Designs for subsampling tree cores for biomass estimation

John Tipton, Mevin Hooten

February 25, 2014

Introduction

Goals

- Estimate plot level biomass with increased precision
- Increase the number of large dbh trees in the sample
- Easily implementable in the field

Biomass Model

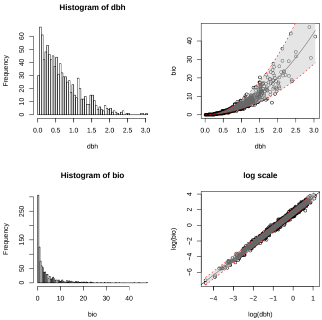
DBH is used as a predictor for biomass using an allometric equation. Under simple random sampling, the model of interest is the allometric tree growth equation with exponential error given by

$$y_i = x_i^{\beta_1} e^{\beta_0 + \epsilon} \quad (1)$$

where y_i is biomass for tree i, x_i is dbh for tree i, $\epsilon \sim N(0, \sigma^2)$ is a random error, and β_0 and β_1 are coefficients to be estimated.

The Model

A plot of the allometric relationship using simulated data

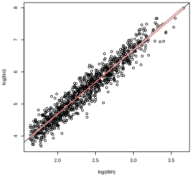


The Model

The allometric model is estimated by using ordinary least squares on the log-log model

$$\log(y_i) = \beta_0 + \beta_1 \log(x_i) + \epsilon \tag{2}$$

Figure : Plot of simulated log-allometric relationship typical of a sample plot (assumes one species only)



Questions of Interest

Main questions:

1. What is the effect of sampling design on the estimate of total biomass and its variance estimator
2. How to optimize the sampling design for the desired sampling goals of estimating the total plot biomass with high precision while selecting large trees

Horvitz-Thompson Estimation

For a finite population with elements indexed by $i = 1, \ldots, N$ with a probability π_i of being included in the sample s, the unbiased estimate of the population total is

$$\hat{t}_\pi = \sum_{k \in s} \frac{y_k}{\pi_k}.$$

where y_i is the biomass(or other variable of interest) for tree i.

Estimation of the variance of the population total is calculated as

$$\hat{V}(\hat{t}_\pi) = \sum_{k \in s} \sum_{l \in s} \frac{1}{\pi_{kl}} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) y_k y_l$$

- The log model can be estimated by using weighted least squares regression with the weight matrix $W = diag(\pi_1, \ldots, \pi_N)$
- Weighted least squares regression model estimate for β is $\hat{\beta} = (X^T WX)^{-1} X^T Wy$

Five different sampling schemes were considered

1. Simple random sampling (SRS)
2. Empirical cumulative distribution function (ECDF)
3. Probability proportional to size sampling (PPS)
4. Adaptive estimation of the ECDF (AECDF)
5. Stratified sampling with simple random sampling within each stratum (STSI)

Simple random sampling (SRS)

- Easy to implement in the field
- Fails to preferentially select large dbh trees
- Baseline for comparison of precision
- Example sampling protocol:
  - Count the number N of trees at the plot and sample n without replacement

Probability Proportional to Size (PPS)

- PPS sampling assigns sample inclusion probabilities based on a size variable known for all elements in the sample, in our case dbh
- PPS sampling assigns sample weights for element i of $\pi_i = \frac{dbh_i}{\sum_{i \in U} dbh_i}$
- Highly efficient. This scheme will increase precision more than any other method
- PPS sampling will preferentially sample larger dbh trees
- Example sampling procedure
  - Measure dbh and label all trees in the plot to allow re-visit. Decide which tree to core using a tablet or cellphone with probability of being sampled proportional to size.

Empirical Cumulative Distribution Function
Sampling ECDF

- ECDF sampling assigns sample inclusion probabilities based on the finite population empirical distribution function $F(\cdot)$ where $F(x_{(n)}) = \frac{n}{N}$ for $x_{(n)}$ the n^{th} order statistic
- ECDF provides a point of comparison of the AECDF method as the ECDF is a best case scenario of the AECDF method
- ECDF will preferentially sample larger trees
- Example sampling procedure:
    - Measure dbh and label all trees in the plot to allow re-visit. Decide which tree to core using a tablet or cellphone with probability of being sampled proportional to the ECDF of dbh.

Stratified Sampling (STSI)

- STSI groups similar elements of the population and thus increases the precision of estimates by reducing within group variances
- STSI sampling can be made to preferentially sample larger trees
- Example sampling procedure:
  - Decide on a criteria for separating trees into class sizes $h = 1, \ldots H$ (e.g. small, medium, large). At the plot estimate the number of trees in each class size N_h (this does not have to be exact, just close). Using a cellphone/tablet, randomly sample n_h trees from each class h.

Adaptive ECDF sampling (AECDF)

- AECDF asymptotically approaches the ECDF design
- AECDF preferentially samples larger dbh trees
- AECDF variance estimator will need to be derived mathematically (not currently solved to the best of my knowledge)
- Example sampling procedure:
  - Measure the dbh of the first tree at the plot. Sample a tree core for this first tree. At the second tree, measure dbh and sample according to up-down decision using smartphone or computer. Continue through the plot in this fashion. This design produces a random (but controllable) sample size.

Using the simulated data presented earlier, I repeatedly sampled $n = 10, 40$, and 100 trees from the simulated population of $N = 400$ trees using the proposed designs and calculated Monte Carlo bias and relative efficiencies for each design based on 10,000 replicates.

Results $n = 20$

	Bias	Relative Efficiency
SRS	-0.00	1.00
ECDF	-0.01	0.71
PPS	-0.04	0.13
AECDF	0.06	1.40
STSI	0.00	0.48

Results $n = 40$

	Bias	Relative Efficiency
SRS	-0.02	1.00
ECDF	-0.00	0.73
PPS	-0.07	0.14
AECDF	0.01	1.41
STSI	-0.00	0.37

Results $n = 100$

	Bias	Relative Efficiency
SRS	0.01	1.00
ECDF	0.00	0.54
PPS	-0.27	0.10
AECDF	0.04	1.45
STSI	0.00	0.55

Results for sampling for biomass

1. Relative efficiency is a measure of the expected decrease in variance for a given sample size and design relative to SRS. A relative efficiency of 0.25 implies the design of interest has $1/4$ of the variance of SRS design (equivalently the design requires $\sqrt{1/4} = 1/2$ the sample size to achieve the same efficiency).
2. Two designs perform well in the simulation, PPS, and STSI. Of these designs, STSI is the easiest to implement in field while PPS is more powerful. STSI can be tuned to sample the distribution of dbh (and hence biomass and age)
3. Fixed radius nested plot sampling is similar to STSI sampling and Bitterlich sampling is similar to PPS sampling.

Sampling for age (longer chronologies)

- The results for sampling older trees should be comparable to these designs sampling biomass, as long as the correlation between age and dbh is similar to the correlation between biomass and dbh. As this level of correlation decreases, the efficiency gains due to STSI and PPS sampling decrease as well
- Currently we are investigating a sub-sampling scheme using real data to show similar results for tree age as we have for biomass.