

Stratifying a Forest Removes Between-Stratum Differences from Sampling Error, and Allocating More Plots to Variable Strata Reduces It Further

— with the same thirty plots, allocating by area share cuts the variance of the estimated mean to under a quarter, and allocating by stratum variability cuts it to about two thirds of that —the separator is whether plots are allocated by area share or by the variability of each stratum — [Paper 938]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

YOI: <https://www.yamagishiyuuki.com/yoi/938>

— *From thinking to syncing.* —

Abstract

When the mean timber volume of a forest is estimated by measuring a limited number of plots, the way plots are chosen changes the estimation error. This paper takes a forest with three strata, young, middle and old, derives the variance of the estimate for thirty plots taken at random, allocated by area share, and allocated by stratum variability, and confirms the formulas by random sampling from a synthetic forest. No new theorem or law is claimed. First, most of the forest's variance is difference between strata: with area shares 0.5, 0.3 and 0.2 and stratum means of 50, 150 and 300 cubic metres per hectare, the within-stratum variance was 2680.000 against a between-stratum variance of 9100.000 (Section 2). Second, allocating by area share removes between-stratum differences from the error: taking thirty plots at random gives a variance of the estimated mean of 392.666667; allocating 15, 9 and 6 plots by area share gives 89.333333, a ratio of **0.227504** (Section 3). Third, allocating by variability reduces it further: allocating 7.143, 8.571 and 14.286 plots in proportion to area times variability gives a variance of 58.800000, a ratio of **0.149745** to random (Section 3). Fourth, random sampling needs many times more plots for the same precision: matching the area-share allocation takes 131.866 plots, and matching the variability allocation **200.340**. Taking 20000 samples from a synthetic forest, the three variances matched the formulas within ratios of 0.95 to 1.05 (Section 4). Fifth, the separator is whether plots are allocated by area share or by the variability of each stratum: allocating by area share removes the luck of how many plots fall in each stratum. Allocating by variability moves plots toward the strata that are hard to estimate, cutting error further with the same number of plots (Section 5). **Keywords:** stratified sampling, proportional allocation, Neyman allocation, variance of the estimate, survey sampling, forest inventory.

Scope of this paper (scope note): No new theorem or law is claimed: stratified sampling, proportional allocation, allocation matched to stratum variability (Neyman allocation), and the variance formulas for each are known in sampling theory. No real forest is claimed: area shares, stratum means and variabilities, and the 30 plots are set values corresponding to no particular forest or species. Stratum boundaries and costs are not treated: how to draw stratum boundaries, strata with different measurement costs, and two-stage sampling within plots are not treated. Stratum variability is assumed known: the allocation matched to variability is the value when the variability of each stratum is known in advance. Relation to earlier papers: Paper 891, in its scope note, did not enter the whole forest, treated only the volume of a single tree, and did not treat how to sample a whole forest; this paper sits in that seat of how to sample a whole forest (no earlier occupant). Paper 581 showed that the age at which a forest takes up carbon fastest is not the age at which it holds the most. Here it is counted how strata divided by stand age change the error in estimating mean volume. What is added: splitting the forest's variance into within and between strata, deriving the variances and ratios for three allocations, finding the number of random plots needed for the same precision, confirming the formulas by random sampling from a synthetic forest, and placing the separator on "allocating by area share, or by variability".

1. Notation and the Question

Setting: divide the forest into strata h with area shares W_h , stratum means μ_h and within-stratum variabilities S_h . When the forest mean is estimated from n plots, the variances of the estimate under random sampling, area-share (proportional) allocation and allocation matched to variability (Neyman allocation) are:

$$V_{\text{random}} = \frac{\sum W_h S_h^2 + \sum W_h (\mu_h - \mu)^2}{n}, \quad V_{\text{prop}} = \frac{\sum W_h S_h^2}{n}, \quad V_{\text{Neyman}} = \frac{(\sum W_h S_h)^2}{n}$$

Neyman allocation gives stratum h $n W_h S_h / \sum W_k S_k$ plots. The check sets $W = (0.5, 0.3, 0.2)$, $\mu = (50, 150, 300)$, $S = (20, 40, 100)$, $n = 30$, and samples 20000 times from a synthetic forest of 100000 trees, normally distributed within each stratum.

Question: with the same number of plots, how should they be allocated to reduce the error in estimating the forest mean?

2. The Breakdown of the Forest's Variance

Computation 2.1: the overall mean was 130.000, the within-stratum variance 2680.000 and the between-stratum variance 9100.000.

Observation 2.2

most of the forest's variability is not within strata but between them. Mean volume differs sixfold between young and old stands, and that difference makes up most of the forest's variance: taking plots at random, the luck of how many plots fall in each stratum brings this between-stratum difference into the error.

3. Three Allocations

Computation 3.1: 30 plots.

allocation	plots per stratum	variance of estimate	standard error	ratio to random
random	by chance	392.666667	19.815819	1
area share	15, 9, 6	89.333333	9.451631	0.227504
matched to variability	7.143, 8.571, 14.286	58.800000	7.668116	0.149745

Observation 3.2

allocating by area share removes between-stratum differences from the error, and matching to variability reduces it further. Allocating by area share drops the between-stratum variance 9100.000 from the formula, making the variance 0.227504 times as large: matching to variability moves plots toward the more variable old stratum, and even with fewer plots in the young stratum the overall variance falls to 0.149745 times, which is 0.658209 of the area-share allocation.

4. Plots Needed for the Same Precision, and a Random Check

Computation 4.1: random sampling needs 131.866 plots to match area-share allocation and **200.340** plots to match allocation by variability.

Computation 4.2: variances of the estimate from 20000 samples of the synthetic forest (the variability allocation rounded to 7, 9 and 14 plots).

allocation	random variance	formula	mean estimate
random	387.607516	392.666667	130.057
area share	89.270611	89.333333	129.795
matched to variability (rounded)	59.152657	58.857143	129.834

For all three, the ratio of random variance to formula lay between 0.95 and 1.05, and the synthetic forest's mean was 129.865.

5. The Separator

Observation 5.1

the separator is whether plots are allocated by area share or by the variability of each stratum. Allocating by area share removes a large source of error, the difference between strata, and with the same 30 plots the variance becomes 0.227504 times as large: allocating by variability moves plots toward strata that are hard to estimate even if small in area, and the variance becomes 0.149745 times. Random sampling needs 200.340 plots for the same precision.

Observation 5.2

what is not said. Real forests, how to draw stratum boundaries, differing costs by stratum, two-stage sampling, and the case where stratum variability is unknown are not treated. What is shown is that dividing into strata and allocating plots greatly reduces error for the same number of plots, and that the reduction differs by allocation.

6. Roll Call and Closing

1. variance formulas for three allocations (Section 1): within and between strata.
2. within 2680.000, between 9100.000 (Computation 2.1): mostly between.
3. area share: ratio **0.227504** (Computation 3.1): between-stratum differences removed.
4. matched to variability: ratio **0.149745** (Computation 3.1): reduced further.

5. **200.340 random plots** (Computation 4.1): **needed for the same precision.**
6. **random and formula within 0.95 to 1.05** (Computation 4.2): **confirmed on a synthetic forest.**
7. **area share, or variability** (Observation 5.1): **the separator; the scope note of Paper 891; Paper 581.**

Closing: Stratifying a forest removes between-stratum differences from sampling error, and allocating more plots to variable strata reduces it further. In a three-stratum forest, most of the overall variance was difference between strata: with the same 30 plots, allocating by area share made the variance of the estimated mean 0.227504 times that of random sampling and allocating by variability 0.149745 times, and random sampling needed 200.340 plots for the same precision. The variances from random samples of the synthetic forest matched the formulas for all three: the separator is whether plots are allocated by area share or by the variability of each stratum. Placed among the earlier papers: the volume estimation that Paper 891 kept to a single tree was extended to how a whole forest is sampled, and the forest divided by stand age in Paper 581 was counted as strata. To be honest: no real survey design is claimed. How much the error of the same number of plots changes with allocation, and no further.

References

1. Cochran, W. G. (1977). *Sampling Techniques*. Wiley.
 2. Neyman, J. (1934). On the two different aspects of the representative method. *Journal of the Royal Statistical Society*.
 3. Yamagishi, Y. (2026). *The Age at Which a Forest Takes Up Carbon Fastest Is Not the Age at Which It Holds the Most* [Paper 581]. YOI: <https://www.yamagishiyuuki.com/yoi/581>; DOI: 10.5281/zenodo.23014343
 4. Yamagishi, Y. (2026). *The Error in a Standing Tree's Volume Stacks Three Assumptions, and the Hardest to Measure, the Form Factor, Counts Most* [Paper 891]. YOI: <https://www.yamagishiyuuki.com/yoi/891>
-

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)