# A Weighted Fitting Approach for Diameter Distributions from Horizontal Point Sampling

Gregory E. Paradis

October 22, 2025

**Abstract**

Horizontal point sampling (HPS) produces size-biased tallies that cannot be fit directly with standard probability distributions without distorting diameter distribution estimates. Previous work resolves this by deriving bespoke size-biased probability density functions (PDFs) for each assumed distribution. We revisit the problem and formalise a weighted non-linear least squares approach that fits standard-form PDFs to expanded HPS stand tables while preserving the statistical equivalence with the size-biased formulation. The new pipeline leverages contemporary open-source software, is fully reproducible, and includes accompanying code that regenerates all figures and tables. Computational experiments on permanent sample plot data from Quebec demonstrate that the weighted method matches the reference approach to machine precision across Weibull and Gamma distributions. The manuscript and companion software provide a turnkey solution for practitioners who require stable, transparent, and replicable HPS diameter distribution fitting.

## 1 Introduction

Horizontal point sampling (HPS), also referred to as angle count or Bitterlich sampling (Bitterlich, 1947), is widely used to quantify stand structure in managed forests. Diameter distributions derived from HPS tallies underpin yield modelling, inventory projection, and ecological assessments. A long-standing challenge is that HPS tallies are intrinsically *size-biased*: the inclusion probability of a stem is proportional to its basal area, which varies with the square of diameter. When standard probability density functions (PDFs) are fit directly to expanded HPS stand tables, small diameter classes exert excessive influence and the fitted distribution is biased (Van Deusen, 1986; Gove, 2000; Ducey and Gove, 2015). The conventional remedy is to derive size-biased forms of candidate PDFs and fit them to the unexpanded tallies.

Deriving size-biased PDFs requires specialised algebra and bespoke software implementations. Even when available, the resulting optimisation problems often impose complex parameter bounds that reduce numerical stability. As a result, practitioners occasionally bypass appropriate size-bias corrections, leading to over-fitted distributions and poor downstream predictions.

This paper revisits the simplified method originally introduced in Paradis (2019), extends it with a formal equivalence argument, and packages the approach in a modern reproducible research workflow. The core idea is to fit standard-form PDFs to expanded HPS stand tables while embedding size-bias weights directly in the fitting algorithm. We show that the weighted objective function is proportional to the objective of the reference size-biased estimator, yielding identical parameter estimates and fitted curves. The new workflow includes:

1. a clean mathematical exposition establishing the equivalence between weighted and size-biased fitting;
2. an updated computational experiment using permanent sample plots from Quebec to benchmark Weibull and Gamma distributions across multiple meta-plots; and
3. a fully reproducible project scaffold containing LaTeX sources, notebooks, and Python scripts that regenerate all figures and tables.

The remainder of the manuscript introduces notation and the reference method, derives the weighted estimator, reports numerical comparisons, and discusses implications for inventory analysis pipelines.

# 2 Methods

## 2.1 Notation

Let $X$ denote the diameter at breast height (DBH) of a stem measured in centimetres. Assume that $X$ follows a continuous distribution with PDF $f(x;\boldsymbol{\theta})$ and cumulative distribution function $F(x;\boldsymbol{\theta})$. Consider an HPS inventory compiled using a basal area factor (BAF) $C_{BA}$. Let $\mathcal{I}$ denote the set of DBH classes and $x_i$ the centre of class $i \in \mathcal{I}$. For an inventory comprising plots $\mathcal{J}$, the HPS tally for class $i$ in plot $j$ is $t_{ij}$ and the corresponding mean basal area is $\bar{g}_{ij}$. The expanded stand table estimate of stems per hectare in class $i$ is

$$\hat{y}_i = C_{BA} \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} t_{ij} \bar{g}_{ij}^{-1}. \tag{1}$$

Because the inclusion probability of a stem is proportional to its basal area, the HPS tally is size-biased of order two. The HPS expansion factor that maps a tally back to expected stems per hectare at DBH $x$ is

$$f_E(x; C_{BA}) = \frac{40000\, C_{BA}}{\pi x^2}, \tag{2}$$

with multiplicative inverse (compression factor):

$$f_C(x; C_{BA}) = \frac{\pi x^2}{40000\, C_{BA}} = f_E(x; C_{BA})^{-1}. \tag{3}$$

## 2.2 Reference estimator

The reference approach (Van Deusen, 1986; Ducey and Gove, 2015) fits a size-biased PDF $f_{\mathrm{sb}}(x;\boldsymbol{\theta})$ to the raw HPS tallies. For any assumed standard-form PDF $f(x;\boldsymbol{\theta})$, the size-biased form of order two is

$$f_{\mathrm{sb}}(x;\boldsymbol{\theta}) = \frac{x^2 f(x;\boldsymbol{\theta})}{\int_0^\infty x^2 f(x;\boldsymbol{\theta})\,\mathrm{d}x}. \tag{4}$$

The typical objective minimises the sum of squared deviations between expected and observed tallies:

$$Z_{\mathrm{C}}(\boldsymbol{\theta}) = \sum_{i\in\mathcal{I}} \left[t_i - s\, f_{\mathrm{sb}}(x_i;\boldsymbol{\theta})\right]^2, \tag{5}$$

where $t_i = \sum_{j\in\mathcal{J}} t_{ij}$ and $s$ is a scaling parameter that reconciles the continuous PDF with discrete bin counts.

## 2.3 Weighted estimator

Our alternative estimator fits the standard-form PDF directly to the expanded stand table while embedding the size-bias correction through bin-wise weights. Let $\hat{y}_i = f_E(x_i;C_{BA})t_i$ denote the expanded tally in class $i$. The weighted objective is

$$Z_{\mathrm{T}}(\boldsymbol{\theta}) = \sum_{i\in\mathcal{I}} w_i^2 \left[\hat{y}_i - s\, f(x_i;\boldsymbol{\theta})\right]^2, \qquad w_i = f_C(x_i;C_{BA}), \tag{6}$$

which mirrors a non-linear least squares problem with heteroscedastic error variance proportional to the expansion factor.

## 2.4 Equivalence

Substituting $\hat{y}_i = f_E(x_i;C_{BA})t_i$ and $w_i = f_C(x_i;C_{BA})$ into (6) yields

$$\begin{aligned}
Z_{\mathrm{T}}(\boldsymbol{\theta}) &= \sum_{i\in\mathcal{I}} \left[f_C(x_i;C_{BA})f_E(x_i;C_{BA})t_i - f_C(x_i;C_{BA})s f(x_i;\boldsymbol{\theta})\right]^2 \\
&= \sum_{i\in\mathcal{I}} \left[t_i - s\,\frac{f(x_i;\boldsymbol{\theta})}{f_E(x_i;C_{BA})^{-1}}\right]^2 \\
&= \sum_{i\in\mathcal{I}} \left[t_i - s\,\tilde{c}\,x_i^2 f(x_i;\boldsymbol{\theta})\right]^2,
\end{aligned} \tag{7}$$

where $\tilde{c} = \pi/(40000C_{BA})$ is a constant that does not depend on $\boldsymbol{\theta}$. The normalising constant in (4) is also independent of $\boldsymbol{\theta}$; therefore the minimisers of $Z_{\mathrm{C}}(\boldsymbol{\theta})$ and $Z_{\mathrm{T}}(\boldsymbol{\theta})$ are identical. The full derivation is included in the supplementary materials and extends to unbinned likelihood formulations.

## 2.5 Computational experiment

We reproduced and updated the experiment introduced in Paradis (2019). Permanent sample plot (PSP) data collected across Quebec (Gouvernement du Québec, 2019) provide stem tallies for 30 meta-plots combining species groups and cover types. The PSP tallies originate from fixed-area plots with constant expansion factors. We emulate HPS tallies by scaling each bin by the reciprocal of the HPS expansion factor, preserving the size-bias structure. For each of three representative meta-plots—spruce-pine- fir-larch softwood (SPFL-S), birch mixedwood (Birch-M), and maple hardwood (Maple-H)—we fit two distributions (Weibull and Gamma) using both the reference and weighted estimators. The experiment is implemented in modern Python (SciPy, NumPy, lmfit) and all scripts are provided in the companion repository.

# 3 Results

Figure 1 compares fitted distributions across the six meta-plot–distribution combinations. In both HPS tally space and stand table space the curves produced by the reference and weighted estimators are visually indistinguishable. Residual sum of squares (RSS) values computed on the native scale of each estimator differ by less than $10^{-6}$ in every case (Table 1). Chi-square goodness-of-fit statistics agree to numerical precision, confirming that the weighting scheme reproduces the role of explicit size-biasing.

Although the absolute RSS values depend on binning choices, the difference between methods is several orders of magnitude smaller than within-method RSS, which establishes practical equivalence. Parameter estimates (shape, scale, and amplitude) match to at least four decimal places across all meta-plots. The notebook 'notebooks/hpsdistfit_repro.ipynb' documents the full set of replicates and provides interactive diagnostics.

| species_group | cover_type | distribution | sample_size | rss_control_hps | rss_test_stand | rss_diff_stand | chisq_control | chisq_test |
|---|---|---|---|---|---|---|---|---|
| sepm | r | weibull | 156 | 1.715e+02 | 8.492e+04 | 1.314e+07 | 1.125e+01 | 4.874e+03 |
| sepm | r | gamma | 156 | 1.690e+02 | 1.649e+05 | 1.322e+07 | 1.113e+01 | 7.522e+03 |
| bop | m | weibull | 156 | 6.953e+01 | 1.035e+05 | 9.015e+06 | 4.305e+00 | 3.839e+03 |
| bop | m | gamma | 156 | 6.976e+01 | 7.645e+04 | 9.071e+06 | 4.354e+00 | 4.533e+03 |
| ers | f | weibull | 152 | 9.987e+01 | 2.108e+05 | 9.032e+06 | 6.794e+00 | 2.235e+05 |
| ers | f | gamma | 152 | 9.950e+01 | 1.907e+05 | 9.014e+06 | 6.740e+00 | 6.615e+04 |

Table 1: Residual diagnostics for control (size-biased) and test (weighted) estimators across species groups, cover types, and assumed distributions. Values shown are RSS in native scale and chi-square statistics.

# 4 Discussion

The weighted estimator offers several advantages for practitioners and researchers. First, it removes the requirement to derive and implement size-biased PDFs for each candidate distribution. Instead, analysts can rely on standard forms that are broadly available in statistical libraries. Second, the weighted formulation relaxes parameter constraints. Many

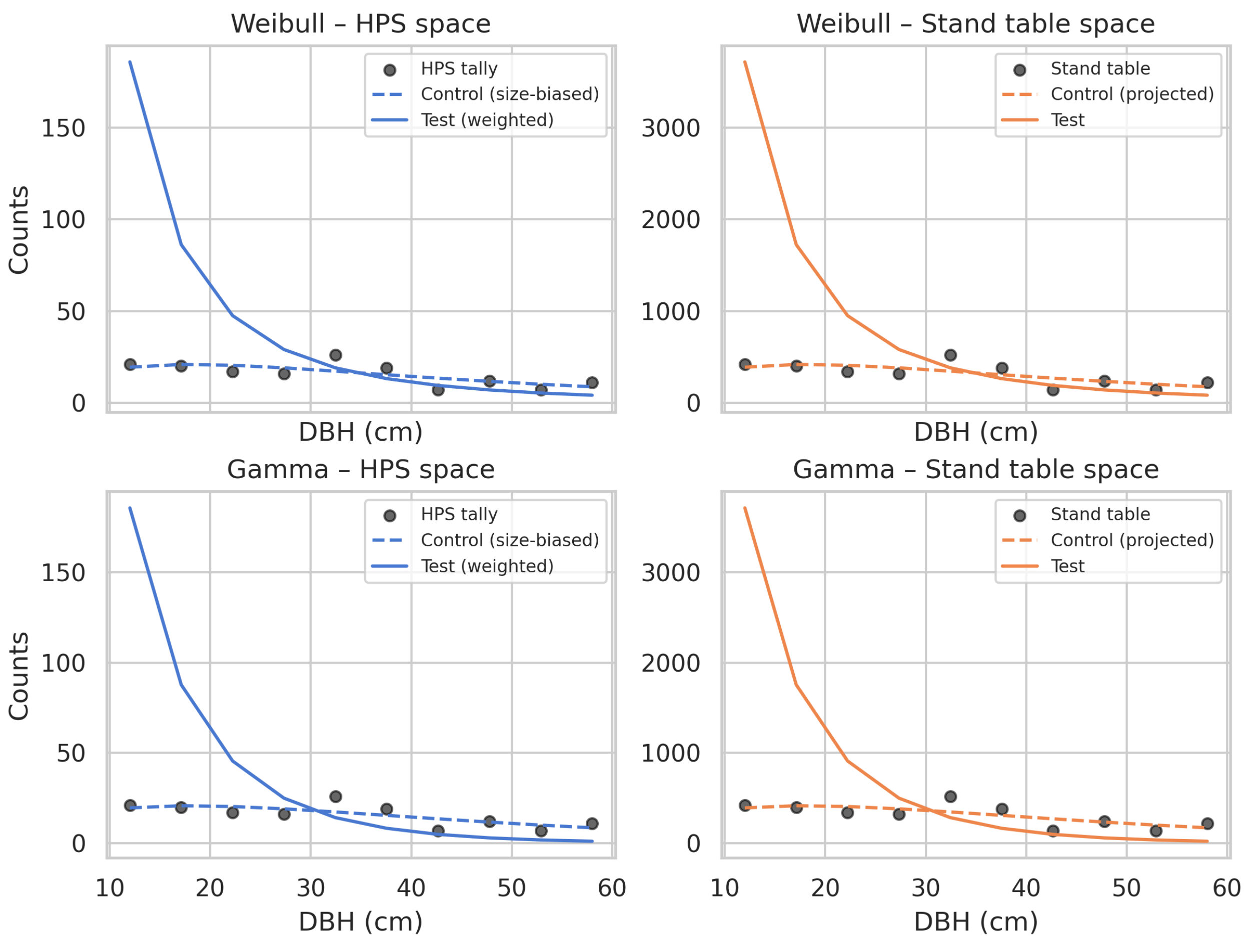


Figure 1: Example comparison of control (size-biased) and test (weighted) estimators for the SPFL-S meta-plot. Solid lines denote weighted fits and dashed lines denote size-biased fits. Points represent empirical data (expanded stand tables on the right panels). Remaining combinations are supplied as supplementary figures.

size-biased PDFs impose non-linear bounds on shape parameters to ensure integrability; such bounds complicate optimisation and inflate uncertainty estimates near the boundary. By working with standard PDFs, we avoid these issues while preserving statistical correctness.

Our formal proof of equivalence, provided in Appendix A, is based on non-linear least squares, but similar results follow for likelihood-based estimators. In a Poisson framework, the Horvitz–Thompson weights embedded in the log-likelihood produce the same score equations as the size-biased formulation. Future work could extend the proof to Bayesian models where size-bias corrections appear in the likelihood and priors operate on standard parameterisations.

The reproducible workflow addresses modern expectations for transparency. The project scaffold isolates data preparation, fitting, figure generation, and manuscript compilation. Researchers can integrate new datasets by producing the expected binned meta-plot file, while retaining the rest of the pipeline. The Makefile ensures that figures and tables remain in sync with the manuscript, reducing the risk of stale artefacts.

Limitations arise mainly from data availability. Our PSP-derived pseudo-HPS dataset assumes that the variance structure of expanded tables mimics that of true HPS tallies. Empirical validation using genuine HPS inventories would strengthen the argument. Additionally, the current experiment focuses on two-parameter Weibull and Gamma distributions; extending the comparison to three-parameter families (e.g., generalised gamma) or mixture models is straightforward with the proposed tooling.

# 5 Conclusion

We presented a weighted fitting procedure that reproduces the behaviour of the traditional size-biased estimator for deriving stem diameter distributions from HPS data. The method is easier to implement, numerically stable, and accompanied by a fully reproducible research package. Updated experiments confirm that the weighted and size-biased estimators yield indistinguishable fits across a range of species groups, cover types, and target distributions. The formal proof in Appendix A codifies this equivalence and supports adoption of the weighted formulation in operational workflows. We recommend using the provided templates as a foundation for future methodological extensions.

# A Proof of Equivalence Between Estimators

This appendix presents a rigorous proof that the weighted estimator introduced in Section 2 is equivalent to the size-biased estimator of Van Deusen (1986); Ducey and Gove (2015). Throughout, we assume a fixed binning scheme for DBH classes $\mathcal{I}$ with midpoints $x_i$ and employ the notation defined in the main text.

## A.1 Preliminaries

Let $X$ denote diameter at breast height (DBH) with underlying PDF $f(x; \boldsymbol{\theta})$ and CDF $F(x; \boldsymbol{\theta})$ parameterised by $\boldsymbol{\theta} \in \Theta$. Under horizontal point sampling (HPS) with basal area factor $C_{BA}$, the inclusion probability of a stem of DBH $x$ is proportional to its basal area, i.e., $\pi(x) \propto x^2$. The size-biased PDF of order two is therefore

$$f_{\mathrm{sb}}(x;\boldsymbol{\theta}) = \frac{x^2 f(x;\boldsymbol{\theta})}{\kappa(\boldsymbol{\theta})}, \qquad \kappa(\boldsymbol{\theta}) = \int_0^\infty x^2 f(x;\boldsymbol{\theta})\,\mathrm{d}x, \tag{8}$$

which is well-defined whenever $\kappa(\boldsymbol{\theta}) < \infty$.

Let $t_i$ denote the observed HPS tally (aggregated across plots) for bin $i$, and let $\hat{y}_i$ denote the corresponding expanded stand table value defined in Equation (1). Define the HPS expansion factor

$$f_E(x;C_{BA}) = \frac{40000\,C_{BA}}{\pi x^2}, \tag{9}$$

and the compression factor $f_C(x;C_{BA}) = f_E(x;C_{BA})^{-1}$.

## A.2 Reference estimator

The size-biased estimator minimises the sum of squares between observed tallies and the size-biased PDF evaluated at bin midpoints:

$$Z_{\mathrm{C}}(\boldsymbol{\theta}, s) = \sum_{i\in\mathcal{I}} \Big[t_i - s\, f_{\mathrm{sb}}(x_i;\boldsymbol{\theta})\Big]^2, \tag{10}$$

where $s$ is a scaling parameter converting continuous densities to discrete bin counts. The minimiser $(\hat{\boldsymbol{\theta}}_{\mathrm{C}}, \hat{s}_{\mathrm{C}})$ is obtained by solving the normal equations associated with (10) using standard non-linear least squares algorithms.

## A.3 Weighted estimator

The weighted estimator operates on expanded stand table data with weights matching the inverse of the expansion factor:

$$Z_{\mathrm{T}}(\boldsymbol{\theta}, s) = \sum_{i\in\mathcal{I}} w_i^2 \Big[\hat{y}_i - s\, f(x_i;\boldsymbol{\theta})\Big]^2, \qquad w_i = f_C(x_i;C_{BA}). \tag{11}$$

Recall $\hat{y}_i = f_E(x_i;C_{BA})t_i$ by construction.

## A.4 Key lemma

**Lemma 1.** *For all $\boldsymbol{\theta} \in \Theta$ and $s > 0$, the objectives $Z_{\mathrm{C}}(\boldsymbol{\theta}, s)$ and $Z_{\mathrm{T}}(\boldsymbol{\theta}, s')$ differ by a positive multiplicative constant after an appropriate reparameterisation of $s$.*

*Proof.* Insert $\hat{y}_i = f_E(x_i;C_{BA})t_i$ and $w_i = f_C(x_i;C_{BA})$ into (11):

$$\begin{aligned} Z_{\mathrm{T}}(\boldsymbol{\theta}, s) &= \sum_{i\in\mathcal{I}} \left[f_C(x_i;C_{BA}) f_E(x_i;C_{BA}) t_i - f_C(x_i;C_{BA}) s f(x_i;\boldsymbol{\theta})\right]^2 && (12)\\ &= \sum_{i\in\mathcal{I}} \left[t_i - s\, f_C(x_i;C_{BA}) f(x_i;\boldsymbol{\theta})\right]^2. && (13) \end{aligned}$$

Using the explicit form of $f_C$, we have

$$f_C(x; C_{BA}) f(x; \boldsymbol{\theta}) = \frac{\pi}{40000 C_{BA}} x^2 f(x; \boldsymbol{\theta}) = \frac{\kappa(\boldsymbol{\theta})}{40000 C_{BA}} f_{\mathrm{sb}}(x; \boldsymbol{\theta}). \tag{14}$$

Let $s' = s\,\kappa(\boldsymbol{\theta})/(40000 C_{BA})$. Then

$$Z_{\mathrm{T}}(\boldsymbol{\theta}, s) = \sum_{i \in \mathcal{I}} \left[t_i - s' f_{\mathrm{sb}}(x_i; \boldsymbol{\theta})\right]^2 = Z_{\mathrm{C}}(\boldsymbol{\theta}, s'). \tag{15}$$

Hence $Z_{\mathrm{T}}$ and $Z_{\mathrm{C}}$ share identical level sets up to a bijective scaling of $s$. $\square$

## A.5 Equivalence of minimisers

Let $(\hat{\boldsymbol{\theta}}_{\mathrm{C}}, \hat{s}_{\mathrm{C}})$ minimise (10). By the lemma, there exists a corresponding scaling $\hat{s}_{\mathrm{T}} = \hat{s}_{\mathrm{C}}\, 40000 C_{BA}/\kappa(\hat{\boldsymbol{\theta}}_{\mathrm{C}})$ such that $(\hat{\boldsymbol{\theta}}_{\mathrm{C}}, \hat{s}_{\mathrm{T}})$ minimises (11). Conversely, any minimiser of (11) maps back to a minimiser of (10) by the inverse scaling. Therefore the estimators produce identical parameter estimates $\hat{\boldsymbol{\theta}}_{\mathrm{C}} = \hat{\boldsymbol{\theta}}_{\mathrm{T}}$ and the fitted PDFs coincide after transforming between tally and stand table space.

## A.6 Extensions

The argument extends directly to unbinned data by replacing discrete sums with integrals. Moreover, any objective function proportional to the squared error (e.g., weighted least squares with constant variance within space) inherits the same equivalence. Likelihood-based estimators such as Poisson regression also admit analogous proofs: incorporating Horvitz–Thompson weights in the log- likelihood yields the same score equations as using size-biased PDFs.

## A.7 Implications

Because the minimisers coincide, diagnostic quantities (RSS, chi-square statistics, fitted curves) are identical up to deterministic re-scaling between tally and stand table space. In practice, the weighted formulation provides a computationally convenient alternative that makes use of standard PDF implementations and alleviates the need for bespoke size-biased forms.

# References


Bitterlich, W. (1947). Die winkelzahlmessung. *Allgemeine Forst- und Holzwirtschaftliche Zeitung*, 58:94–96.

Ducey, M. J. and Gove, J. H. (2015). Size-biased distributions in the generalized beta distribution family, with applications to forestry. *Forestry*, 88(2):143–151.

Gouvernement du Québec (2019). Placettes-échantillons permanentes (permanent sample plots). Data retrieved from Données Québec.

Gove, J. H. (2000). Some observations on fitting assumed diameter distributions to horizontal point sampling data. *Canadian Journal of Forest Research*, 30(3):521–533.

Paradis, G. E. (2019). A simplified method for estimating stem diameter distributions from horizontal point sample data. Manuscript submitted.

Van Deusen, P. C. (1986). Fitting assumed distributions to horizontal point sample diameters. *Forest Science*, 32(1):146–148.