Journal of Tropical Forest Science 15(1): 117–135 (2003)

A NEW TYPE OF SAMPLE UNIT FOR THE EFFICIENT ASSESSMENT OF DIVERSE TREE COMMUNITIES IN COMPLEX FOREST LANDSCAPES

D. Sheil,

Center for International Forestry Research (CIFOR), P. O. Box 6596 JKPWB, Jakarta 10065, Indonesia
E-mail: D. SHEIL@CGIAR.ORG

M. J. Ducey,

Department of Natural Resources, 215 James Hall, University of New Hampshire, Durham, NH 03824, USA

K. Sidiyasa,

Wanariset Samboja, Kilometer 38 Balikpapan-Samarinda, 76100, Kalimantan Timur, Indonesia

&

I. Samsoedin

FERDA (Pusat Litbang Hutan dan Konservasi Alam), Jalan Gunung Batu 5, Bogor 16610, Indonesia

Received June 2000

SHEIL, D., DUCEY, M. J., SIDIYASA, K. & SAMSOEDIN, I. 2003. A new type of sample unit for the efficient assessment of diverse tree communities in complex forest landscapes. We present a new and versatile sample unit suitable for rapid assessments of tropical forest in heterogeneous areas. The method uses multiple applications of small and easy-to-apply variable-area subunits, in which the area is defined by simple and objective rules. Compared with any fixed-area approach the sample unit is quick and easy to apply even in difficult terrain, and the amount of information collected varies little with stem densities. Unlike most variable-area methods difficult judgements are rare. Further, it cannot be extended to arbitrary size, but remains compact, allowing data to be linked to local-site variables. Useful data will generally result even in patchy and divided environments. Here we describe the method, discuss the nature of the resulting data and show how various stand characters can be calculated. The calculation of basic stand parameters from the sample data does not require any sophisticated analyses, and some worked examples are provided to ensure that the calculations are accessible. To demonstrate the statistical theory underlying this class of methods, and the good performance of the estimation methods, a more formal theoretical treatment is included as an appendix. The approach offers considerable promise for efficient forest assessments.

Key words: Bias - biodiversity - ecology - error-estimation - variable-area - inventory - survey - transect - estimator - maximum-likelihood

SHEIL, D., DUCEY, M. J., SIDIYASA, K. & SAMSOEDIN, I. 2003. Unit sampel jenis baru untuk penilaian lebih cekap bagi pelbagai komuniti pokok di landskap hutan yang kompleks. Kami membentangkan unit sampel baru dan serba guna yang sesuai untuk penilaian cepat hutan tropika di kawasan heterogen. Kaedah ini menggunakan aplikasi berganda bagi subunit kecil dengan kawasan berubah-ubah yang senang digunakan. Kawasan ditentukan oleh peraturan yang mudah dan objektif. Berbanding dengan pendekatan yang menggunakan kaedah kawasan yang ditetapkan, unit sampel ini cepat dan senang digunakan walaupun di bahagian rupa bumi yang sukar, dan maklumat yang dikumpul hanya berubah sedikit dengan ketumpatan batang. Tidak seperti kebanyakan kaedah kawasan berubah-ubah, penilaian sukar adalah jarang. Tambahan pula, ia tidak dapat digunakan untuk saiz sembarangan, tetapi kekal padat. Ini membolehkan data dihubungkan dengan pemboleh ubah tapak tempatan. Data berguna dapat dihasilkan walaupun dalam persekitaran berkelompok dan terbahagi. Kami menerangkan tentang kaedah, membincangkan tentang data yang dihasilkan dan menunjukkan bagaimana pelbagai ciri dirian dapat dikira. Pengiraan parameter dirian asas daripada data sampel tidak memerlukan analisis yang sukar, dan beberapa contoh kerja disediakan untuk menjelaskan pengiraan. Untuk menunjukkan teori statistik berasaskan kelas kaedah ini, dan prestasi kaedah anggaran yang baik, rawatan secara teori yang lebih formal dilampirkan. Kaedah ini menawarkan penilaian hutan yang lebih cekap.

Introduction

Ecologists and foresters have devised many ways to assess tree communities (e.g. Cochran 1977, Anonymous 1981, Gillison & Brewer 1985, Austin & Hyligers 1989, Kenkel *et al.* 1989, Thompson 1992, Roesch 1993, Schreuder *et al.* 1993, Morrison *et al.* 1995). Each approach has strengths and weaknesses making them more or less suited to any given objective and context. Choice of sampling procedure is a classical concern and numerous factors are relevant (see Kenkel *et al.* 1989 for a short review). Given a specific sampling objective or goal, choice of method will relate to costs and effort, to data quality, and to statistical aspects such as accuracy, precision and ease of analysis and interpretation. There is increasing interest in assessing forest composition and species richness. Such activities can be extremely demanding on both time and resources; this has led to a search to find rapid but effective methods.

One of the main factors to consider in tree sampling, when individual stems are to be recorded, is the range of stem densities encountered and the amount of data required versus the amount actually collected. Tree density can vary greatly, causing fixed-area plots to vary in per-plot recording effort and number of records. When fixed-area plots are placed in a forest that has an average density of m trees per plot, and those trees are distributed according to a uniform random distribution, the number of trees tallied on an individual plot will be distributed as a Poisson deviate with mean m and variance m (e.g. de Vries 1986). Taking the coefficient of variation as a measure of the relative accuracy of the estimate, we find it is proportional to $1/\sqrt{m}$. Thus, the relative accuracy of fixed area plots increases as density increases. At low densities the accuracy may be inadequate, while at high densities the effort required per plot may be unacceptable. This provides a simple rule-of-thumb criterion for univariate assessments, as many stand parameters show a similar variance relationship in normal closed forest (Dawkins 1952, 1985).

Analysis of biodiversity and community relationships presents additional challenges for sampling methods, that go beyond those encountered in the estimation of parameters such as density, basal area and volume. Interest in evaluating forest composition and richness has led to the development and use of many multivariate analytical procedures for ecological characterisations based on multiple species identities. The analyses for exploring compositional patterns are often computationally intensive (e.g. multi-dimensional scaling, canonical correspondence analysis, TWINSPAN, nonmetric-multi-dimensional scaling; see Ter Braak & Prentice 1988) and quantified relationships to define data adequacy are hard to derive except by trial (Hall 1991). Such methods attempt to distinguish patterns at the level of the differences between individual sample points. Most multivariate algorithms can only be computed with sufficient counts and species (i.e. the species composition of zero, one or few trees or species will simply not compute), and give little useful information until there is some genuine pattern to similarities and differences between individual sample compositions. Ordination patterns can be dominated by outliers and swamped by "noise". Hall (1991) indicated that 15 to 20 stems per plot are sufficient to reveal the major compositional patterns in one African forest. One solution to ensure usable data is to use plots big enough to include sufficient stems even in sites with low stem densities. It would be more efficient, however, to develop approaches which ensure each sample provides adequate, but not excessive, amounts of information.

Here we present a new type of sample-unit, describe some of its merits and outline how some basic parameters can be estimated. The method can be viewed as a refinement of the variable-area transect method developed by Parker (1979). Novelty lies in combining several small and easy-to-apply variable-area subunits, and an efficient combination of criteria to define the accepted stem-count within each. This procedure, as described here, is especially suitable for rapid and robust assessments of tropical forest composition, but the general approach could be adapted for wider application.

Fixed-count distance methods

Principally because of concerns about data-quantity and efficiency, fixed count (or "distance") approaches, employing variable plot areas, have been developed and used by ecologists in temperate forest. These methods provide a constant amount of information from each sample location. The point-centred-quarter method (PCQ), providing four tree records per sample point, is probably the best known of these approaches (Morisita 1954). Such methods promise many advantages. Cottam and Curtis (1956) noted some of these: "A number of methods have been described in recent years which utilise spacing distances instead of fixed area plots for the sampling of plant communities in compositional studies. All of these methods posses certain obvious advantages when compared with the standard plot techniques, since they are all more efficient in terms of results obtained per man-hour expended. They are faster, require less equipment and fewer workers, and are much more flexible, in that there is no need to adjust the sample size for the particular density of the vegetation under study. The advantage in speed is particularly great, with savings of 90% or more commonly obtained." However, these methods have remained relatively unexplored in rich tropical vegetation (see Hall 1991 for a short review) despite the fact that this is where efficient assessments are most needed.

There are two main approaches to fixed-count plots. In the circular-approach a centre point is selected and a circle (or segment as in the PCQ) of increasing radius is extended until a predetermined number of stems, r, is included. The horizontal radius (d_r) to the middle of the rth tree is recorded, and the area of the plot (containing $r - 1/2$ stems) is given by πd_r^2. The benefits of this approach are its simplicity and compactness. Compactness is useful if discrete site properties are to be recorded on a per plot basis and related to the vegetation (e.g. soil nutrient status sampled at the plot centre may be assumed as "representative"). In angle-delimited methods, such as the PCQ, judging sectors can prove difficult. In all methods, identifying the rth stem can be slow and clumsy when r is high. Linking small plots into networks is one solution to developing larger samples but can lead to multiple counting of some stems, and complex data.

The transect approach is to extend a single rectangular plot until it includes the required number of stems. For a transect of horizontal width w and horizontal distance to the rth stem d_r, the area is $w \times d_r$. Benefits of this approach are that, given a narrow enough transect width (< 20 m), it is relatively simple to accumulate any number of stems with little search effort or likelihood of omission errors. Engeman *et al.* (1994) found that variable-area transects performed well when compared with other distance-based methods in a simulation study. However, in low density tree cover, the sample may ultimately extend far from its origin, cross vegetation and site types and complicate the "best" location of any accessory site data (e.g. soil sampling, topographic parameters). Such factors may confound any search for relationships between composition, density and distribution. For example, if we were to look at species richness in a fixed number of trees in compact areas, and also in long transects, we should not be surprised that the second is "richer" – as it is likely to cover a wider range of site conditions. This "bigger habitat variation in more extended units" problem effects all variable-area methods, but is most evident in a transect approach.

In either the circular or transect approach, the areas can be used in two ways to estimate density X. Suppose n plots are tallied. Morisita's (1957) estimator is $X = (r - 1)\, S_{i=1..n}(1/A_i)/n$. A maximum likelihood estimator, developed by Parker (1979) for the transect case, is $X = (rn - 1)/\, S_{i=1..n}A_i$. Both estimators are unbiased when trees are distributed according to a uniform random distribution. The maximum likelihood estimator performs better (lower variance when unbiased, and lower bias and variance when biased) across a range of spatial patterns (Engeman & Sugihara 1998).

Sampling patchy clumped and divided distributions can lead to difficulties. For example, consider the PCQ being located on the edge of an open area–the first trees recorded on the "open side" may be far from the sample origin, be difficult to find and reach–and difficult to reconcile with the notion that they represent the vegetation at the sample point. Such "problem" locations may not be common, but they are troublesome, causing confusion and wasted effort. Field crews need to know how to proceed, but clear instructions are difficult to formulate without compromising the fixed counts approach.

Similar, but less marked, problems may relate to more subtle patterns of stem distributions and densities. The influences of population pattern on density estimation remain poorly explored even though the closely related topic of estimating dispersion (stem pattern) was one of the original emphases in the development of these distance methods in temperate ecology (Cottam & Curtis 1956). Nevertheless, such influences are generally slight, and should not lead to significant bias when samples are replicated. Numerous studies have in any case indicated that stem distributions in continuous forest can, when viewed from a sampling perspective, be considered to deviate little from random (e.g. Dawkins 1952, 1985, Lieberman & Lieberman1994 and references therein). Even where populations are not uniformly distributed, distance-based sampling methods remain useful, and biases are generally negligible when compared with normal sampling variability (Lessard *et al.* 1994).

A new approach

Our method is a modification of the variable-area transect method. First, rather than employing a series of independent transects, it uses short variable transect "cells" directed sideways from a central baseline (Figure 1). In this sense, the sample unit is a cluster of cells, each of which is a modified variable-area transect. The modification to each transect employs a decision-tree approach to terminating sampling effort on that transect, in order to make a good trade-off between the goals of similar number of trees sampled, compact sample area, and ease of implementation. In each cell of width w, sampling proceeds as follows.

(1) If a horizontal distance L_{min} is travelled without encountering any trees, that cell is tallied (i.e. scored) as empty (a zero).

(2) If at least one tree is tallied before reaching L_{min}, and some maximum number of trees r_{max} is tallied before reaching a maximum horizontal distance L_{max}, the cell is recorded as containing r_{max} trees, and its length is recorded as the distance from the centre-line to tree r_{max} (this measure is recorded to the stem centre, not the nearest point).

(3) If L_{max} is reached before r_{max} trees have been tallied, then sampling stops. The cell is recorded as containing the number of trees tallied so far, and its length is recorded as L_{max}.

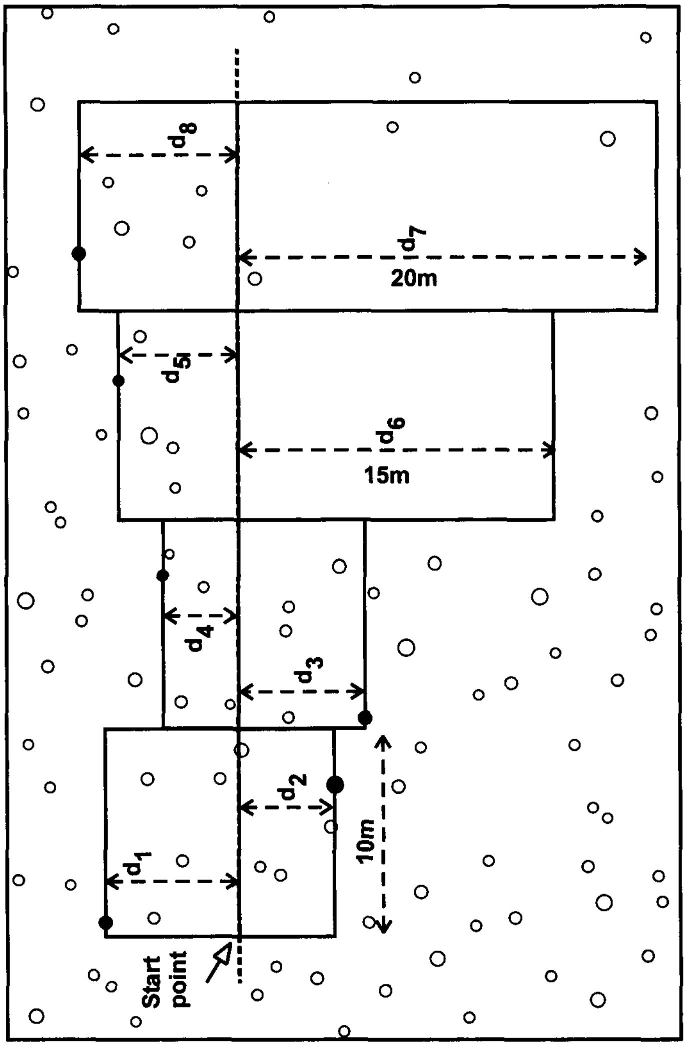

Figure 1 The sample unit. This is composed of eight variable cells of 10-m-wide that extend from the 40-m transect line. All distances are defined horizontally. Each cell captures five trees, or less, and the distance to the most distant tree included (filled in the figure) is recorded (d_1, d_2,... etc.). The maximum distance to search in each cell before deciding it is "empty" is 15 m (see d_6). The maximum distance to search to collect up to five stems is 20 m (see d_7)

In the specific example outlined here (as drawn in Figure 1) and now used in our studies in a rugged landscape in East Kalimantan, the centre line is 40 m long, and each cell is 10 m wide intended to include five trees ≥ 10 cm diameter at 1.3 m (dbh), so $r_{max} = 5$. We used $L_{min} = 15$ m and $L_{max} = 20$ m. These distances and measures were based on what was felt to be practical and workable in the field coupled with some general analyses as below. We initially field tested this method with only four cells, and later extended this to eight cells, providing a maximum of 40 stems per completed sample-unit.

Field aspects and numerical properties

Variable-area methods are usually defined by a fixed count and limitless search zone. In our method, the count requirement is relaxed when specific criteria are met. It is true that having *exactly* equivalent quantities of data from each sample unit has some convenient properties, particularly for analyses, but we contend that this is often unrealistic to attain, and sometimes confusing to interpret, e.g. when a single plot extends over a considerable area of bare ground. Exactly equivalent sample unit data is unrealistic in extensive surveys of tropical forest landscapes. Even setting aside the problem of fragmentary tree cover, some stems cannot be confidently identified to the species level in any case; this shortcoming alone meaning that stems are omitted from compositional analyses (even if technically they should not be). We contend that *approximately* equivalent data quantity is the most effective general option when all practicalities and limitations are considered.

Our field trials showed that the procedures were easily applied in all but the thickest vegetation. Very little time was spent on plot establishment, or measuring distances (about 3 to10 minutes depending on terrain and the need to cut marker sticks). In general the distances were easily observed and there was little trouble in identifying which stems are closest to the baseline, with few cases requiring discrimination of the 5th stem by comparative measurement (c. one in five).

We next show how the procedure collects consistent quantities of data independently of a wide range of stem densities. Considering the expected number of stem records in a randomly distributed population serves as an illustration that is amenable to direct analysis. If we examine the situation where stems are distributed randomly, uniformly and independently, the expected number of empty cells is described by an exponential relationship:

$$E(N_{empty}) = N_{total}\, e^{-Ax}$$

where

$E(N_{empty})$ is the mean number of empty cells per location,
N_{total} is the number of individual cells to be searched,
A is the per-cell search area (here the fraction of a hectare), and
x is the local stem density ($n\ ha^{-1}$).

For example given eight 10-m-wide cells to be searched over 15 m (i.e. N = 8, A = 10 m × 15 m/10 000 m^2 = 0.015 ha):

$$E(N_{empty}) = 8\, e^{-0.015x}$$

This horizontal 15-m search-limit relation, and that for each of a 5- and 20-m limits, are illustrated in Figure 2.

Similarly we can assess the mean number of records per sample when cell counts are not zero but are below five within the given search distance. Averaging each stem-count weighted by their relative probabilities indicates how the search distance influences data quantities at any given mean stem-density. This relationship for limiting horizontal distances of 10, 20 and 50 m for eight 10-m-wide cells, is plotted in Figure 3.

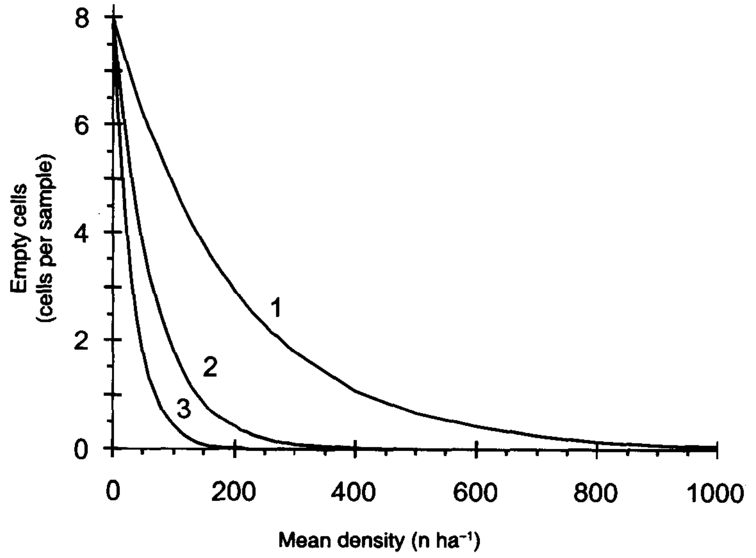

Figure 2 How the mean number of empty cells relates to stem-densities when the members of the sample population are randomly distributed. Examples are given where the maximum search distances are 5 m (1), 15 m (2) and 30 m (3).

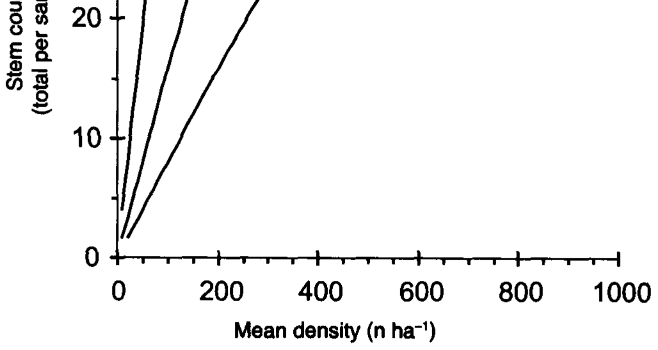

Figure 3 How the mean number of stem-records relates to stem-densities when the members of the sample population are randomly distributed. Examples show results for 10-m-wide transects (cells) searched to a maximum of 10 m (1), 20 m (2) and 50 m (3).

Many mature closed-canopy tropical forests have a stem density ($\geq$ 10 cm dbh) between 400 and 600 ha^{-1}. When we consider forests in this range, using a 20-m upper cut-off, the expected number of trees measured remains between 39 and 40, and the sample area remains quite compact. As density drops, the number of trees remaines above Hall's (1991) suggestion of 20 trees per sample, until the density falls below 150 ha^{-1}, at which point the forest is quite open. Within this range, the sample area begins to approach the maximum, which is a 40×40 m^2 (= 0.16 ha). As density falls further, the number of trees falls into and then below Hall's (1991) range. However, extending the sample to acquire larger numbers of trees could make the sample area quite large and, therefore, heterogeneous and difficult to relate to any localised site characteristics. The choice of a 15-m minimum horizontal distance, 20-m maximum distance, and maximum five trees sampled per cell reflected a pragmatic trade-off between uniformity of sample effort and compactness of sampled area.

Basic calculations: mean density, basal area and composition

The method was derived from a practical standpoint to allow efficient assessments of both stand parameters (density, basal area, etc.) and composition. A complete worked example is provided in Appendix 1. Any potential readers should be able to follow the simple arithmetic required. A fuller analytical account of the estimators and their statistical properties are provided in Appendix 2. Such an analysis is not required to use the method. The main conclusions are that these estimators are nearly unbiased under the idealised case of a uniform, independent random stem distribution, with bias small relative to variance even at low densities where the bias is most extreme.

Estimation of stand parameters proceeds in a fashion analogous to Morisita's (1957) approach. For each cell (individual variable-length transect), we calculate a density estimate, and for a group of cells (eight in our implementation) the density estimate is the average of the cell density estimates. For a cell, we can distinguish three situations:

(1) The transect is run out for a horizontal distance L_{empty}, and no trees are encountered. The cell is tallied as empty. In this case, the total density estimate for the cell, X_i, equals 0.

(2) The maximum number of trees r_{max} is encountered before L_{max} is reached; the total length of transect run in the cell is L_i. The total density estimate corresponds to the variable-area transect estimate for a single cell: $X_i = (r_{max} - 1)/(wL_i)$. Each individual stem counts as $x_i = ((r_{max} - 1)/r_{max})/(wL_i)$ similar trees per unit area toward calculations of density, basal area, and similar values for that cell; in common forest biometry terms, x_i is the expansion factor (n.b. x_i is per tree, X_i is per cell).

(3) The cell is extended to its maximum horizontal length L_{max}, and less than the maximum number (in our implementation, five) trees are tallied. If r_i trees are tallied, the total density estimate is $X_i = r_i / (wL_{max})$. The expansion factor for individual trees is $1/(wL_{max})$.

Regardless of the proportion of cells meeting the three criteria above, the estimate of total density based on the group of cells is the taken as the mean of the individual cell estimates.

For stand parameters to which individual trees contribute additively, such as basal area or biomass per hectare, we recommend the following procedure:

(1) Calculate the value of the variable of interest for each tree j in cell i. Call this value y_{ij}.

(2) Multiply the y_{ij} values by their corresponding x_i values to obtain per unit area estimates for each tree. Sum these values of $(x_i y_{ij})$ over all trees in the cell to obtain the per unit area estimate for the cell. Call the estimate for cell i Y_i.

(3) Average the Y_i values to obtain the best estimate for a group of cells.

An alternative approach would be to calculate a single average value per tree across the entire group of cells, and multiply this average by the density estimate for the group. However, since this approach discards information about the relationship between the contribution of a tree and the local density immediately around that tree, we speculate that it will produce less desirable estimates when trees are not distributed according to a uniform random process.

Calculating compositional information such as relative density or relative basal area, or the fraction of stems or basal area in a diameter class, involves a straightforward generalisation of the method for calculating stand parameters. First, total density and total basal area per hectare are calculated as presented above. Second, the density and basal area of a species are calculated following the same procedures as for a whole-stand parameter, except that the y_{ij} value for a tree is treated as zero unless the tree is in the species or diameter class of interest. Once the y_i values have been calculated, obtaining stems per hectare and basal area per hectare for each species follows directly from calculation of the Y_i values and their averages as outlined above. Finally, relative values or fractional contributions may be obtained by dividing the individual species densities and basal areas by the estimates of total density and basal area.

Species richness, unlike the other parameters, is best assessed at the scale of the entire sample–providing integer species counts quoted against the complete stem count (species per stems). A simple index to make such richness data comparable over a limited range of count variation is provided by Z = log(species counts)/ log(stem counts), from the power approximation suggested by Sheil *et al.* (1999). If more exact estimates of richness are required alternative approaches (e.g. Hurlbert 1971) are available.

Conclusions

The benefits of the new type of sample unit, especially when compared with more conventional sampling methods, fixed and variable-area, are as follows:

(1) Ease and speed of application. Even in open forests, the search distances are short, and accurate judgements can be made rapidly without numerous measurements and comparisons.

(2) The method is easily applied in difficult terrain. There is no need to define numerous plot edges and angles to be assured of the correct sample geometry. Accurate horizontal measurement of the baseline plus the horizontal measure to the fifth tree, or limiting distances, in each cell suffices.

(3) Unlike other variable-area method the sample unit cannot be extended arbitrarily far from the central baseline, but remains compact even in sparsely populated, divided, or patchy environments. The central baseline provides a reference line on which to assess accessory data such as soil properties or other site characters.

(4) The amount of information collected varies little even when forest densities vary greatly (unlike fixed area or angle-gauge approaches). The amount of effort required for each plot is not dependent upon stem densities.

(5) Each stem can only be recorded once and the resulting data is simple to compile.

(6) Open patches may impact some of the short variable-area cells but this will generally not result in unusable sample information.

Issues for consideration include:

(1) Variation in plot size and shape – in some studies this may not be appropriate and could confound relationships (e.g. when within plot variance is being compared with some site characteristics or stem density itself).

(2) Parameter biases are present. For precise assessments, more time-consuming methods are needed. There will also be compositional biases in patterned communities, i.e. over-representation of "gap-edge" stems, and reduced-representation of "clump-interior" stems. When totally unbiased estimates are required with a relatively small number of samples in very patchy environments other more sophisticated, but also more technically demanding, approaches may be warranted (e.g. see Thompson 1992, Roesch 1993).

The new approach is easy to use and reduces many problems faced when assessing vegetation in complex landscapes. Most of the field time is occupied with measuring, identifying and vouchering the trees themselves rather than with establishing plot boundaries, or collecting surplus records. The variable-area approach generally ensures appropriate quantities of information. However, the difficulties that can arise by insisting on fixed counts are greatly reduced by the distance criteria. Application is simple and involves little special effort. Biases arising within the method cannot be totally eliminated but appear negligible compared with normal sampling variance. Further field trials of the method, particularly in comparison with alternatives such as Bitterlich (prism) sampling (Grosenbaugh 1958, Bitterlich 1984), would be helpful in establishing its range of applicability and relative efficiency. Our new method offers high efficiency and guaranteed compactness. Rapid and broad characterisation rather than detailed evaluation of specific sites is enabled; such an approach is particularly well suited to rapid vegetation assessments. The approach might be adapted to any static populations scattered over large areas as long as satisfactory measurement procedures and criteria can be devised, dimensions and counts being selected to ensure adequate information is collected easily and efficiently.

Acknowledgements

These activities have been supported by ITTO through the project "Forest, Science and Sustainability: The Bulungan Model Forest" PD 12/97 Rev. 1 (F). We thank G. Applegate and K. MacDicken for commenting upon earlier texts. We are grateful to the field crews involved in the initial field trials, particularly I. Rahman, J. J. Afriastini, Z. Arifin and Y. Purwanto. We also wish to gratefully acknowledge the support of K. Kartawinata and the assistance of R. Go, M. Wan and I. Susilanasari.

References

ANONYMOUS. 1981. *Manual of Forest Inventory.* Forestry Paper 27. FAO, Rome. 200 pp.

AUSTIN, M. P. & HYLIGERS, P. C. 1989. Vegetation survey design for conservation: gradsect sampling of forests in north-eastern New South Wales. *Biological Conservation* 50: 13–32.

BITTERLICH, W. 1984. *The Relascope Idea.* Commonwealth Agricultural Bureaux, Farnham Royal, England. 242 pp.

COCHRAN, W. G. 1977. *Sampling Techniques.* 3rd edition. Wiley, New York. 428 pp.

COTTAM, G. & CURTIS, J. T. 1956. The use of distance measures in phytosociological sampling. *Ecology* 37: 451–460.

DAWKINS, H. C. 1952. Experiments in low percentage enumerations of tropical high-forest. *The Empire Forestry Review* 31: 131–149.

DAWKINS, H. C. 1985. Sampling errors in extensive forest inventory. *Commonwealth Forestry Review* 64: 78–79.

DE VRIES, P. G. 1986. *Sampling Theory for Forest Inventory.* Springer-Verlag, Berlin. 446 pp.

ENGEMAN, R. M. & SUGIHARA, R. T. 1998. Optimization of variable area transect sampling using Monte Carlo simulation. *Ecology* 79(4): 1425–1434.

ENGEMAN, R. M., SUGIHARA, R. T., PANK, L. F., & DUSENBERRY, W. E. 1994. A comparison of plotless density estimators using Monte Carlo simulation. *Ecology* 75(6): 1769–1779.

Gillison, A. N. & Brewer, K. R. W. 1985. The use of gradient directed transects or gradsects in natural resource surveys. *Journal of Environmental Management* 20: 103–127.

Grosenbaugh, L. R. 1958. *Point-sampling and Line Sampling: Probability Theory, Geometric Implications, Synthesis.* U.S.D.A. Forest Service, Southern Forest Experiment Station, Occasional Paper 160. 34 pp.

Hall, J. B. 1991. Multiple-nearest-tree sampling in an ecological survey of Afromontane catchment forest. *Forest Ecology and Management* 42: 245–266.

Hurlbert, S. H. 1971. The non-concept of species diversity: a critique and alternative parameters. *Ecology* 52: 577–586.

Kenkel, N. C., Juhász-Nagy, P. & Podani, J. 1989. On sampling procedures in population and community ecology. *Vegetatio* 83: 195–207.

Lessard, V., Reed, D. D. & Monkevich, N. 1994. Comparing N-tree distance sampling with point and plot sampling in northern Michigan forest types. *Northern Journal of Applied Forestry* 11(1): 12–16.

Lieberman, M. & Lieberman, D. 1994. Patterns of density and dispersion of forest trees. Pp. 106–119 in McDade, L. A., Bawa, K. S., Hespenheide, H. & Hartshorn, G. S. (Eds.) *La Selva: Ecology and Natural History of a Neotropical Rainforest.* University of Chicago Press, Chicago, Illinois, USA.

Morisita, M. 1954. Estimation of population density by spacing method. *Memoirs of the Faculty of Science Kyushu University, Series E* 1: 187–197.

Morisita, M. 1957. A new method for the estimation of density by the spacing method applicable to nonrandomly distributed populations. *Physiology and Ecology* 7: 134–144 (Japanese; U.S.D.A. Translation 11116).

Morrison, D. A., Le Brocque, A. F. & Clarke, P. J. 1995. An assessment of some improved techniques for estimating the abundance (frequency) of sedentary organisms. *Vegetatio* 120: 131–145.

Parker, K. R. 1979. Density estimation by variable area transect. *Journal of Wildlife Management* 43(2): 484–492.

Roesch, F. A. 1993. Adaptive cluster sampling for forest inventories. *Forest Science* 39: 655–669.

Schreuder, H. T., Gregoire, T. G. & Wood, G. B. 1993. *Sampling Methods for Multi-resource Forest Inventory.* Wiley & Sons Inc., New York.

Sheil, D., Sayer, J. A. & O'Brien, T. 1999. Tree diversity and conservation in logged rainforest. *Science* 284: 1587.

Ter Braak, C. J. F. & Prentice, I. C. 1988. A theory of gradient analysis. *Advances in Ecological Research* 18: 271–317.

Thompson, S. K. 1992. *Sampling.* Wiley-Interscience Publication, New York. 358 pp.

Appendix 1 A worked example

A hypothetical data set from one eight-cell sample unit is provided in Table 1. The data are presented as if it were in a spread-sheet. The calculations follow the procedures provided in the main text as calculations for density and of basal area. Table 2 shows the summary data, in which the cell means have been calculated. The total sample unit values are the (equally weighted) means of the eight cell estimates.

Table 1 Hypothetical data set from one eight-cell sample

Cell	Species	dbh (cm)	Stem BA (cm²)	Cell-class and estimator (per stem)	Length of cell L_i (m)	Area (m²) L_i 10 m	x_i (N m^{-2})	BA x_i
1	Sp 1	19	283.53	B: $x_i = 4/(5 \text{ Area})$			0.0050	1.418
1	Sp 1	14	153.94				0.0050	0.770
1	Sp 1	22	380.13				0.0050	1.901
1	Sp 2	10	78.54				0.0050	0.393
1	Sp 2	11	95.03		16	160	0.0050	0.475
2	Sp 1	35	962.11	B: $x_i = 4/(5 \text{ Area})$			0.0053	5.131
2	Sp 1	32	804.25				0.0053	4.289
2	Sp 2	14	153.94				0.0053	0.821
2	Sp 2	11	95.03				0.0053	0.507
2	Sp 2	11	95.03		15	150	0.0053	0.507
3	-		0.00	A: $x_i = 0$			0	0.000
3	-		0.00					
3	-		0.00					
3	-		0.00					
3	-		0.00		15	150		
4	Sp 1	44	1520.53	C: $x_i = 1/A$			0.0050	7.603
4	Sp 1	50	1963.49				0.0050	9.817
4	Sp 1	17	226.98				0.0050	1.135
4	Sp 2	12	113.10				0.0050	0.565
4	-				20	200		0.000
5	Sp 1	13	132.73	C: $x_i = 1/A$			0.0050	0.664
5	Sp 1	16	201.06				0.0050	1.005
5	-							
5	-							
5	-				20	200		
6	Sp 2	10	78.54	B: $x_i = 4/(5\text{Area})$			0.0067	0.524
6	Sp 2	10	78.54				0.0067	0.524
6	Sp 2	14	153.94				0.0067	1.026
6	Sp 2	16	201.06				0.0067	1.340
6	Sp 2	20	314.16		12	120	0.0067	2.094
7	Sp 1	23	415.48	B: $x_i = 4/(5\text{Area})$			0.0053	2.216
7	Sp 1	11	95.03				0.0053	0.507
7	Sp 2	12	113.10				0.0053	0.603
7	Sp 2	16	201.06				0.0053	1.072
7	Sp 2	15	176.71		15	150	0.0053	0.942

continued

Table 1 (continued)

Cell	Species	dbh (cm)	Stem BA (cm^2)		Cell-class and estimator (per stem)	Length of cell L_i (m)	Area (m^2) L_i 10 m	x_i ($N\ m^{-2}$)	BA x_i
8	Sp 1	80	5026.54	B:	$x_i = 4/(5\text{Area})$			0.0042	21.164
8	Sp 1	23	415.48					0.0042	1.749
8	Sp 2	11	95.03					0.0042	0.400
8	Sp 2	12	113.10					0.0042	0.476
8	Sp 2	16	201.06			19	190	0.0042	0.847

Table 2 Per cell summary and final values for the worked example

	Cell mean density ($N\ ha^{-1}$)*			Cell mean basal area $m^2\ ha^{-1}$ (= $cm^2\ m^{-2}$)		
Cell	Sp 1	Sp 2	All	Sp 1	Sp 2	All
1	150.0	100.0	250.0	4.088	0.868	4.956
2	106.7	160.0	266.7	9.421	1.835	11.255
3	0.0	0.0	0.0	0.000	0.000	0.000
4	150.0	50.0	200.0	18.555	0.565	19.121
5	100.0	0.0	100.0	1.669	0.000	1.669
6	0.0	333.3	333.3	0.000	5.508	5.508
7	106.7	160.0	266.7	2.723	2.618	5.341
8	84.2	126.3	210.5	22.914	1.723	24.637
Mean**	87.2	116.2	203.4	7.421	1.640	9.061

* Converted from stem per m^2 to per ha by multiplying by 10 000.

**These are the values that would be taken for the sample plot as a whole.

Appendix 2 Derivation of the estimators for density

Assume that there are X trees per hectare, distributed according to a uniform, independent process across the area to be sampled ("Poisson forest", de Vries 1986). If a single fixed-area plot of horizontal width w and length L is sampled, the number of trees tallied (i.e. recorded within the sample) will be distributed according to a Poisson distribution with parameter XwL:

$$p_L(r) = \frac{(XwL)^r \exp(-XwL)}{r!}$$

The minimum-variance unbiased estimator of X in this case is $Z_1 = r/(wL)$.

Now consider the case of a single cell of width w. In the basic variable-area transect case, such that $L_{min} = L_{max}$ = infinity, the distance L travelled to tree r_{max} is distributed according to a gamma distribution with parameters X, r_{max}:

$$p_{r_{max}}(L) = \frac{X(XwL)^{r_{max}-1} \exp(-XwL)}{w\Gamma(r_{max})}$$

and its reciprocal will be distributed according to an inverse gamma distribution. The mean of the inverse gamma distribution is $Xw/(r-1)$, so a natural and unbiased estimator for X is $Z_2 = (r-1)/(wL)$. If many such estimates, taken from several transects, are averaged, this yields Parker's (1979) adaptation of Morisita's (1957) estimator.

Now suppose we perform variable-area transect sampling, but with an upper limit L_{max} to the distance travelled. In other words, if we reach L_{max} before tallying r_{max} trees, we record the number of trees tallied so far r, and the distance L_{max}. If we encounter r_{max} trees before reaching L_{max}, we record the number of trees as r_{max} and the distance as L, the distance travelled to tree r_{max}. A tempting estimator to use in this case is Z_1 if the distance is recorded as L_{max}, and Z_2 if it is not. However, is such an estimator – call it Z_3 – unbiased? To see that it is, let us examine its expectation based on the probability distributions of Z_1 and Z_2:

$$E(Z_3) = \int_0^{L_{max}} \frac{(r_{max}-1)X^{r_{max}}(wL)^{r_{max}-2}\exp(-XwL)}{w\Gamma(r_{max})}\,dl + \sum_{r=0}^{r_{max}-1} \frac{rX^r(wL_{max})^{r-1}\exp(-XwL_{max})}{r!}$$

A series of simplifications gives

$$E(Z_3) = X\left[\frac{\gamma(r_{max}-1, XwL_{max})}{\Gamma(r_{max}-1)} + \sum_{r=1}^{r_{max}-1} \frac{(XwL_{max})^{r-1}\exp(-XwL_{max})}{(r-1)!}\right]$$

where γ is the incomplete gamma function. Noting that the summand in the right-hand expression within the brackets is of the same form as $p_L(r)$, and that it equals the probability that $r_{max} - 2$ or fewer trees would be tallied within distance L_{max}, i.e. that the distance to tree $r_{max} - 1$ is greater than L_{max}, we may further write

$$E(Z_3) = X\left[\frac{\gamma(r_{max} - 1, XwL_{max})}{\Gamma(r_{max} - 1)} + 1 - \frac{\gamma(r_{max} - 1, XwL_{max})}{\Gamma(r_{max} - 1)}\right] = X$$

which proves that Z_3 is unbiased. Furthermore, since Z_3 is unbiased, the mean of a series of independent estimates for n cells calculated using Z_3 is also an unbiased estimator of X.

Now suppose we sample as before, but sampling stops if no trees have been encountered before travelling distance L_{min}. If that occurs, we use 0 as our estimate of density; otherwise, we use Z_3. Call this estimator Z_4. It is clear that if Z_3 is unbiased, Z_4 must be biased, and underestimate X, because the sampling situation is identical except that some trees actually occurring between L_{min} and L_{max} are not tallied. We can distinguish two possibilities. One possibility is that even though no trees were encountered before reaching L_{min}, r_{max} trees would have been encountered had the sampling continued toward L_{max} (the sample "should have been" a variable-area transect sample). The second possibility is that even though no objects were encountered before reaching L_{min}, $r < r_{max}$ objects would have been encountered had the sampling continued to L_{max} (the sample "should have been" a quadrat sample). The probabilities and adjustments to the estimator are relatively easy to define. Call the first D_1 and the second D_2. The bias of Z_4 is the negative of the sum of these two components.

$$D_1 = \exp(-XwL_{min}) \int_{L_{min}}^{L_{max}} \frac{(r_{max} - 1)X^{r_{max}}(wL - wL_{min})^{r_{max}-1}\exp[-Xw(L - L_{min})]}{wL\Gamma(r_{max})} dL$$

$$D_2 = \exp(-XwL_{min}) \sum_{r=1}^{r_{max}-1} \frac{[Xw(L_{max} - L_{min})]^r \exp[-Xw(L_{max} - L_{min})]}{wL_{max}[(r-1)!]}$$

Both of these quantities are difficult to assess analytically, but straightforward to obtain numerically. The bias will be a function of L_{min}, L_{max}, r_{max}, and unfortunately, the unknown density X. Results across a range of density are shown in Figure 4 for the case $L_{min} = 15$ m, $L_{max} = 20$ m, $r_{max} = 5$ used by the authors. Bias only becomes apparent in a relative sense when the density becomes quite low, and outside the range typically encountered in closed tropical forests. Furthermore, when one unit of eight cells forms the estimate for a tract, the bias is negligible and is easily swamped by the variance. When ten units (80 cells total) are employed to form the estimate, bias still accounts for a small fraction of mean squared error except when $X < 100$ trees ha^{-1}.

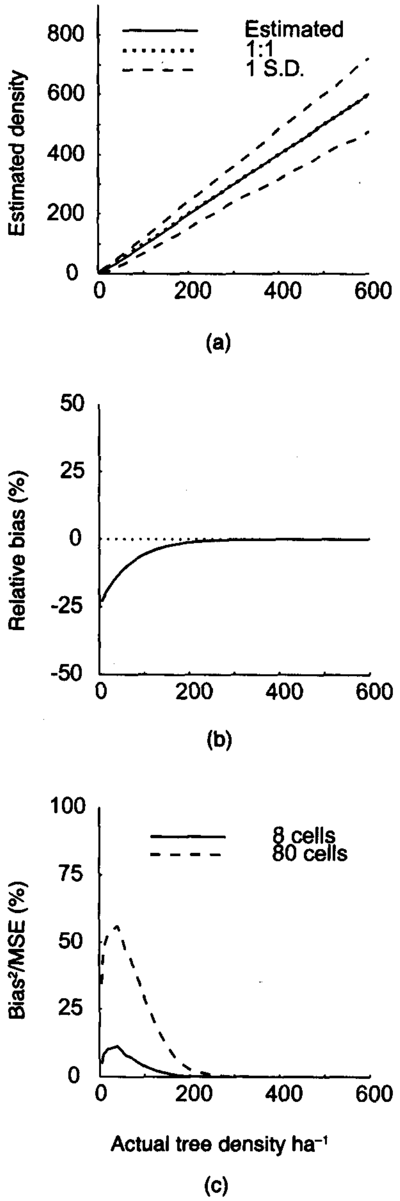

Figure 4 Bias and variance of the sampling technique when the members of the sample population are randomly distributed. (a) Expected value (solid line) plus or minus one standard deviation (dashed line) when eight cells form the sample unit. The true density is shown for reference as a dotted line. (b) Percent relative bias, equal to 100 × (expected estimate – true density)/true density. The relative bias is negligible except at very low densities. (c) Squared bias as a per cent of mean squared error, for one eight-cell sample unit (solid line) and ten eight-cell sample units (dashed line). When one group of cells is used to form the estimate, bias is negligible even at very low densities. If larger numbers of cells are used (e.g. 80), bias contributes appreciably to mean squared error but is not the dominant term unless density is very low.

A maximum-likelihood estimation approach might appear a desirable alternative. This is especially true given Engeman and Sugihara's (1998) results showing that Parker's (1979) original ML estimator dominates his analog to the Morisita (1957) estimator, even for non-Poisson distributions of objects. However, the likelihood formulas for the case employed here are complex, and do not yield a simple closed-form solution.