

$$V(\bar{y}) = \frac{k}{A^\beta}.$$

The parameter  $\beta$  is usually less than 1, the value in the absence of spatial correlation. This indicates a positive intracluster (within-plots) correlation (Sukhatme and Sukhatme 1970). Whittle (1956, 1962) and Matérn (1960, §6.12) relate Smith's formula to the correlation function of a spatial stochastic process.

In Table 1 I have collected  $\beta$  values for volumes or basal areas from a number of forest sampling studies. The  $\beta$ 's were given by the authors, or have been calculated from regressions of  $\log V(\bar{y})$  over  $\log A$  with the published data (actually, mostly coefficients of variation or variances for means expressed on a per hectare basis are given). CV is the estimated coefficient of variation for 0.1ha plots. In almost all instances the log-log relationships are reasonably linear, although often the number of points is too small to be sure. Hatheway and Williams (1958), however, in re-analyzing the data from Johnson and Hixon

Table 1: Estimates of  $\beta$ 

Source	Variable	Sizes (ha)	CV	$\beta$
Johnson and Hixon (1952), square plots	bd.feet	0.04–0.16	56	0.53
Kinashi (1954) (Strand 1957), plantations	volume	-	-	0.50
Kinashi (1954) (Strand 1957), natural	volume	-	-	0.92
Mesavage and Grosenbaugh (1956), cp.15	bd.feet	0.04–0.32	130	0.43
Mesavage and Grosenbaugh (1956), cp.14	bd.feet	0.04–0.32	140	0.85
Mesavage and Grosenbaugh (1956), cp.28	bd.feet	0.04–0.32	95	0.52
Mesavage and Grosenbaugh (1956), cp.16	bd.feet	0.04–0.32	95	0.52
Mesavage and Grosenbaugh (1956), cp.27	bd.feet	0.04–0.32	85	0.42
Mesavage and Grosenbaugh (1956), cp.26	bd.feet	0.04–0.32	95	0.29
Mesavage and Grosenbaugh (1956), cp.23	bd.feet	0.04–0.32	85	0.29
Mesavage and Grosenbaugh (1956), La.	bd.feet	0.04–0.32	52	0.70
Strand (1957)	b.area	-	-	0.50
Prodan (1958) stand 122	b.area	0.01–0.25	12	0.81
Prodan (1958) stand 123	b.area	0.01–0.25	16	0.73
Prodan (1958) stand 124a	b.area	0.01–0.25	27	0.38
Prodan (1958) stand 124b	b.area	0.01–0.25	9	0.95
Prodan (1958) stand 125	b.area	0.01–0.25	28	0.48
Prodan (1958) stand 126.1	b.area	0.01–0.25	14	0.76
Prodan (1958) stand 126.2	b.area	0.01–0.25	7	0.99
Prodan (1958) stand 126.3	b.area	0.01–0.25	18	0.48
Prodan (1958) stand 126.4	b.area	0.01–0.25	19	0.84
Prodan (1958) stand 127.1	b.area	0.01–0.25	18	1.07
Prodan (1958) stand 127.2	b.area	0.01–0.25	19	0.96
O'Regan and Palley (1965), LS	b.area	0.02–0.16	42	0.74
O'Regan and Palley (1965), WP	b.area	0.04–0.10	19	0.98
O'Regan and Palley (1965), BM	b.area	0.02–0.32	49	0.98
Tardif (1965)	volume	0.02–0.12	18	0.54
Wensel and John (1969), Cutfoot	b.area	-	-	0.82
Wensel and John (1969), Portage Lk	b.area	-	-	0.60
Wensel and John (1969), Cloquet	b.area	-	-	1.08
Grayet (1977), regular	b.area	0.04–0.25	16	1.07
Grayet (1977), irregular	b.area	0.04–0.25	24	1.06

(1952), found a deviation indicating competitive effects for the smallest (1/40 acre) plots. A similar situation occurs with the data from O'Regan and Palley (1965), where two different lines had to be fitted.

Assume that Fairfield Smith's *law* holds, at least for plots that are not too small. That is,

$$V(\bar{y}) = \frac{\alpha\sigma^2}{n^\beta},$$

where  $\alpha$  is the ratio of the predicted stand variance (from extrapolation to  $n = 1$ ) to the actual stand variance  $\sigma^2$ . Then, the percentage "bias" of  $s^2$  would be

$$100 \frac{\overline{s_w^2} - \sigma^2}{\sigma^2} = 100 \frac{1 - \alpha n^{1-\beta}}{n - 1}.$$

Table 2 shows some values for this expression, for  $\beta = 0.5$ . Obviously, without further information we cannot say much about the magnitude of the discrepancies. It is clear, however, that there is a potential for significant differences in variance with plot size.

Table 2: % difference between plot and stand variance ( $\beta = 0.5$ )

$\alpha$	trees per plot				
	10	25	50	100	200
0.10	8	2	1	0	0
0.50	-6	-6	-5	-4	-3
1.00	-24	-17	-12	-9	-7
2.00	-59	-37	-27	-19	-14

## Estimating plot and stand variances from inventory data

Given inventory data containing a simple random sample of  $m$  plots from a stand, we can easily estimate the tree size variance for the plots by its sample mean. It is possible to obtain an estimate of the stand variance with a method used in cluster sampling (Cochran 1963 §9.3, Sukhatme and Sukhatme 1970 §6.4). The approach was developed for equal cluster sizes, and has been used by Prodan (1958) to investigate the effect of plot size on the precision of forest inventories, taking as elementary population units the basal areas for 0.01ha plots. We deal here with the individual trees and the associated variable-size clusters (fixed area plots containing variable numbers of trees).

Consider a stand partitioned into  $M$  potential sample plots, with plot  $p$  containing  $n_p$  trees. Let  $y_{pi}$  be the size of tree  $i$  in plot  $p$ . The sum of squares

can be partitioned according to the following analysis of variance:

	D.F.	S.S.	M.S.
Within plots	$M(\bar{n} - 1)$	$\sum_p^M \sum_i^{n_p} (y_{pi} - \bar{y}_p)^2$	$\overline{S_w^2}$
Between plots	$M - 1$	$\sum_p^M n_p (\bar{y}_p - \bar{y})^2$	$\bar{n} S_b^2$
Total for stand	$M\bar{n} - 1$	$\sum_p^M \sum_i^{n_p} (y_{pi} - \bar{y})^2$	$S^2$

In the mean squares,  $S^2$  is the required stand variance,  $\overline{S_w^2}$  is an average within-plots variance, and  $S_b^2$  is a weighted variance of the plot means. Note that

$$S^2 = \frac{(\bar{n} - 1)\overline{S_w^2} + (1 - 1/M)\bar{n}S_b^2}{\bar{n} - 1/M} \quad (3)$$

(compare to (1) and (2)).

We can write down a similar table for the sample:

	D.F.	S.S.	M.S.
Within plots	$m(\hat{n} - 1)$	$\sum_p^m \sum_i^{n_p} (y_{pi} - \bar{y}_p)^2$	$\overline{s_w^2}$
Between plots	$m - 1$	$\sum_p^m n_p (\bar{y}_p - \hat{y})^2$	$\hat{n} s_b^2$
Total for stand	$m\hat{n} - 1$	$\sum_p^m \sum_i^{n_p} (y_{pi} - \hat{y})^2$	$s^2$

where  $\hat{n}$  and  $\hat{y}$  are sample averages. The sample stand mean square  $s^2$  is not a good estimate for  $S^2$ , since the elements on which it is based are not a simple random sample from the population. It is found, however, that with equal cluster sizes the sample “between” and “within” mean squares are unbiased estimates of the corresponding population values. Therefore, an estimate for  $S^2$ , unbiased in the equal clusters case, is obtained by substituting the sample mean squares for the population ones in the right-hand-side of (3). Substituting also  $\hat{n}$  for  $\bar{n}$  and simplifying,

$$\hat{S}^2 = \frac{\frac{1}{m\hat{n}} \sum \sum y_{pi}^2 + \frac{1-m/M}{m\hat{n}(m-1)} \sum n_p \bar{y}_p^2 - \frac{1-1/M}{m\hat{n}^2(m-1)} (\sum n_p \bar{y}_p)^2}{1 - \frac{1}{M\hat{n}}}. \quad (4)$$

The denominator is essentially 1, and in most practical situations the other terms containing  $M$  can also be neglected (these terms are zero in sampling with replacement).

In general, this estimator is not unbiased, due to fluctuations in  $n_p$ . The bias is likely to be small, at least for stands where the number of trees per plot does not vary too much. More robust variants of (4) might be useful for natural stands with very irregular stocking.

Table 3 shows results from using this method in three inventories of radiata pine plantations.

Table 3: Estimated plot and stand variances

Plot size (ha)	0.05	0.06	0.10
Stand area (ha)	10	84	-
Plots	27	25	10
Stems/ha	441	463	240
<i>Diameters (cm):</i>			
Mean	46.8	40.6	33.5
Plot variance	312.4	125.3	17.5
Stand variance	323.1	147.2	21.8
% difference	-3.3	-14.9	-20.0
<i>Basal areas (m<sup>2</sup>):</i>			
Mean	0.1974	0.1412	0.0898
Plot variance	0.01946	0.00472	0.000491
Stand variance	0.01998	0.00530	0.000608
% difference	-2.6	-10.9	-19.3

## Spatial pattern in a radiata pine stand

Some of the spatial relationships discussed above can be illustrated with data from a radiata pine plantation. The experimental site is described in detail by Beets and Brownlie (1987). We use two 0.1 ha plots in an unthinned stand, where all trees have been mapped and their diameters measured annually since planting in 1974. The mean diameter and stocking are shown in Figure 3.

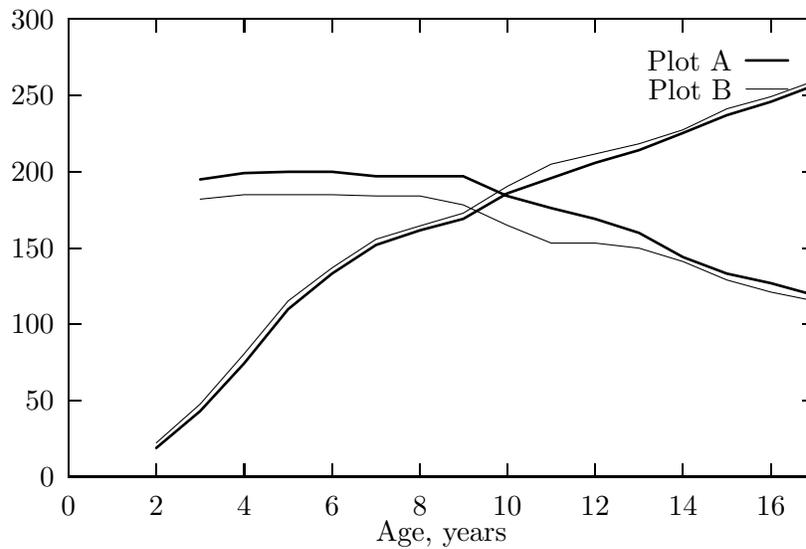


Figure 3: Mean diameter (mm), and number of trees per plot.

Figure 4 displays the relative diameters and diameter correlations at ages 4, 8, 12, and 16 years. The correlation functions were calculated by grouping the distances between all pairs of trees into 1m classes. Similar values were

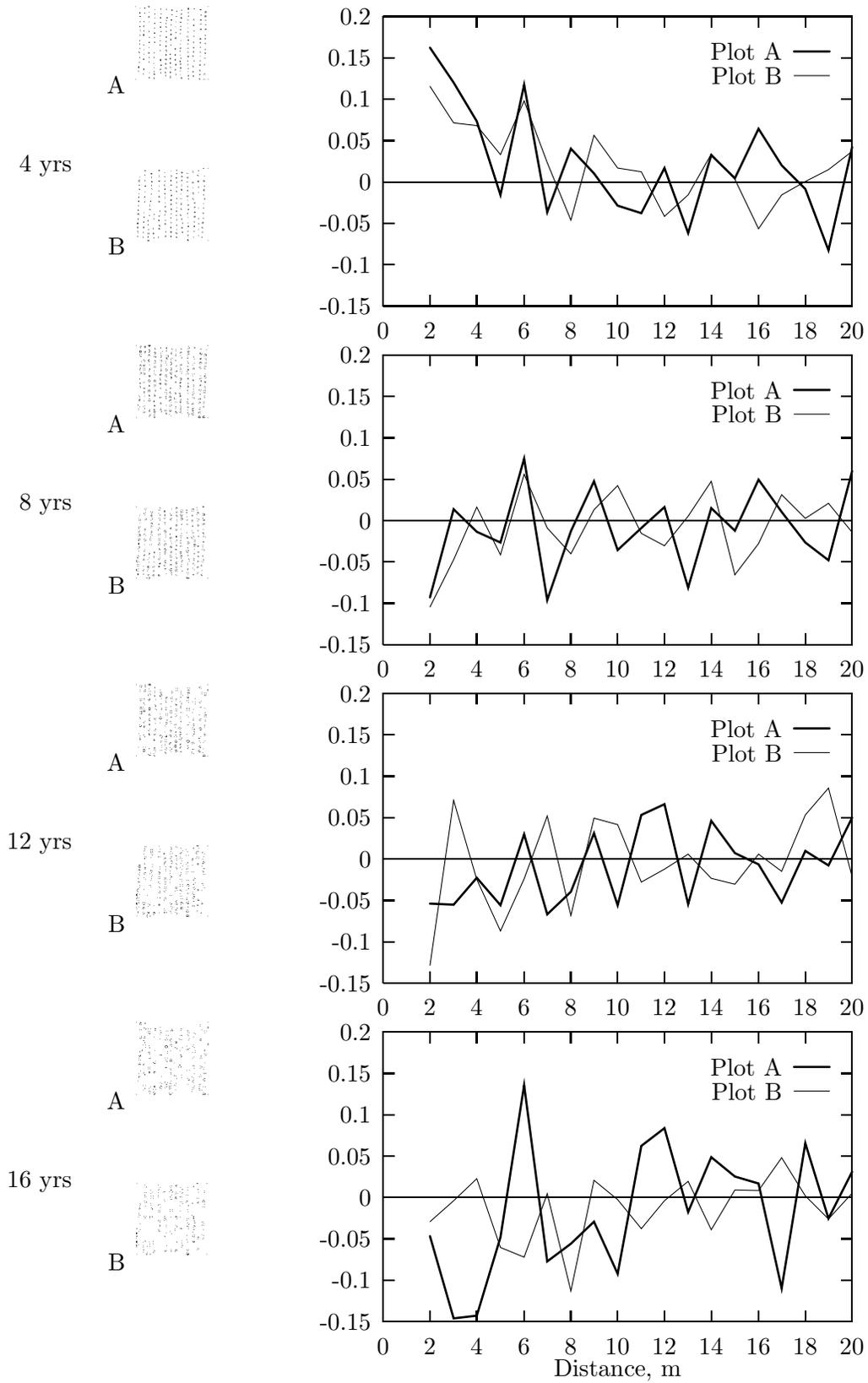


Figure 4: Tree positions and relative diameter(scaled  $\times 5$ ), and correlation functions for ages 4, 8, 12, and 16 years

obtained for tree basal area correlations.

It is clear that there is initially a positive correlation between the sizes of nearby trees, presumably due to microsite similarities. Later, competition overcomes this effect, and the correlation becomes negative. Any competition-induced oscillations in the correlation function, as described by Matérn (1960) and Bachacou and Decourt (1976), would be obscured here by the irregular and nominally rectangular ( $1.8 \times 2.4\text{m}$ ) spacing. Figure 5 demonstrates the changes with age in the correlation for the 2m distance.

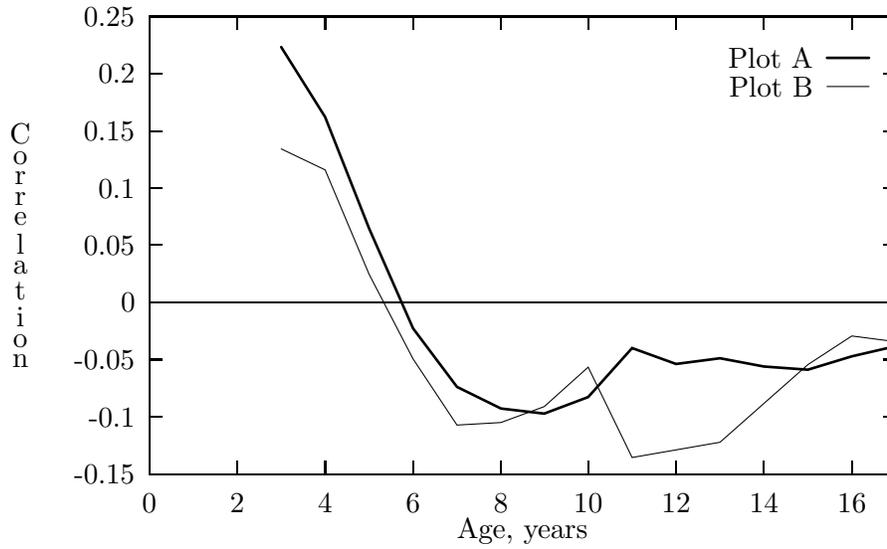


Figure 5: Change in 2m correlations with age

It is important to understand that, as with the variances, a correlation function derived from plot data will tend to be somewhat different from the one applicable to the whole stand. It seems plausible that stand correlations for the longer distances would be higher than those calculated here, due to positive soil/topography correlations at ranges above the plot size.

The trees in this unthinned stand are subject to intense competition, as can be seen from the heavy mortality in Figure 3. Thinning reduces competition and disturbs the spatial pattern (Schoonderwoerd and Mohren 1988), and would have produced less negative, or perhaps predominantly positive correlations.

These mapped areas are somewhat small for studying the relationship between variance and plot size, and no estimate of the stand variance is available. For illustration, though, values for small subplots, relative to the plot variance, are shown in Table 4. These were calculated by subdividing the plots into 4, 9, and 16 subplots, and taking the average of the variances calculated with (2).

## Discussion and conclusions

Totals per hectare, such as stocking and basal area, obtained from random sample plots are unbiased estimates of the corresponding values for a stand.

Table 4: % variance differences relative to 0.1ha plot

Age	Plot A			Plot B		
	0.0250ha	0.0111ha	0.0062ha	0.0250ha	0.0111ha	0.0062ha
4	-3.7	-10.9	-8.3	-2.1	-7.8	-8.6
8	-0.3	1.7	2.6	-0.2	1.3	4.6
12	2.2	5.0	9.5	0.0	-0.6	8.7
16	1.7	4.9	8.6	0.8	2.4	4.6

More generally, this is true for any linear functions of the size distributions or stand tables. Due to the correlation of tree sizes within plots, however, the expected values of nonlinear functions such as variances, shape parameters and percentiles must vary with the plot size, and, in particular, must differ from the stand values. In other words, a stand is more than a list of trees.

It is curious that these facts, although closely related to well-known issues in Forest Inventory, have not been widely appreciated by growth modellers. Apart of the warnings by García (1984, 1988), the only specific reference to these problems that I have found is a comment by Arvanitis in the discussion of Zöhrer (1972). Stage (1973, see p.27) may also have been aware of some of the difficulties. Fries (1967) mentioned that the coefficient of variation within 25 m<sup>2</sup> plots tended to be higher than within 100 m<sup>2</sup> plots, but he did not pursue the matter further.

The magnitude of the difference between plot and stand variances in any particular instance may or may not be important. The effect will depend, among other things, on the interactions between soil/topography homogeneity, competition intensity, age, genetics, and thinning and other treatments. Although it seems clear that substantial differences are possible, more research is needed to establish their extent under various circumstances. The method for estimating variances from inventory data presented here can be useful for this purpose.

The influence of competition on the correlations and variances may vary depending on the predominant competition mechanisms. Competition may be “diffuse” (Schoonderwoerd and Mohren 1988), where it is exercised indirectly through sharing of substrate resources, or “sharp”, arising from direct neighbour-to-neighbour interference. The diffuse competition concept is implicit in distance-independent models (in an extreme form, independent of distance), and in many distance-dependent competition indices (most clearly in Spurr (1962) and Pukkala (1988), for example). Sharp competition assumptions are most obvious in models that partition the growing space into polygons (Brown 1965, Gates et al. 1979). With sharp competition, irregularity in the tree locations might obscure the influence of competition in the correlation functions, the correlations being more related to neighbourhood relationships than to distance (Reed and Burkhart 1985). There is some evidence to suggest that underground competition for nutrients and water tends to be diffuse, while above-ground competition for light is “sharper” (Oliver and Larson 1990).

If the spread (variance) of a diameter distribution changes with the exten-

sion of the piece of land considered, the same can be expected of the distribution shape, perhaps to a greater extent. In addition, reliable estimation of shape parameters (higher moments, etc.) requires unrealistically large samples. The accuracy of the highly detailed information produced by some growth forecasting and analysis systems may be largely illusory. In this context, there may be truth in the observation that “the fitting of distributions to observational data has a certain intrinsic interest which is apt to outrun its statistical usefulness” (Kendall and Stuart 1976, p.185).

We can conclude that the applicability to whole stands of distributions derived from plot data is, at best, uncertain. It is also conceptually inconsistent to use pooled stand inventory data as input to distance-independent growth models. Strictly speaking, the models are not really *distance-independent*. In practice, growth predictions may well be insensitive to the exact diameter distribution variance and shape. If this is so, it can be argued that the use of this redundant state information presents disadvantages (García 1988).

There are implications also for distance-dependent growth models. Microsite-induced correlation may explain the poor performance of competition indices that is often reported. More importantly, ignoring this correlation can produce unrealistic and misleading spatial patterns and distributions.

## References

- Bachacou, J. and Decourt, N. (1976). Etude de la compétition dans des plantations régulières à l'aide de variogrammes. *Annales des Sciences Forestiers*, 33(4), 177–198.
- Beets, P. N. and Brownlie, R. K. (1987). Puruki experimental catchment: Site, climate, forest management, and research. *New Zealand Journal of Forestry Science*, 17(2/3), 137–160.
- Borders, B. E. (1990). Projecting stand and stock tables: A comparison of three methods. In Burkhart, H. E. (Ed.), *Research in Forest Mensuration, Growth and Yield, Proceedings of Sessions of S4.01 at the 1990 IUFRO World Congress*, pages 13–22. Publication FWS-2-90, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University.
- Bouchon, J. (1979). Structure des peuplements forestiers. *Annales des Sciences Forestiers*, 36(3), 175–209.
- Brown, G. S. (1965). Point density in stems per acre. New Zealand Forestry Research Notes 38, NZ Forest Service.
- Cochran, W. G. (1963). *Sampling Techniques* (Second ed.). Wiley.
- Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer-Verlag.
- Depta, D. J. (1984). Integrated forest planning systems at Weyerhaeuser Company. In Nagumo, H. et al. (Eds.), *Proceedings IUFRO Symposium on*

- Forest Management Planning and Managerial Economics*, pages 184–194. University of Tokyo.
- Fries, J. (1967). Variation in forest stands as a basis for the planning of experiments. In *XIV IUFRO Congress*, volume VI, pages 37–54.
- García, O. (1984). New class of growth models for even-aged stands: *Pinus radiata* in Golden Downs Forest. *New Zealand Journal of Forestry Science*, *14*(1), 65–88.
- García, O. (1988). Growth modelling — A (re)view. *New Zealand Forestry*, *33*(3), 14–17.
- Gates, D. J., O'Connor, A. J., and Westcott, M. (1979). Partitioning the union of disks in plant competition models. *Proceedings Royal Society London, A*, *367*, 59–79.
- Grayet, J.-P. (1977). Contribution à l'étude de l'échantillonnage en futaie feuillue. *Annales des Sciences Forestiers*, *34*(1), 59–75.
- Hatheway, W. H. and Williams, E. J. (1958). Efficient estimation of the relationship between plot size and the variability of crop yields. *Biometrics*, *14*, 207–222.
- Hyink, D. M. and Moser, Jr., J. W. (1983). A generalized framework for projecting forest yield and stand structure using diameter distributions. *Forest Science*, *29*(1), 85–95.
- Johnson, F. A. and Hixon, H. J. (1952). The most efficient size and shape of plot to use for cruising in old-growth Douglas-fir timber. *Journal of Forestry*, *50*(1), 17–20.
- Kendall, M. and Stuart, A. (1976). *The Advanced Theory of Statistics* (Fourth ed.), volume 1: *Distribution Theory*. Griffin.
- Kinashi, K. (1954). Forest inventory by sampling methods. Bulletin 23, Kyushi University Forest. (Cited by Strand 1957).
- König, D., Carvajal-Gonzalez, S., Downs, A. M., Vassy, J., and Rigaut, J. P. (1991). Modelling and analysis of 3-D arrangements of particles by point processes with examples of application to biological data obtained by confocal scanning light microscopy. *Journal of Microscopy*, *161*(3), 405–433.
- Matérn, B. (1960). Spatial variation. *Meddelanden från Statens Skogsforskningsinstitut*, *49*(5).
- Matheron, G. (1971). The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau 5, École Nationale Supérieure des Mines de Paris.
- Mesavage, C. and Grosenbaugh, L. R. (1956). Efficiency of several cruising designs on small tracts in North Arkansas. *Journal of Forestry*, *54*, 569–576.
- Millier, C., Poissonnet, M., and Serra, J. (1972). Morphologie mathématique et sylviculture. In *3e Colloque des Statisticiens Forestiers, IUFRO*, pages 287–307, Jouy-en-Josas, 7-11 sept. 1970. INRA.

- Oliver, C. D. and Larson, B. C. (1990). *Forest Stand Dynamics*. McGraw-Hill.
- O'Regan, W. G. and Palley, M. N. (1965). A computer technique for the study of forest sampling methods. *Forest Science*, 11(1), 99–114.
- Prodan, M. (1958). Untersuchungen über die durchführung von repräsentativeaufnahmen. *Allg.Forst- u.Jagdztg.*, 129, 15–33.
- Pukkala, T. (1988). Methods to describe the competition process in a tree stand. *Scandinavian Journal of Forest Research*, 4(2), 187–202.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (Second ed.). Wiley.
- Reed, D. D. and Burkhart, H. E. (1985). Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science*, 31(3), 575–587.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley.
- Schoonderwoerd, H. and Mohren, G. M. J. (1988). Autocorrelation and competition in even-aged stands of Douglas-fir in the Netherlands. In Ek, A. R., Shifley, S. R., and Burke, T. E. (Eds.), *Forest Growth Modelling and Prediction*, pages 619–626. USDA Forest Service, General Technical Report NC-120.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- Smith, T. M. F. (1976). The foundations of survey sampling: a review (with discussion). *Journal of the Royal Statistical Society, A*, 139(2), 183–204.
- Spurr, S. H. (1952). *Forest Inventory*. Ronald Press.
- Spurr, S. H. (1962). A measure of point density. *Forest Science*, 8, 85–95.
- Stage, A. R. (1973). Prognosis model for stand development. Research Paper INT-137, USDA Forest Service.
- Stoyan, D., Kendall, W. S., and Mecke, J. (1987). *Stochastic Geometry and its Applications*. Wiley.
- Strand, L. (1957). Virkningen av flatestørrelsen på nøyaktigheten ved prøveflatetakster. *Meddelelser fra det Norske Skogforsøksvesen*, XIV(48), 621–649.
- Sukhatme, P. V. and Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications*. Iowa State University Press.
- Tardif, G. (1965). Some considerations concerning the establishment of optimum plot size in forest survey. *Forestry Chronicle*, 41(1), 93–102.
- Wensel, L. C. and John, H. H. (1969). A statistical procedure for combining different types of sampling units in a forest inventory. *Forest Science*, 15(3), 307–317.
- Whittle, P. (1956). On the variation of yield variance with plot size. *Biometrika*, 43, 337–343.

Zöhner, F. (1972). The beta-distribution for best fit of stem-diameter-distributions. In *3e Colloque des Statisticiens Forestiers, IUFRO*, pages 91–106, Jouy-en-Josas, 7-11 sept. 1970. INRA.