SYSTEMATIC SAMPLING OF SCANNING LIDAR SWATHS

A Thesis

by

WESLEY TYLER MARCELL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2009

Major Subject: Forestry

SYSTEMATIC SAMPLING OF SCANNING LIDAR SWATHS

A Thesis

by

WESLEY TYLER MARCELL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Marian Eriksson
	Sorin Pospescu
Committee Members,	Cristine Morgan
	Ross Nelson
Head of Department,	Steven Whisenant

December 2009

Major Subject: Forestry

ABSTRACT

Systematic Sampling of Scanning Lidar Swaths. (December 2009)

Wesley Tyler Marcell, B.S., Texas A&M University

Co-Chairs of Advisory Committee: Dr. Marian Eriksson
Dr. Sorin Popescu

Proof of concept lidar research has, to date, examined wall-to-wall models of forest ecosystems. While these studies have been important for verifying lidars efficacy for forest surveys, complete coverage is likely not the most cost effective means of using lidar as auxiliary data for operational surveys; sampling of some sort being the better alternative. This study examines the effectiveness of sampling with high point-density scanning lidar data and shows that systematic sampling is a better alternative to simple random sampling. It examines the bias and mean squared error of various estimators, and concludes that a linear-trend-based and especially an autocorrelation-assisted variance estimator perform better than the commonly used simple random sampling based-estimator when sampling is systematic.

DEDICATION

I dedicate my thesis to my parents and grandmother for all of their financial and moral support throughout my undergraduate and graduate career. After nine years of school I am sure they had their doubts, but they never gave up the hope that I would finally finish with my graduate degree. I love you all so much!

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

Page

1. INTRODUCTION

1.1 Common Lidar Sensors: Profiling Versus Scanning Lidar

Traditional methods for forest measurement can require hundreds of man-hours of fieldwork to estimate parameters such as stem count, stand basal area (cross-sectional area at 1.3-m above ground) and volume for inventories of relatively large areas. New technologies in remote sensing such as light detection and ranging (lidar) allow these estimates to be made over large areas in a short period of time, requiring fewer resources than ground-only surveys. Lidar systems use lasers to measure the distance between the sensor and a target surface. Measurements are made by recording the time it takes for a laser pulse to travel from the source to the target surface and back to the sensor, and then calculating distance based on the speed of light (Lefsky et al., 2002). There are two common types of lidar systems: profiling lidars (e.g., Nelson et al., 2003), which create one-dimensional height profiles along slices of the forest, and scanning lidars (e.g., Næsset, 2004), which use an array of sensors to scan relatively wide swaths of the terrain along flight lines.

Profiling lidars use sequential lidar pulses acquired along linear transects to create a one-dimensional height profile. Each laser pulse is recorded when it returns to the sensor. Each pulse can have multiple returns. The first return will give the height of the canopy

_____
This thesis follows the style of *Photogrammetric Engineering and Remote Sensing.*

and subsequent returns will represent understory or ground level readings. Strong secondary returns often indicate the presence of a dense understory or represent the ground below the canopy. A succession of laser pulse measurements can be used to develop a profile of a forest canopy and the underlying terrain along a flight line (Nelson et. al, 1984).

Profiling lidar data can be used to estimate certain forest measurements such as biomass and merchantable volume based on canopy height density measurements. However, they cannot be used to infer individual tree characteristics. To estimate individual tree characteristics such as tree count, average crown width, and average height, a scanning lidar system must be used.

Scanning lidar data can be used to make height and crown width measurements at the individual tree level while profiling lidar only provides height data along a small slice of the forest. Another benefit of scanning lidar is that a complete coverage of data can be obtained for the entire area of interest. Forest height is a crucial forest inventory attribute for calculating timber volume, site potential, and silvicultural treatment scheduling (Popescu and Wynne, 2004). Tree height, tree count, and crown width measurements can be used to estimate many forest parameters including: average tree diameter at breast height, biomass, and merchantable volume. Scanning lidar data, in conjunction with specialized software allows researchers to create a database of thousands, even millions, of individual tree measurements in a fraction of the time

required by traditional field survey techniques. Forest parameters can then be estimated from the database using common statistical techniques. The focus of this research will concentrate on scanning lidar systems.

1.2 The Need for Development of a Sampling Method

To date, most lidar research has focused on the development of supporting technologies and on proof-of-concept studies, and not on using the data to enhance operational inventories; most have used complete lidar coverage of the study areas. There is considerable interest in moving away from proof-of-concept studies and toward the use of lidar to enhance operational inventories. Andersen (2009), for example, used lidar collected over a 300 × 300-m area centered on Forest Inventory and Analysis (Bechtold and Patterson, 2005) plots and concluded that lidar may be useful for characterizing stand condition for operational inventories. Other examples will be noted below.

 It is unlikely that lidar will be used to the exclusion of traditional plot measurements, but rather to enhance and extend them. While complete coverage can be obtained of an entire area of interest, this comes at a cost. An advantage of, indeed the premise of, conducting a laser-based forest inventory on the other hand, is that reliable estimates of the variable(s) of interest and the quality of those estimates may be obtained by measuring only a portion of the area; that is, by sampling.

At the design-phase of an inventory, decisions include whether to stratify, the size and number of field plots to use, and whether auxiliary information is or can be made available and how it might be used to improve the quality of any final estimate, as measured by its variance. These decisions are made in an effort to ensure unbiased or nearly unbiased estimators and to reduce variance for a given cost, or to minimize cost for a specified maximum variance. Technical details, such as flying-height and scanner settings, aside, the questions surrounding the operational use of lidar will include how it will fit into the overall design and how its use will influence the choice of estimators.

Lidar measurements are proxy measurements and they are known to be imperfect due, for example, to (*i*) the fact that lidar pulses rarely intercept tree tips, (*ii*) current algorithms used to extract crown dimensions do not adequately model overlapping crowns, (*iii*) GPS error, and (*iv*) the fact that intermediate and suppressed trees are undercounted. While imperfect, lidar measurements are also known to be "good" in the sense that correlations between the proxy measures and ground-based measurements are generally high; indeed, there are indications (e.g., Næsset, 2004) that the accuracy of the lidar measurements may, on average, actually be better than those made on the ground.

It is therefore probable that lidar will be used in an auxiliary sense—quite likely as one of the phases in a multiphase design. To date, examples of the use of lidar in operational double (2-phase) sampling designs include, Næsset (2004), Parker and Evans (2004), Parker and Mitchel (2005), and Andersen and Breidenbach (2007). A written review of

double- and single-phase lidar sampling can be found in Parker and Evans (2009). If a sample of flight lines is to be flown, then flight plan considerations argue that lines at regular intervals would operationally be more desirable than lines taken at irregular intervals. This is an example of *systematic sampling* (e.g., Cochran, 1977). All three of the double sampling designs cited earlier had a systematic component.

1.3 Early Studies Using Lidar

The first natural resource application of lidar was a 1968 study of bathymetry or oceanic depth (Hickman and Hogg, 1969). In that study an aerial profiling laser was used to make bathymetric measurements in shallow offshore areas. That method was found to be an effective means of collecting bathymetric measurements because of the lasers' ability to penetrate the surface of the water and record sea floor depth. The same principles that allowed bathymetry researchers to measure ocean floor depth were soon applied to terrestrial applications.

One of the first studies involving terrestrial application of lidar was that of Krabill et al. (1980) which was mentioned in Nelson et al. (1984). Studies were conducted to utilize the ability of the airborne laser system to penetrate vegetation and record ground measurements. In those original studies vegetation was seen as a source of noise when creating terrain models. Soon researchers realized that this noise could be used to determine the canopy height of vegetation. Nelson et al. (1984) used a profiling lidar

system to find tree heights in an oak-hickory forest in south-central Pennsylvania. He found that mean tree height estimates were underestimated by 60 cm when compared to data derived using photogrammetric methods. However, his study found lidar estimates to be more precise than photogrammetric estimates.

Further studies were conducted to estimate plant biomass and to determine the repeatability of lidar observations. Nelson et al (1988) found a 7% and 8% difference between laser derived biomass and volume estimates, respectively, when compared to ground measurements. They concluded their study by stating that there are two advantages to estimating forest parameters using profiling lidar. The first advantage is that canopy height data can be collected quickly along transects hundreds of miles long. The second advantage is that lidar can be used to sample areas that may not be easily accessible using traditional ground inventory methods.

1.4 Scanning Lidar Research

Advances in technology have allowed lidar to become a more powerful tool than ever. Newly developed scanning lidar systems are capable of full data coverage with up to twenty points per square meter (Ackermann, 1999). These new systems in combination with more powerful GIS and remote sensing software have allowed lidar research to go far beyond the profiling studies of the 1980's.

In Norway remote sensing techniques are an integral part of forest surveys and there are ongoing studies involving scanning lidar. Forest characteristics for approximately 50% of all area surveyed annually is derived using aerial photo interpretation. In one of the first studies involving scanning lidar, Næsset (1997a, b) was able to estimate mean tree height and timber volume effectively using only the data provided from the scanning lidar system.

Further studies in Norway have looked at complete coverage datasets with relatively low intensity scanning lidar (Næsset, 2004). The dataset used in that study had a sampling rate of about one pulse per square meter. This is not intense enough to make measurements on the individual tree level; however, Næsset's study did provide very good stand-level forest estimates.

1.5 Review of Systematic Sampling

Figure 1 depicts a systematic arrangement for a finite population of size $N = 12$ with $n = 3$ and $k = 4$, where n is the number of sample units to be measured and k is the number of possible samples, only one of the possible samples would actually be selected; indices h, i, and i', run from 1 to k, n, and N, respectively. That is, randomization would be on the k possible samples, so one and only one h would be randomly selected from the $k = 4$ possible samples. Systematic sample $h = 3$ is highlighted to indicate the $n = 3$ units, swaths for us, that would be measured were $h = 3$ to be selected. In the systematic

sampling literature, units with common index *i* are sometimes referred to as a stratum. For us, this is consistent with the land managers' idea of strata being regions of relatively homogenous cover. For simplicity we assume throughout that *N* is an integer multiple of *k*. This amounts to assuming that *n* is fixed. We will sometimes refer to the different *h*'s as "levels."

The degrees of freedom for an estimate is basically the number of times we randomize minus the number of parameters estimated. With classical "single start," systematic sampling (SyS) one randomizes once resulting in zero degrees of freedom. This means that for SyS no statement can be made, with any confidence, about the quality of estimates. Yet SyS is one of the most commonly implemented sample designs due to the ease with which samples can be selected (eg., Cochran, 1977). Intuitively SyS is also "good" for land-based applications because a SyS can be viewed as being more "even" than a simple random sample (SRS) of the same population. Statistically this is reflected by the fact that the true variance of the mean of a SyS is often less than the true variance of the mean of a SRS of equal size from the same population (Cochran, 1977). The problem is that there is no unbiased design-based estimator of the true SyS variance.
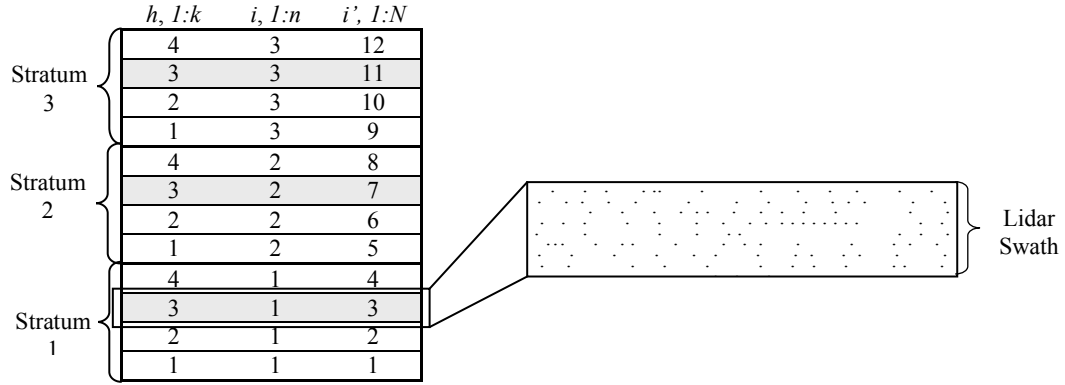
Figure 1. *Schematic diagram depicting systematic sampling. N=12 is the total number of units in the population, n = 3 is the sample size, and k=4 is the number of possible systematic samples; i', i and h, respectively, are indices over these values.*

Let y_{hi} be i^{th} of the n values of the variable Y observed in the hth of the k possible samples, and let $\bar{y}_h = \sum_{i=1}^{n} y_{hi} / n$ be its mean. Further, let $\bar{\bar{Y}} = (1/nk)\sum_{h=1}^{k}\sum_{i=1}^{n} y_{hi}$ be the mean of all $N = nk$ units in the population. Then

$$V = \mathrm{Var}(\bar{y}_{sys}) = \frac{1}{k}\sum_{h=1}^{k}(\bar{y}_h - \bar{\bar{Y}})^2 \tag{1}$$

is the true SyS variance of the mean for systematic samples of size n from a population of size $N = nk$. Notice that the summation is taken over all k of the possible systematic samples, only one of which is observed. Though biased, often the standard finite population variance estimator,

$$s^2_{\bar{y}_h,srs} = v_1 = \frac{(1-f)}{n} \sum_{i=1}^{n} \frac{(y_{hi} - \bar{y}_h)^2}{n-1} = (1-f)\frac{s^2_h}{n}, \qquad (2)$$

is used. Here $f = n / N$ is the sampling fraction, $(1 - f)$ is the finite population

correction, $s^2_h = \sum_{i=1}^{n} (y_{hi} - \bar{y}_h)^2 / (n - 1)$ is the usual estimator for the variance among all

N population units based on the n units in systematic sample number h, and division by

n, of course, because we are considering variances of means. The $v_\#$ notation is Wolter's

notation for his numbered estimators and is included because diagrams in the appendices

use this shorter notation for simplicity. The longer $s^2_{\bar{y}_h,xyz}$ notation is used as a mnemonic

device indicating the origin of the estimator.

We refer to equation (2) as the SRS-based estimator because it is appropriate for simple

random samples of size n selected without replacement from a population of known size

N, is used. If the design is in fact, SyS, then (2) is a conservative estimator, and is

sometimes justified on those grounds—in repeated sampling confidence intervals would

be too wide; they would contain the true population mean more frequently than the

stated level of confidence (p-value) would indicate. The bias of $s^2_{\bar{y}_h,srs}$ under classical SyS

arises from fact that SyS induces within-level (h) and between-level components of

variance and that the between-level component is unmeasured (Cochran, 1977).

The SRS-based estimator (2) is unbiased if the order of units in the population is random

with respect to the variable under consideration. For example, if persons' height is the

variable of interest and the systematic sample is selected by choosing the first person on

every fifth page in a phone book, then (2) should be approximately unbiased. The deviations from random that have received the most attention are when there is (*i*) autocorrelation, in our case spatial autocorrelation, in the variable of interest, (*ii*) when the "evenness" alluded to above can be considered as being associated with strata, as defined above, and (*iii*) a linear trend between the variable of interest and the index *i'* ordering the population. Case (*i*), and to a lesser extent (*ii*) and (*iii*) are expected for landscape variables. Numerous alternatives to (2) have been proposed to reduce bias in estimating (1) if population units are arranged according to (*i*)-(*iii*).

The estimator Cochran (1977) uses for stratification effects is based on successive differences. It is given by

$$s^2_{\bar{y}_h,se} = v_2 = \frac{(1-f)}{n}\sum_{i=1}^{n}\frac{a^2_{hi}}{2(n-1)} \tag{3}$$

where the $a_{hi} = y_{hi} - y_{h,i-1}$ are the successive differences between values separated by k units.

Equation (3) is used when the mean is consistent within each stratum of k units. It is a biased estimator. Cochran (1977) states that $s^2_{\bar{y}_h,se}$ contains an unwanted contribution for the difference of neighboring strata's means which causes the first and last strata of the model to carry too much weight in estimating the random component of variance. In a reasonably large sample the estimate of $s^2_{\bar{y}_h,se}$ will generally be too high (Moore, 1955 and Meyer, 1956). Another stratification based estimator is

$$s^2_{\bar{y}_h,st} = v_3 = \frac{(1-f)}{n} \sum_{i=1}^{n/2} \frac{a^2_{h,2i}}{n} \tag{4}$$

where the $a_{h,2i}$ are, again, successive differences of units, but now treating the strata as successively grouped in pairs with two units per pair. That is, it pools the data from each pair of strata. Estimator (4) is based on fewer degrees of freedom than $s^2_{\bar{y}_h,se}$.

The next trio of estimators are due to Yates (1949) and each is based on successive differences, and are similar in concept to (3), but with end-corrections intended to equalize the influence of all units regardless of position in the systematic sequence of observations. According to Wolter (1984), "v_4 is based upon second differences, which annihilate a linear trend in the population." It can be derived by assuming that the $y_{i'} = \beta_0 + \beta_1 i' + \varepsilon_{i'}$ where, again, i' represents ordering 1, 2, ... N, in the population. Estimators v_5 and v_6 generalize the concept to higher-order differences and, in turn, higher-order trends. The estimators are:

$$s^2_{\bar{y}_h,lt1} = v_4 = \frac{(1-f)}{n} \sum_{i=3}^{n} \frac{b^2_{hi}}{1.5(n-2)}, \tag{5}$$

$$s^2_{\bar{y}_h,lt2} = v_5 = \frac{(1-f)}{n} \sum_{i=5}^{n} \frac{c^2_{hi}}{3.5(n-4)}, \tag{6}$$

$$s^2_{\bar{y}_h,lt3} = v_6 = \frac{(1-f)}{n} \sum_{i=9}^{n} \frac{d^2_{hi}}{7.5(n-8)}, \tag{7}$$

where

$$b_{hi} = \frac{y_{hi}}{2} - y_{h,i-1} + \frac{y_{h,i-2}}{2},$$

$$c_{hi} = \frac{y_{hi}}{2} - y_{h,i-1} + y_{h,i-2} - y_{h,i-3} + \frac{y_{h,i-4}}{2},$$

$$d_{hi} = \frac{y_{hi}}{2} - y_{h,i-1} + y_{h,i-2} - y_{h,i-3} + y_{h,i-4} - \ldots + \frac{y_{h,i-8}}{2},$$

respectively. Note the alternating plus and minus components in these equations. Cochran (1977) states that for more complex populations containing continuous variation estimators based off the quadratic formula often provide better results than those based off successive differences such as estimator (3).

We did not consider Wolter's estimator v_7, due to Koop (1971), as Wolter found it to be "unpredictable, and its variance is generally too large to be useful. This estimator cannot be recommended…" (p. 790), and we anticipated that autocorrelation-based estimators would be better behaved.

Our final two estimators are model-assisted, being based on ("assisted by" in the terminology of Särndal, et. al, 1992) a "superpopulation model." That is on the assumption that actual, observed, population is a realization of a process for which $y_{hi} = \mu + \varepsilon_{hi}$, where the errors are assumed to be autocorrelated, in our case spatially autocorrelated, in the direction normal to the flight lines. The estimators are due to Cochran (1946) and make the further assumption that the autocorrelation is of the exponential form, $\rho_k = e^{-\lambda k}$, where k is the separation ("lag") distance between two successive systematic observations.

The first of the two autocorrelation-based estimators is

$$s^2_{\bar{y}_h,\rho1} = v_8 = \begin{cases} s^2_{\bar{y}_h,srs}\left(1+\dfrac{2}{\ln(\hat{\rho}_{hk})}+\dfrac{2\hat{\rho}_{hk}}{1-\hat{\rho}_{hk}}\right) & \text{if } \hat{\rho}_{hk} > 0 \\ \\ s^2_{\bar{y}_h,srs} & \text{if } \hat{\rho}_{hk} \leq 0, \end{cases} \tag{8}$$

where the correlation is estimated using

$$\hat{\rho}_{hk} = \frac{\sum_{i=1}^{n-1}(y_{hi}-\bar{y}_h)(y_{h,i+1}-\bar{y}_h)}{(n-1)s_h}.$$

It is a large n and large k approximation to Cochran's theoretical

$$s^2_{\bar{y}_h,\rho2} = c_{47} = \begin{cases} s^2_{\bar{y}_h,srs}\left(1-\dfrac{N-1}{k-1}g(N,\lambda)+\dfrac{k(n-1)}{k-1}g(n,k\lambda)\right) & \text{if } \hat{\rho}_k > 0, \\ \\ s^2_{\bar{y}_h,srs} & \text{if } \hat{\rho}_k \leq 0, \end{cases} \tag{9}$$

where $\lambda = \log(\hat{\rho}_k)$ and

$$g(N,\lambda) = \frac{2}{N(N-1)}\left(\frac{(N-1)e^\lambda - N + e^{-(N-1)\lambda}}{(e^\lambda-1)^2}\right).$$

Estimator (9) is, in theory, applicable for any n and k,

Equation (8) has been mentioned by other authors but has not seen widespread use, especially in forestry and remote sensing literature. We have never seen (9) applied to real data, presumably because of its seemingly more complex form, but with today's computing power, there is no reason not to consider it as well. Wolter (1984) did not

refer to this estimator in his paper so we refer to it as estimator c_{47} in the appendices—it is Cochran's (1946) equation #47.

Wolter (1984) considered estimators (2–8) for fifteen populations (seven simulated and eight real). We used all eight of the estimators (2–9) on our simulated lidar forest. On theoretical grounds, Wolter determined that $s^2_{\bar{y}_h,\rho1}$ tends to underestimate variance whereas (2–7) tend to overestimate and that $s^2_{\bar{y}_h,\rho1}$ "tends to have the smallest absolute bias except when ρ is small. When ρ is small, the $\ln(\rho^k)$ approximation is evidently not very satisfactory" (p.786). For forestry applications we expect "larger" correlations because the conditions at one location tend to be similar to conditions at nearby locations. After analyzing the fifteen populations, he concluded that $s^2_{\bar{y}_h,srs}$ (equation 2) has reasonably small bias and mean squared error (MSE = bias^2 + variance) only when the populations have no trend, autocorrelation, or stratification effects. He also concluded that $s^2_{\bar{y}_h,lt1}$ is superior to the two similarly motivated trend-based estimators ($s^2_{\bar{y}_h,lt2}$ and $s^2_{\bar{y}_h,lt3}$). He wasn't particularly impressed by the performance of $s^2_{\bar{y}_h,\rho1}$:

> I like *v2*, *v3* and possibly *v4* (p. 789) … Estimator *v8* has remarkably good properties for the artificial populations with linear trend or autocorrelation; otherwise it is quite mediocre … This estimator [*v8*] seems too sensitive to the form of the model to be broadly useful in real applications. (p. 790).

By "model" he was referring to the assumed superpopulation model. In his concluding paragraph he did, however, note that his findings were primarily applicable to surveys of establishments and people and that stronger autocorrelation patterns are likely to exist in forestry applications.

1.6 Purpose of Study

We were asked the question: is sampling, and in particular is systematic sampling, a reasonable approach when using scanning lidar? It is known that the mean of a systematic sample is unbiased, and we have seen that systematic flight lines are preferred over random flight lines, so the question becomes, would variance estimates for systematically sampled data be reasonable? The first step of this thesis is to simulate a lidar forest to address these questions.

The mean of a SyS is known to be more precise than the mean of an SRS of the same population when the true variance among units within systematic sample is greater than the variance among all units in the population (Cochran, 1977). The first objective of this thesis, then, is to determine whether, on average, over combinations of n and k for the simulated lidar forest, the variance among units within systematic samples is greater than the variance among all units. The second objective of this thesis is to compare the relative performance of estimators (2-9) over combinations of n and k for the simulated lidar forest. Thus, the overall goal of this study is to analyze sampling strategies using a simulated lidar-forest.

2. METHODOLOGY

2.1 Study Area

A 4800-ha rectangular forested area in eastern Texas was the study area for this project. This area contains stands of various type, size, and density classes. The composition and structure of the study area is similar to that of many forests in the southern United States. Complete scanning lidar coverage of the area was obtained in February, 2004. Sixty two randomly located ground plots within the area were measured in May and June of that year. Other studies using these data include Popescu (2007), Mutlu et al. (2008), and Zhao et al. (2009).

2.2 Creation of the Lidar-Forest

This study is really a study of the lidar-forest associated with the actual forest described above. The lidar-forest is known to deviate from the actual forest due to a variety of errors (eg., GPS location error, measurement error, etc.). The sum of the errors is generally considered to be small (Næsset, 2004). These lidar-derived errors are not an issue for the study, because we are focusing on the sampling error of the lidar-forest— our population of interest. The lidar-forest was created from the raw pulse returns by following the steps as described by Popescu and Wynne (2004).

First, a canopy height model (CHM) was developed from the lidar point cloud. A canopy height model is the difference between tree canopy hits and the corresponding lidar-derived terrain elevation values. A more detailed description of the CHM development is available in Zhao et al. (2009). In the second step, a leaf-off Quickbird image from 2004 was classified to differentiate between pine and hardwood stands and non-forested regions within the study area. The classified image was combined with the original canopy height model to create two separate CHM's; one containing only data for areas classified as pine stands, the other containing only data for areas classified as hardwood stands. This step was necessary to correctly quantify individual trees using TreeVaW, an extension in the Interactive Data Language (IDL) developed by Popescu and Wynne (2004). TreeVaW automatically assigns x-y coordinates (½-m) to individual trees and uses local max filtering and variable windows to determine each tree's height, H, and crown width, CW.

Relationships between CW and H, were developed using the ground data for pine and hardwoods separately. The fitted equations were $CW_P = 0.0024H^2 + 0.1848H + 0.4022$ ($r^2 = 0.66$) and $CW_H = 0.0031H^2 + 0.2076H + 1.6416$ ($r^2 = 0.45$) for pines and hardwoods, respectively. With the two CHMs, these equations are used by TreeVaW to help determine CW by species. Other inputs required by TreeVaW were the minimum expected tree height and the minimum and maximum expected crown width. The minimum expected height was taken to be 7.6-m, which is approximately the minimum height of a merchantable tree given the range of site indices for this region of Texas.

This setting allows TreeVaW to ignore lidar returns below the threshold on the assumption that it was not from the crown of a merchantable tree. The minimum and maximum expected crown widths were set at 1.5-m and 26-m, respectively; these settings control the search windows used to detect individual trees. Since TreeVaW assigns trees to a ½-m × ½-m grid, below we refer to trees having the same y (or x) coordinate as a *TreeVaW-line* or, more simply, a *line*, sometimes qualified by direction.

Some estimated heights were unreasonable—likely due to birds and towers, etc. These points were removed from the dataset by setting a ceiling value for tree height of 47-m. In addition, a number of lidar-trees had crown widths of zero. For these we used the crown width regression equations, above, to replace the zero values. Random variation consistent with the prediction error variance of the equations was added to the predicted values. This is acceptable for our study because once again we are not analyzing the accuracy of lidar, nor of TreeVaW, but rather the ability to correctly and efficiently estimate our lidar-forest parameters using systematic sampling.

2.3 Simulation Study

With scanning lidar, the operator has the ability to adjust (*i*) the scan angle of the sensor and (*ii*) altitude of the aircraft, and (*iii*) the pulse density to affect the ultimate quality of the lidar-forest. The idea behind the simulation study was that by adjusting these settings one could, within limits, obtain actual swath data consistent with our observed

data for different assumed swath widths. That is, we had the ability to treat groups of TreeVaW-lines normal to any assumed flight direction as a "virtual swath." Particulars for any real application will, of course, deviate from those presented herein, but the overall assessment concerning the efficacy of SyS at a lidar phase of sampling and of the overall assessment of estimator usefulness should apply to many real applications. For the simulations we assumed E-W and N-S flight lines. The area was clipped from the original 16923 by 11210 lines to the central 16800 (running from south to north, oriented west to east) by 10800 lines because (*i*) the number of lidar-trees in the outermost few lines was considerably less than for the rest of the area, and (*ii*) for simplicity we chose total line numbers that are even multiples of a few pre-selected sample sizes.

The width of swaths obtainable from current-day sensors on aircraft flying at common altitudes is on the order of about 200 to 800 meters. This would argue that the simulations should try to mimic swaths consistent with this range or, perhaps, a little larger, say 100 to 1000 meters wide. Since we had the dual interest in examining the properties of the SyS estimators from a purely statistical viewpoint, we looked at "swaths" well beyond this range. For extreme widths the idea of angular tolerances, pulse densities, and flying heights consistent with the simulated swaths becomes absurd, but if the observed lidar-forest is a reasonable depiction of reality, this does not diminish the value of looking at the properties of the estimators at the extreme values.

If we let ℓ_T be the total number of TreeVaW-lines (16800 E-W or 10800 N-S) and ℓ_S be the number of lines per swath then we have a population of size $N = \ell_T / \ell_S$. For example, with our ½-m grid, there are $N = 24$ 350-m E-W swaths when $\ell_S = 700$. We generated all even N from 4 to $\ell_T / 4$. From each of the populations we generated all k systematic samples of integer sizes $n = 2$ to $N / 2$. For $N = 24$ these were $n = 2, 3, 4, 6, 8$, and 12 with $k = 12, 8, 6, 4, 3$, and 2. This resulted in 633 (E-W swaths) and 459 (N-S swaths) combinations of n and N. For each combination, the mean number of trees (TPH), basal area (BPH), and stem biomass per hectare (SWH) were computed; the DBH and biomass prediction equations were those of Popescu (2007); for a total of $3(633+459) = 3276$ simulations. For each of the three variables, variances were computed using each of the eight estimators (2-9).

3. RESULTS AND DISCUSSION

The mean of a SyS is known to be more precise than the mean of an SRS of the same population when the true variance among units within systematic sample is greater than the variance among all units (Cochran, 1977). This was the case for more than 95% of the 3276 simulations examined in this. For all of the exceptions, that is for all cases for which the variance among all units was greater than the variance within systematic samples, the simulated sample size was n=2 or n=3. These are almost degenerate cases for which one would not expect SyS to be superior to SRS.

The number of pines per E-W line is presented in Figure 2a. There is a clear long-range linear, tending towards quadratic, trend in the number of pines per line with higher numbers in the south. Had the study area been larger, the apparent trend would diminish, but it is very real *for this* finite population. Corresponding trends for E-W hardwoods and for N-S pines and N-S hardwoods were generally cubic to quartic in nature. The trends for basal area per line and stem wood biomass per line generally followed those for numbers of trees but were somewhat muted and had somewhat larger short-range variation (Appendix A). The presence of linear trend in the E-W swaths indicated that $s^2_{\bar{y}_h, lt1}$ should perform favorably for numbers of trees, and to a lesser extent for basal area per hectare, and stem wood biomass for swaths in this direction.
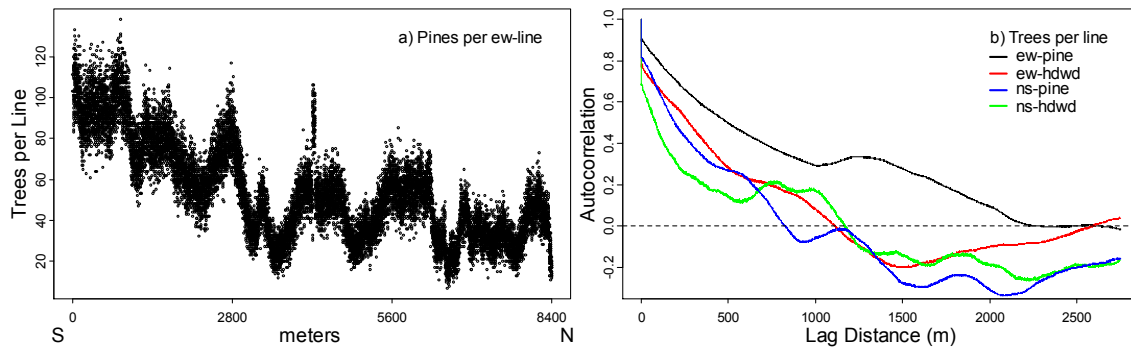
Figure 2. *Trees per line and autocorrelation plots*. a) Number of trees per E-W-line plotted on distance from the southern boundary, b) empirical autocorrelation functions to 2500 meters for the number of trees per line.

All variables indicated a high degree of autocorrelation on a per-line basis. Figure 2b presents the empirical autocorrelation functions for numbers of trees per line for E-W pine, E-W hardwood, N-S pine, and N-S hardwood. Though somewhat muted, the basal area and stem wood biomass trends closely follow the number of trees trends by species group and line direction (Appendix B). For numbers of trees, basal area per line, and stem wood biomass, autocorrelation becomes insignificant at lag distances beyond about 2250 m for E-W pine and about 1200 m for the others; in the geostatistics literature this distance is referred to as the *range*. When summed over lines, then, we would expect that $s^2_{\bar{y}_h,\rho1}$ and $s^2_{\bar{y}_h,\rho2}$ should perform favorably when $k\ell_S$ is less than these values. At 1200-m (2400 E-W lines), for any n greater than about 16800/2400 = 7, autocorrelation is likely to be an issue for E-W hardwood, N-S hardwood, and N-S pine.

It is worth noting that, in theory, all of the autocorrelation functions should be continuous at the origin but all show a discontinuity. In geostatistics, this discontinuity is called a *nugget* and is usually attributed to measurement and micro-scale variance (e.g., Schabenberger and Pierce, 2002). For these data, the discontinuities would indicate measurement error variance for numbers of trees per line of about 10% of the total variance for E-W pine, and of about 20% for E-W hardwood, N-S pine, and N-S hardwood. Tree diameters were predicted from the crown diameter and height estimates, themselves predicted by TreeVaW as indicated in the Methodology section. For basal area and stem wood biomass per line (Appendix B) the measurement plus prediction error variances contributed to about 25% and 50-60% of the total, for E-W pine and the others, respectively.

The upper set of panels in Figure 3 are bias plots of the estimated variances, $s^2_{\bar{y}_h, srs}$, $s^2_{\bar{y}_h, lt1}$, and $s^2_{\bar{y}_h, \rho 2}$, on the true variance V, as calculated from (1) for TPH using E-W swaths. Systematic deviations from the 1:1 line indicate bias. The panels clearly show the bias incurred by using the usual SRS-based estimator when samples are, in fact, selected systematically. It is often stated that the bias of the $s^2_{\bar{y}_h, srs}$ is unpredictable. Color was added for cases of $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink). In each case, increasing N is depicted by darker tones. It is clear from the plots that bias is largest when n is small. The plotted values are means over the k possible samples for each combination of n and N. It is clear from the plots that there is a generally smooth

relationship of variance with N given n. Plots are available for all combinations of estimators, line direction, and species class in Appendix C. The ρ-assisted estimators $s_{\bar{y}h,\rho1}^2$ and $s_{\bar{y}h,\rho2}^2$ are also significantly biased for small n but become largely unbiased as n increases beyond about $n=12$. The poor performance of $s_{\bar{y}h,\rho1}^2$ for small samples is likely attributable to one or both of (*i*) very small sample sizes would indicate samples separated by distances beyond the range and (*ii*) autocorrelation is very poorly estimated for such small sample sizes. Indeed, the algorithm suggests setting $s_{\bar{y}h,\rho1}^2$ and $s_{\bar{y}h,\rho2}^2$ to $s_{\bar{y}h,srs}^2$ when negative correlation estimates are computed. Wolter (1984) commented on this as well (see the quote on page 15 of the introduction section of the thesis).

Given the large variance ranges, bias is perhaps better viewed in the lower panels of Figure 3 in which the $\sqrt[4]{s_{\bar{y}h,srs,lt1,\rho1}^2}$ are plotted on $\sqrt[4]{V}$. The fourth-root transformations both linearized the trends and induced common spread. In these plots we clearly see the relative bias in $s_{\bar{y}h,srs}^2$ across all n and N. We also see that, for most combinations of n and N, $s_{\bar{y}h,lt1}^2$, and $s_{\bar{y}h,\rho1}^2$ are nearly unbiased for these data. Bias plots for all combinations of TPH, BPH, and SWH, pines & hardwoods, and E-W & N-S swaths were qualitatively quite similar and are available in Appendix C and D. For example on average $s_{\bar{y}h,srs}^2$ for TPH was 4.84, 5.98, and 9.08 times that of $s_{\bar{y}h,\rho1}^2$ for pine with samples of size $n = 6$, 8, and 12 for E-W 350-m swaths ($N = 24$); the corresponding values for hardwoods were 1.00, 1.78, and 5.32, respectively. The implications for selecting sample sizes to achieve stated precision goals may be appreciable.
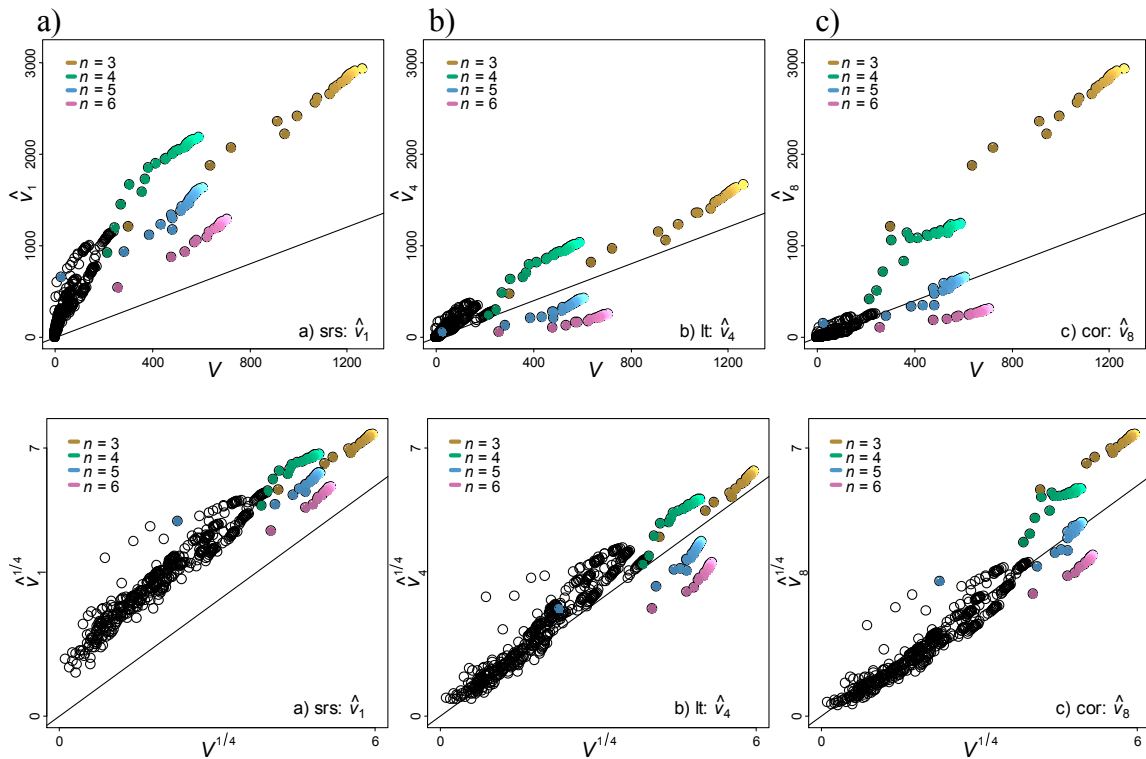
Figure 3. *Estimated variance plotted on true variance.* Estimated variance plotted on true SyS variance for 633 combinations of *N,* and *n* for number of pines per hectare on E-W-swaths using estimators a) $s^2_{\bar{y}_h,srs}$, b) $s^2_{\bar{y}_h,lt1}$, c) $s^2_{\bar{y}_h,\rho1}$. In the upper panels variances are untransformed, in the lower panels they are transformed by $v^{1/4}$.

Figure 4 shows the $\text{MSE}^{1/4}$ of $s^2_{\bar{y}_h,srs}$ and $s^2_{\bar{y}_h,lt1}$ ($n \geq 12$) plotted on $\text{MSE}^{1/4}$ of $s^2_{\bar{y}_h,\rho1}$ for pine trees per hectare on E-W and N-S swaths, panels (a) and (b), and pine basal area per hectare on N-S swaths, panel (c). The scatter of points for $\text{MSE}(s^2_{\bar{y}_h,srs})$ is well above the 1:1 line, again displaying the large bias in $s^2_{\bar{y}_h,srs}$ when samples are systematically selected. Moreover, for the *n* = 6, 8, and 12 example of the previous paragraph, the

average MSE for $s^2_{\bar{y}_h,srs}$ was 2.5, 41, and 93 times that of $s^2_{\bar{y}_h,\rho 1}$, indicating that, not only does $s^2_{\bar{y}_h,srs}$ overestimate the true variance, but that it also more variable than $s^2_{\bar{y}_h,\rho 1}$ for our lidar-forest attributes. For n less than about eight, scatter for both $s^2_{\bar{y}_h,srs}$ and $s^2_{\bar{y}_h,lt1}$ tend to the 1:1 line, again reflecting the poor estimation of ρ based on few data values and the large lag distances for small n (not shown in the diagrams).
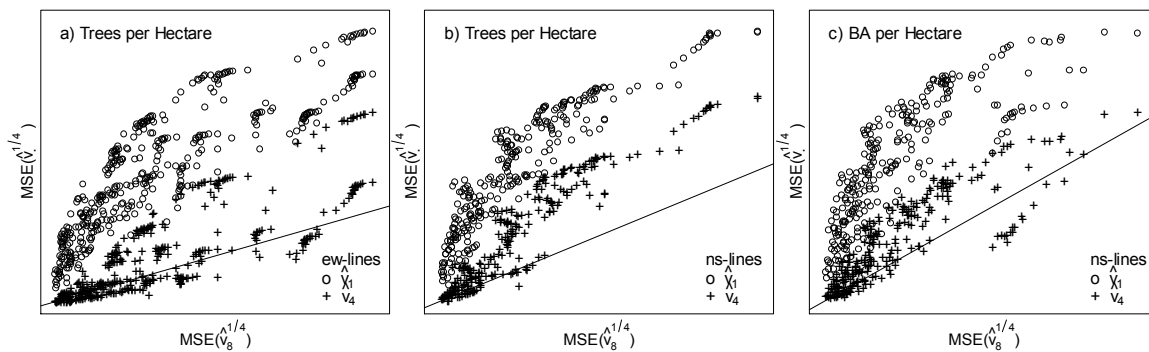

Figure 4. *Transformed mean square error plots.* One-fourth power of the estimated mean squared error for $s^2_{\bar{y}_h,srs}$ and $s^2_{\bar{y}_h,lt1}$ on that of estimated mean square error of $s^2_{\bar{y}_h,\rho 1}$ for (a) pine trees per hectare on E-W swaths (b) pines per hectare on N-S swaths, and (c) pine basal area per hectare on N-S swaths.

The corresponding values for $s^2_{\bar{y}_h,lt1}$ were 1.4, 3.5, and 10.5 times that of $s^2_{\bar{y}_h,\rho1}$; MSE($s^2_{\bar{y}_h,lt1}$) was larger than MSE($s^2_{\bar{y}_h,\rho1}$) for about 70% of the combinations of *n* and *N*. Care must be exercised in interpreting this percentage because of the higher percentages of small *n* in the suite of combinations.

The scatter for MSE($s^2_{\bar{y}_h,lt1}$) is generally centered near, to somewhat above, the 1:1 line for $s^2_{\bar{y}_h,\rho1}$ for most *n* in situations where a linear trend in the population is noted (Figure 4a). This was the case for which $s^2_{\bar{y}_h,lt1}$ was designed and for which we had expected it to perform its best. Yet here too the ρ-assisted estimator $s^2_{\bar{y}_h,\rho1}$ has, on average, smaller MSE. In Figures 4b and 4c, where the presence of nonlinear trend in the population was evident, $s^2_{\bar{y}_h,lt1}$ was clearly intermediate between $s^2_{\bar{y}_h,srs}$ and $s^2_{\bar{y}_h,\rho1}$. This is not surprising because $s^2_{\bar{y}_h,lt1}$ is based on successive differences and should perform reasonably well locally for larger *n*. Average coverage rates for 95% confidence intervals for TPH and BPH for E-W-swaths swaths between 420- and 700-m were virtually 100% for $s^2_{\bar{y}_h,srs}$ and $s^2_{\bar{y}_h,lt1}$ (both) and 92%, 98%, respectively, for $s^2_{\bar{y}_h,\rho1}$. Coverage rate trends were similar for the hardwoods and for the ns-swaths.

Mean squared error is depicted differently in the graphics of Appendix E. On each page the panels are histograms of MSE for all h = 1, … , k and i= 1, 2… , n possibilities, each

panel representing one of our eight estimators. The SRS histogram is displayed in the top right panel and is repeated in red on the other figures for comparative reference. Each page of Appendix E shows different combinations of E-W/N-S and TPH/BPH/SWH for pine/hardwood. Across all the pages we see that (*i*) SRS has the highest mean value (bias), the largest spread (variance), (*ii*) the two stratification estimators are intermediate, (*iii*) the bias in the trend estimators approach zero but they have significant spread, and (*iv*) the correlation estimators have the least bias and the least spread. Numeric tables of the mean variance and MSE for the above combinations of estimators $s^2_{\bar{y}_h,srs}$, $s^2_{\bar{y}_h,lt1}$, and $s^2_{\bar{y}_h,\rho1}$, and only for a set of realistic swath widths. Conclusions drawn from Appendix F are entirely consistent with those already stated.

4. CONCLUSIONS

The within systematic sample variances were larger than the variances among all units. This allows us to state with certainty that, for the number of trees, basal area per hectare, and stem wood biomass in our lidar-forest, the mean of a SyS is more precise than the mean of an SRS of equal size. This was an expected result because these quantities are not spatially random in a forest and are expected to be spatially, positively autocorrelated. Measurement error appears to account for ten to twenty percent of the total per-line variance for trees per hectare and measurement plus model prediction error accounts for between about 25 and 60% of the total per-line variance for basal area per hectare. Calculating "nuggets" may be useful to other scientists as they evaluate the various sources of error in lidar-based inventories and analyze the relative contributions to overall variance.

As noted in the introductory section, Wolter (1984) favored two stratification-based estimators and gave tentative approval to $s^2_{\bar{y}_h,lt1}$; he was not impressed with the performance of $s^2_{\bar{y}_h,\rho1}$. While not discussed explicitly within the text, we found the stratification-based estimators to be intermediate in performance, see Appendices B–E. In natural populations we expect significant spatial autocorrelation. That autocorrelation is reflected in sums over adjacent elements (lines for us) providing the sums are not taken over distances that are too large. In this regard we expected $s^2_{\bar{y}_h,\rho1}$, and $s^2_{\bar{y}_h,\rho2}$ to be at least competitive among our eight estimators. Estimator $s^2_{\bar{y}_h,lt1}$ was designed for

use in populations having linear trends on the unit index. After observing that the lidar-forest had a clear, and nearly linear, trend from south to north, we expected $s^2_{\bar{y}_h, lt1}$ to perform well for estimating the variance of the average basal area and especially for the number of trees per hectare in the case of E-W-swaths. In those cases we found that $s^2_{\bar{y}_h, lt1}$ performed well and that $s^2_{\bar{y}_h, \rho 1}$, and $s^2_{\bar{y}_h, \rho 2}$ were strong competitors, if not better estimators especially when considering MSE. Estimators $s^2_{\bar{y}_h, \rho 1}$, and $s^2_{\bar{y}_h, \rho 2}$ were the best estimator in other cases, with $s^2_{\bar{y}_h, \rho 2}$ often showing slightly better results than $s^2_{\bar{y}_h, \rho 1}$. The linear trend estimator was also superior to SRS when the long-range trends were cubic to quartic. This too was not surprising because those trends are locally approximately linear, and since $s^2_{\bar{y}_h, lt1}$ is based on sequential differences and should perform outperform SRS in such cases.

This study is but a small step toward incorporating SyS in lidar-assisted operational inventories of forested areas. If systematic flight lines occur at distances less than the range of the correlation function, then steps to compensate for the bias due to spatial autocorrelation should be considered, whether it be via $s^2_{\bar{y}_h, \rho 1}$, or some other method. In the presence of autocorrelation some, among a number of other considerations as we move toward operational inventories, include how to appropriately calculate sample sizes, how to deal with (post)stratification, unequal swath lengths (unit sizes), and the fact that pulse densities are not equal across swath widths. The hope is that this research will be of use for future developments in lidar-assisted forest inventories.

REFERENCES

Ackermann, F., 1999. Airborne laser scanning - Present status and future expectations,. *ISPRS Journal of Photogrammetry & RemoteSensing*, 54: 64-67.

Andersen, H-E., and J. Breidenbach, 2007. Statistical properties of mean stand biomass estimators in a lidar based double sampling forest survey design, *Proceedings of the ISPRS workshop: Laser Scanning 2007 and Silvilaser 2007*, 12-14 September 2007, Espoo, Finland (IAPRS Volume XXXVI, Part 3 / W52, 2007), pp. 8-13.

Andersen, H-E., 2009. Using airborne light detection and ranging (lidar) to characterize forest stand condition on the Kenai Peninsula of Alaska, *Western Journal of Applied Forestry,* 24(2): 95-102.

Bechtold, W.A., and P.L. Patterson (eds.), 2005. *The enhanced forest inventory and analysis program – National sampling design and estimation procedures.* US For. Serv. South. Res. Stn. Gen. Tech. Rep. SRS-GTR-80, Asheville, NC. 85 p.

Cochran, W.G., 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations, *Annals of Mathematical Statistic*s, 17: 164-177.

Cochran, W.G., 1977. *Sampling Techniques,* John Wiley & Sons, Inc., New York, 428 p.

Hickman, G.D., and J.E. Hogg, 1969. Application of an airborne pulsed laser for near-shore bathymetric measurements, *Remote Sensing of Environment,* 1: 47-58.

Koop, J.C., 1971. On splitting a systematic sample for variance estimation, *Annals of Mathematical Statistics,* 42: 1084-1087.

Krabill, W.B., J.G. Collings, R.N. Swift, and M.L. Butler, 1980. Airborne laser topographic mapping results from Initial Joint NASA/U.S. Army Corps of Engineers Experiment. NASA Technical Memorandum 73287, Wallops Flight Center.

Lefsky, M.A., W.B. Cohen, G.G. Parker and D.J. Harding, 2002. Lidar Remote Sensing for Ecosystem Studies, *Bioscience,* 52(1): 19-30

Meyer, H.A., 1956. The calculation of the sampling error of a cruise from the mean square successive difference, *Journal of Forestry,* 54(5): 341.

Moore, P.G., 1955. The properties of the mean square successive differences in samples from various populations, *Journal of the American Statistical Association,* 50(270): 434-456.

Mutlu, M., S.C. Popescu, C. Stripling, and T. Spencer, 2008. Assessing surface fuel models using lidar and multispectral data fusion, *Remote Sensing of Environment,* 112(1): 274-285.

Nelson, R., W. Krabill, and G. Maclean, 1984. Determining forest canopy characteristics using airborne laser data, *Remote Sensing of Environment*, 15: 201-212.

Nelson, R., R. Swift, and W. Krabill, 1988. Using airborne lasers to estimate forest canopy and stand characteristics, *Journal of Forestry*, 86: 31-38.

Nelson, R., M. A. Valenti, A. Short, and C. Kelley, 2003. A multiple resource inventory of Delaware using airborne laser data, *Bioscience,* 53(10): 981-992.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small-footprint airborne scanning laser, *Scandinavian Journal of Forest Research*, 19: 164-179.

Næsset, E., 1997a. Determination of mean tree height of forest stands using airborne laser scanner data, *ISPRS Journal of Photogrammetry and Remote Sensing*, 52: 49-56

Næsset, E., 1997b. Estimating timber volume of forest stands using airborne laser scanner data, *Remote Sensing of Environment,* 61: 246-253.

Parker, R.C., and D.L. Evans, 2004. An application of lidar in a double-sample forest inventory, *Western Journal of Applied Forestry,* 19(2): 95-101.

Parker, R.C., and D.L. Evans, 2009. LiDAR Forest Inventory with Single-Tree, Double- and Single-Phase Procedures, *International Journal of Forestry Research,* Vol. 2009, Article ID 864108, 6 pages, doi:10.1155/2009/864108.

Parker, R.C., and A.L. Mitchel, 2005. Smoothed versus unsmoothed lidar in a double-sample forest inventory, *Southern Journal of Applied Forestry,* 29(1): 40-47.

Popescu, S.C., and R.H. Wynne, 2004. Seeing the trees in the forest: Using lidar and multispectral data fusion with local filtering and variable window size for estimating tree height, *Photogrammetric Engineering & Remote Sensing*, 70(5): 589-604.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar, *Biomass & Bioenergy,* 31(9): 646-655.

Särndal, C.-E., B. Swenssen, and J. Wretman, 1992. *Model Assisted Survey Sampling,* Springer-Verlag New York, Inc., New York, 665 p.

Schbenberger and Pierce, 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*, CRC Press, Boca Raton, FL, 738 p.

Wolter, K. M., 1984. An investigation of some estimators of variance for systematic sampling, *Journal of the American Statistical Association*, 79 (388): 781-790.

Yates F., 1949. *Sampling Methods for Censuses and Surveys*, Charles W. Griffin, London, 318 p.

Zhao, K., S. Popescu, and R. Nelson, 2009. Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers, *Remote Sensing of Environment*, 113: 182-196.

APPENDIX A
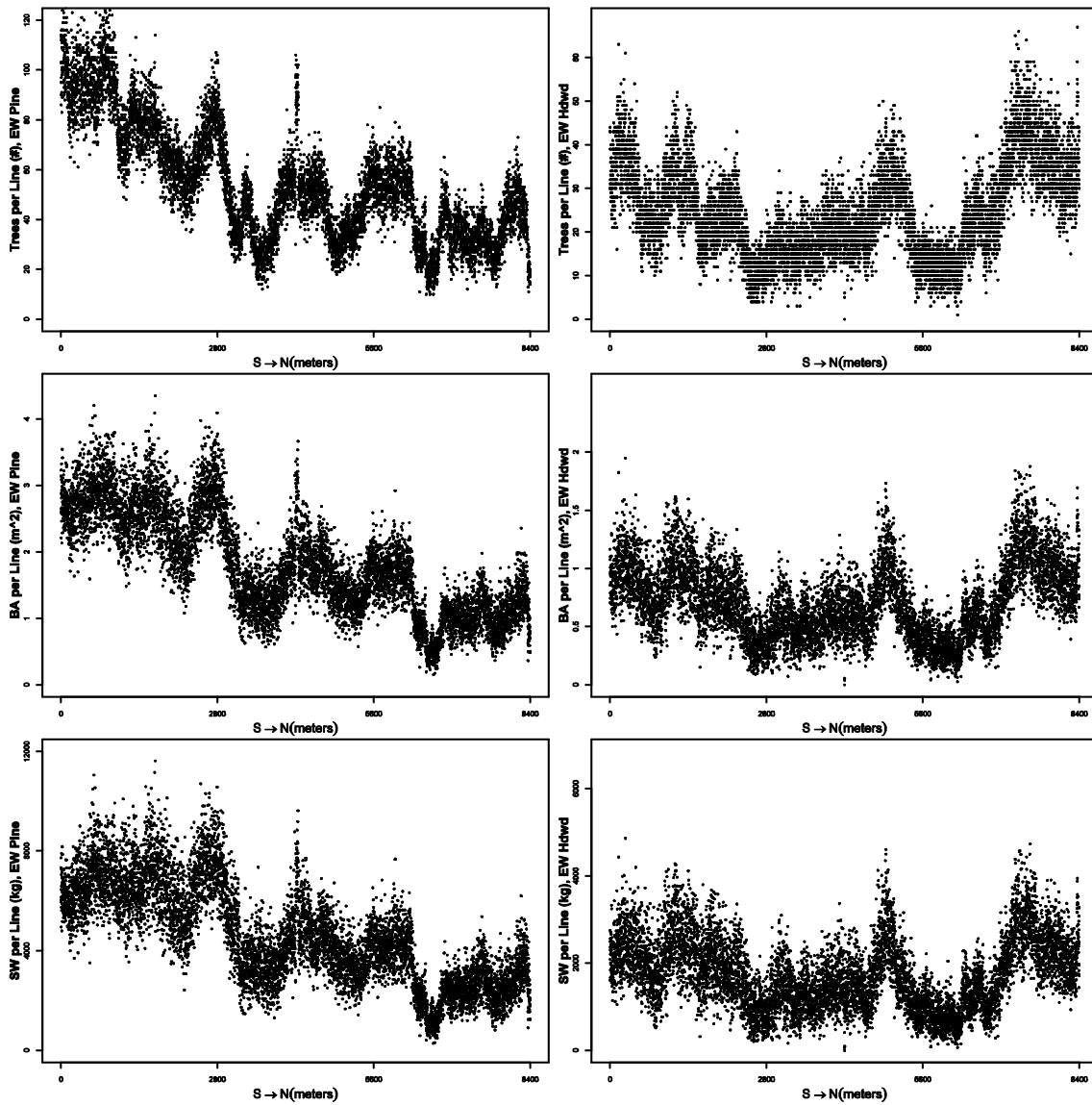
Data plots for E-W lines: TPH (top), BPH (middle), and SWH (bottom); pines (left) and hardwoods (right). Every other point is plotted.
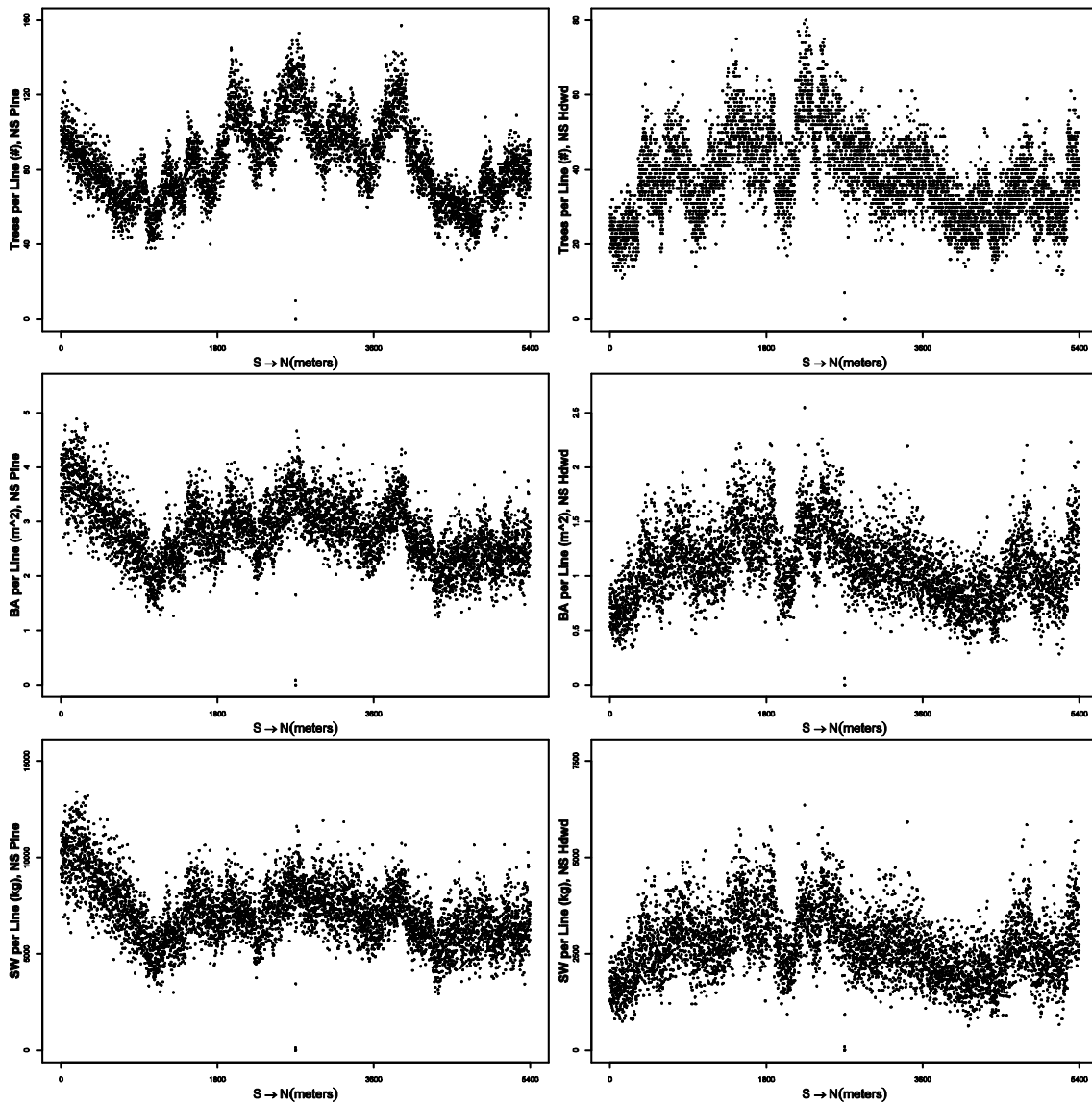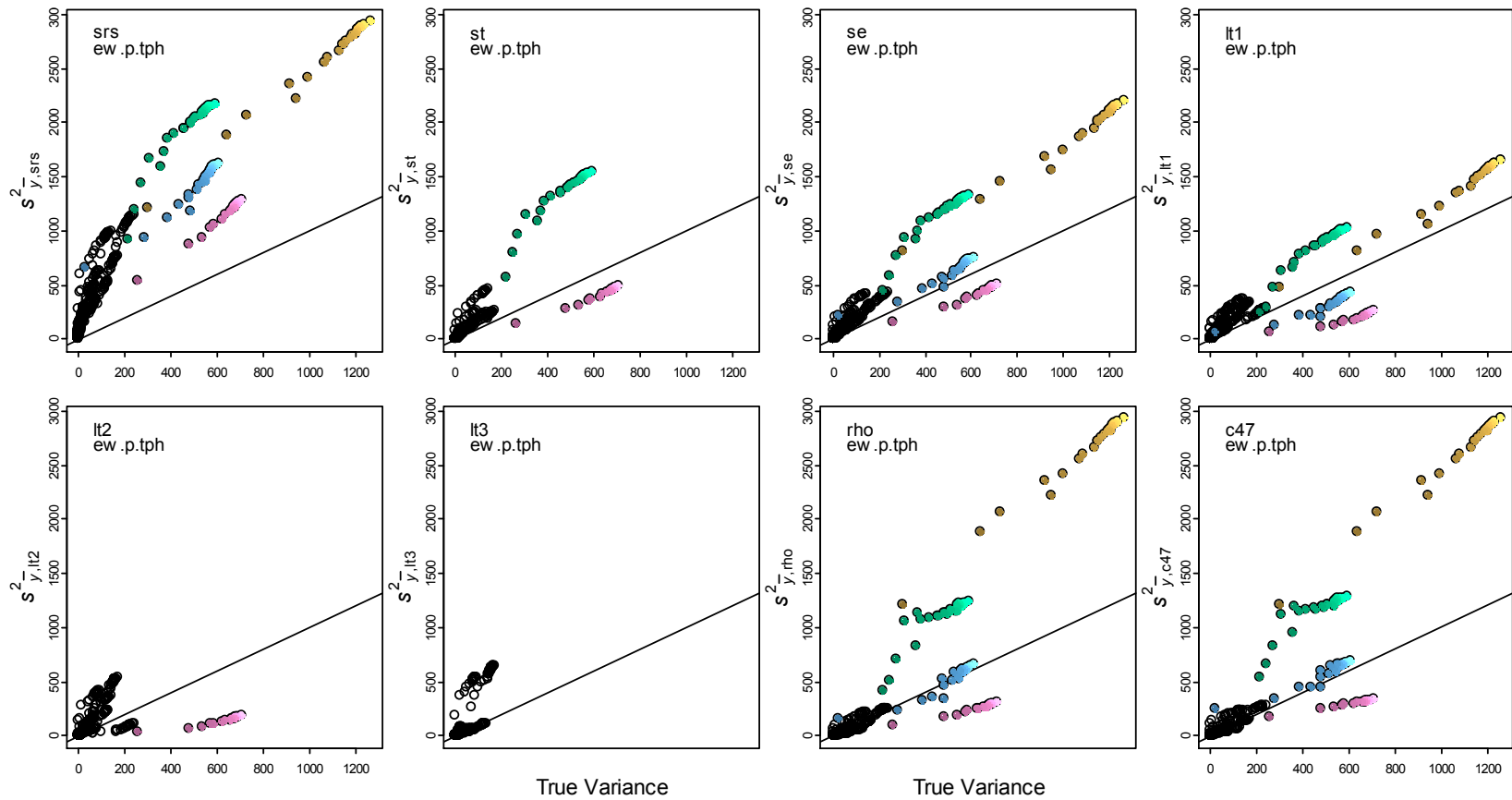
Data plots for N-S lines: TPH (top), BPH (middle), and SWH (bottom); pines (left) and hardwoods (right). Every other point is plotted.
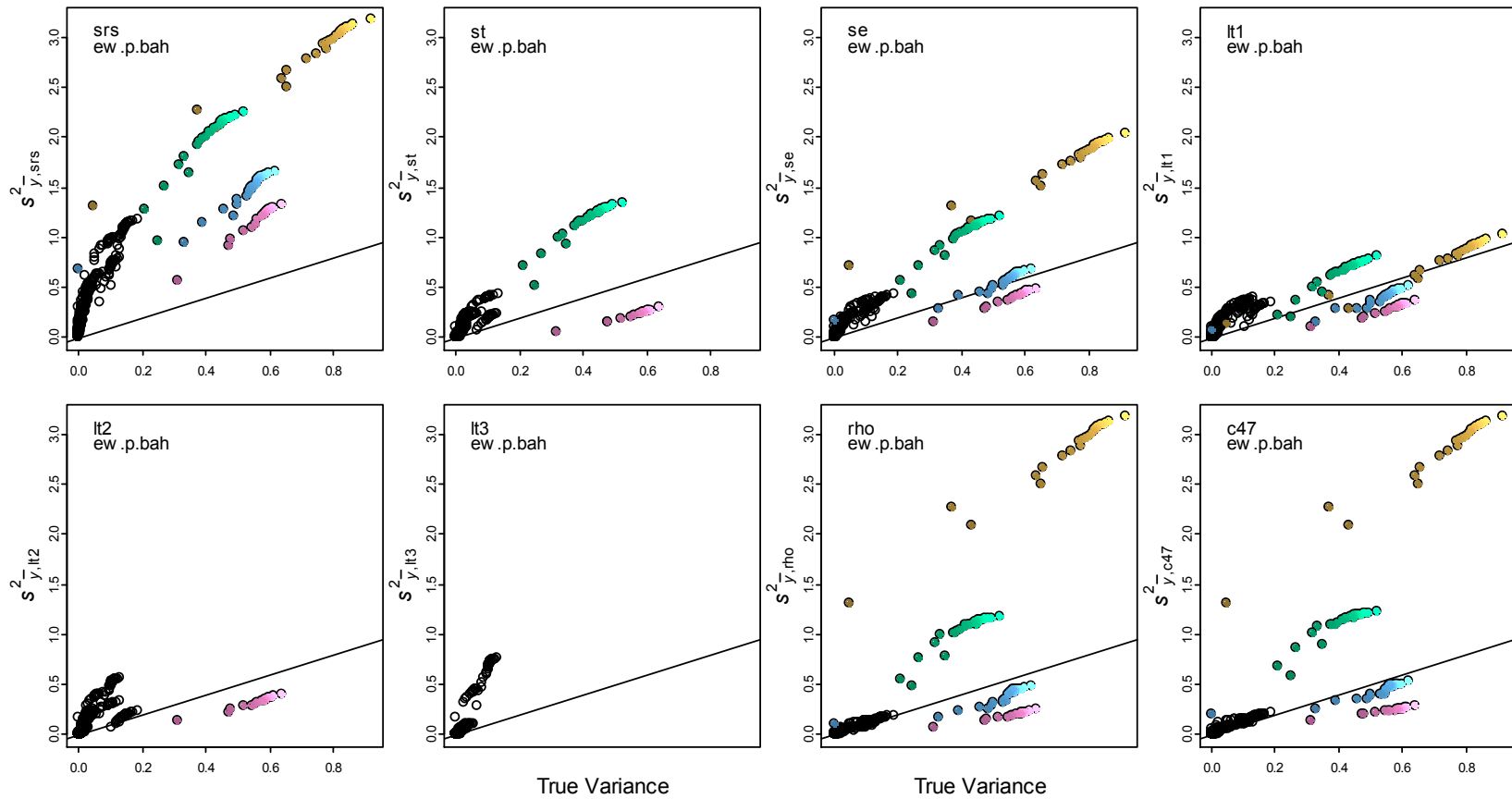
APPENDIX B
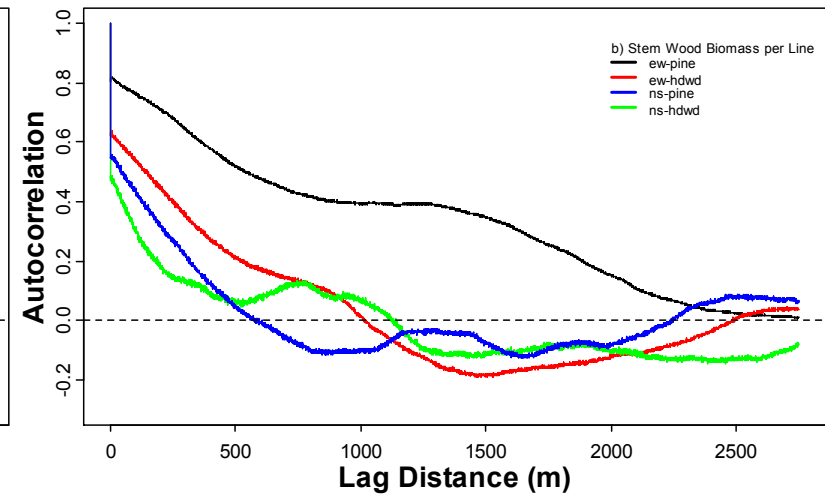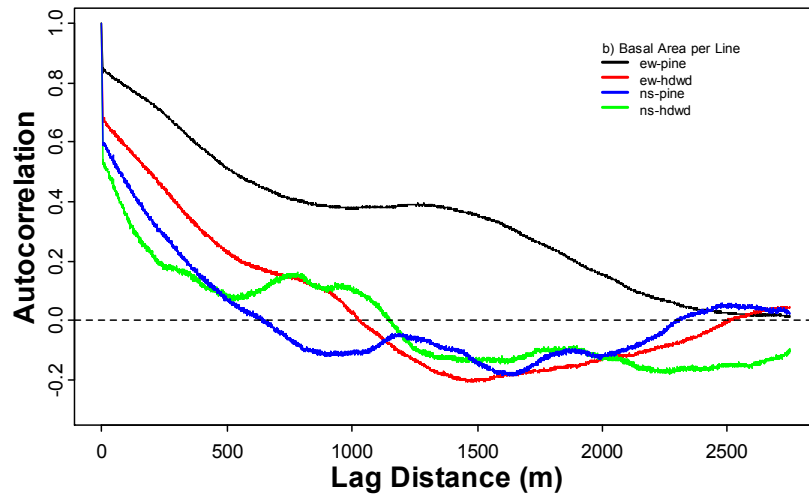
Empirical Autocorrelation functions to 2500 meters for Basal Area (left) and Stem Wood Biomass (right)
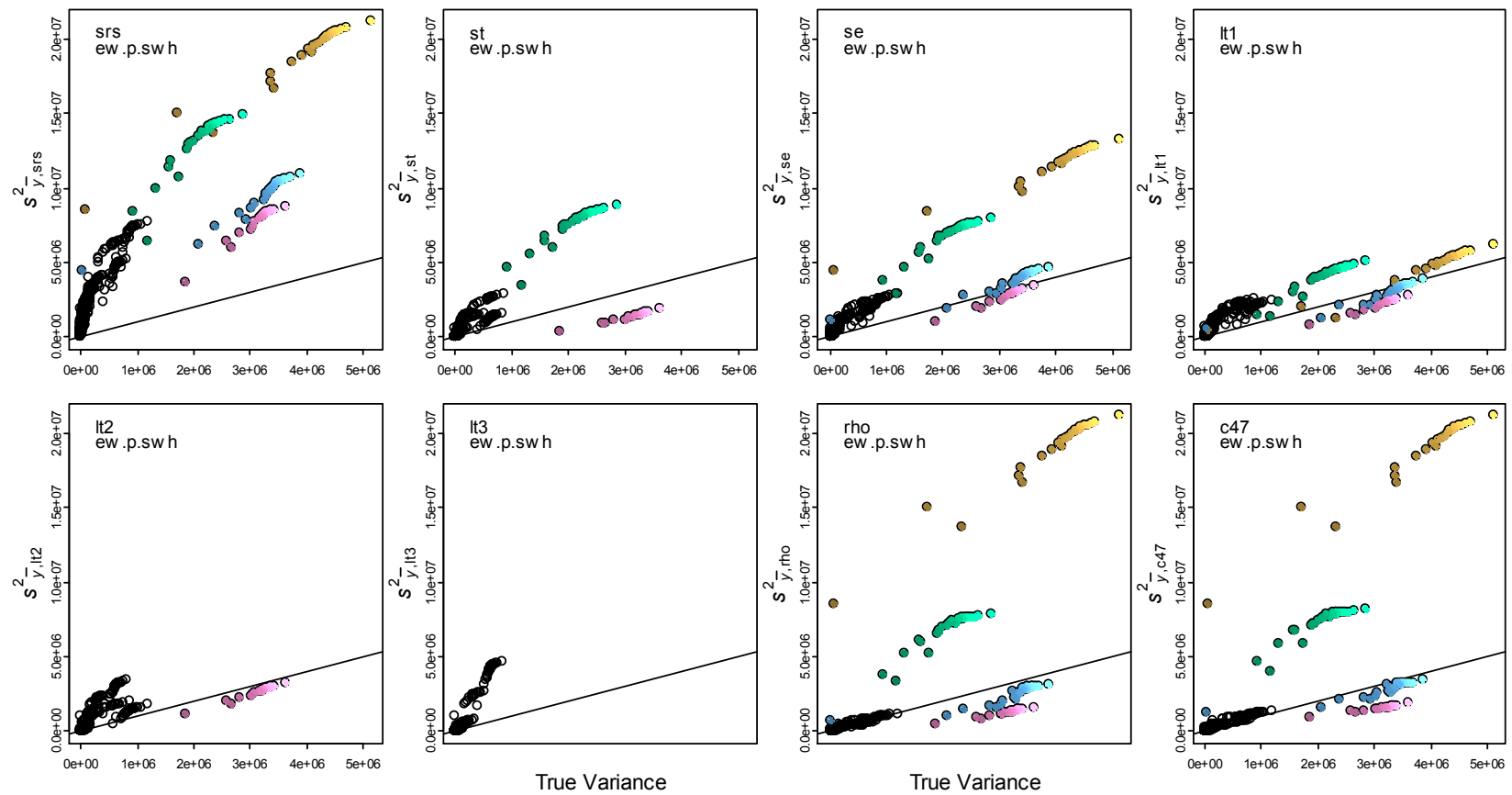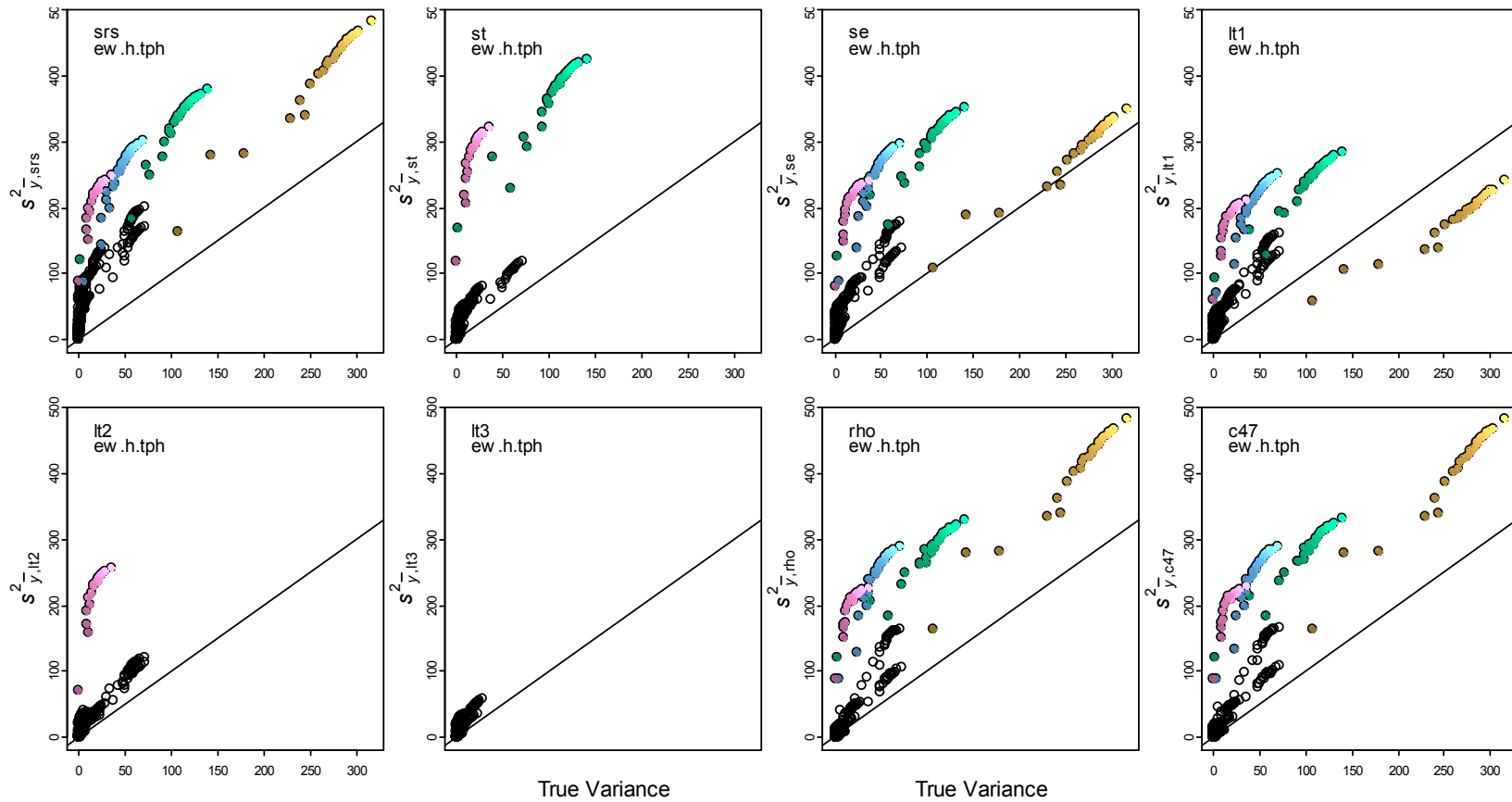
APPENDIX C

Estimated variance plotted on true SyS variance for number of pines per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
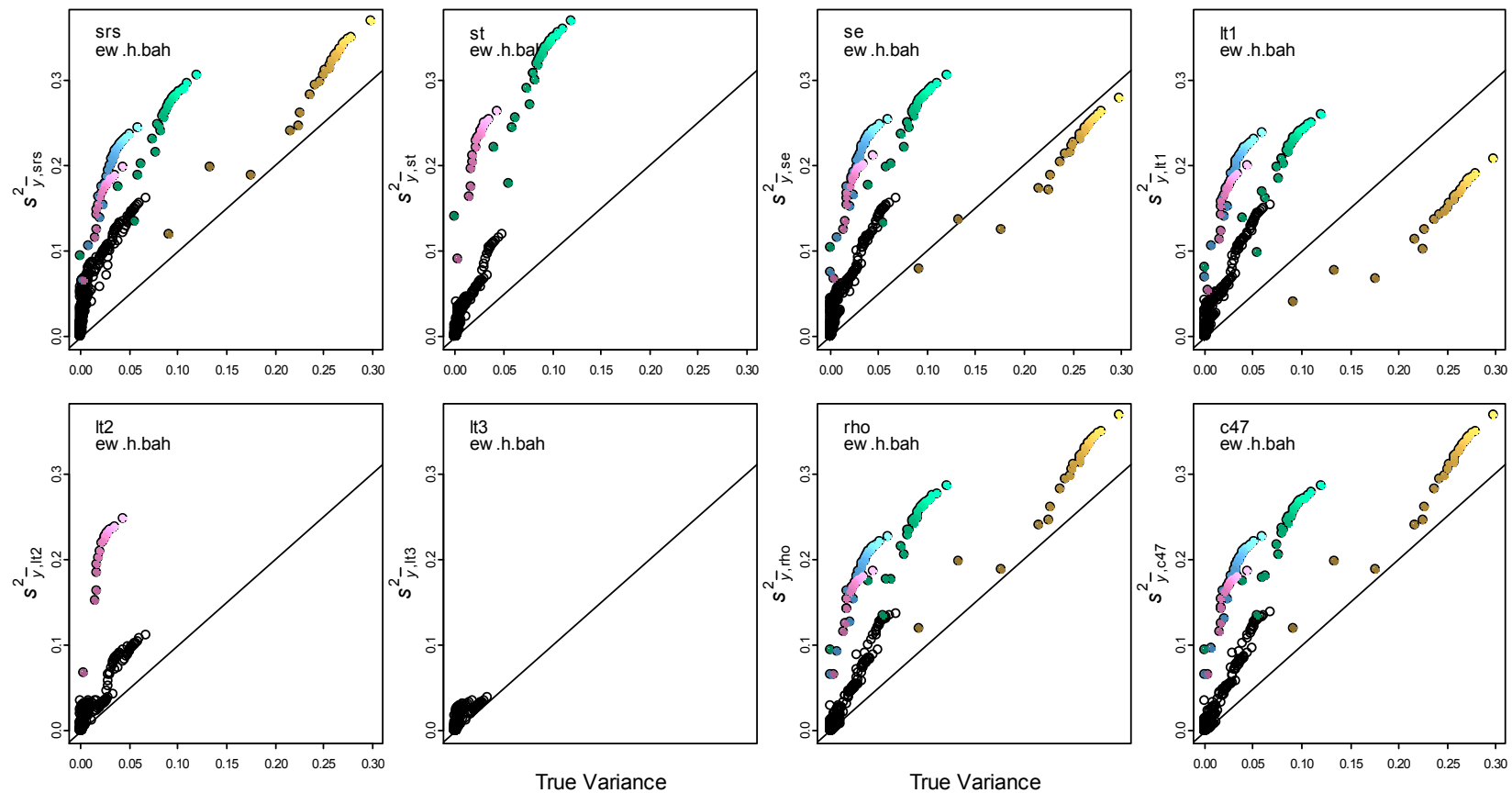
Estimated variance plotted on true SyS variance for basal area of pines per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
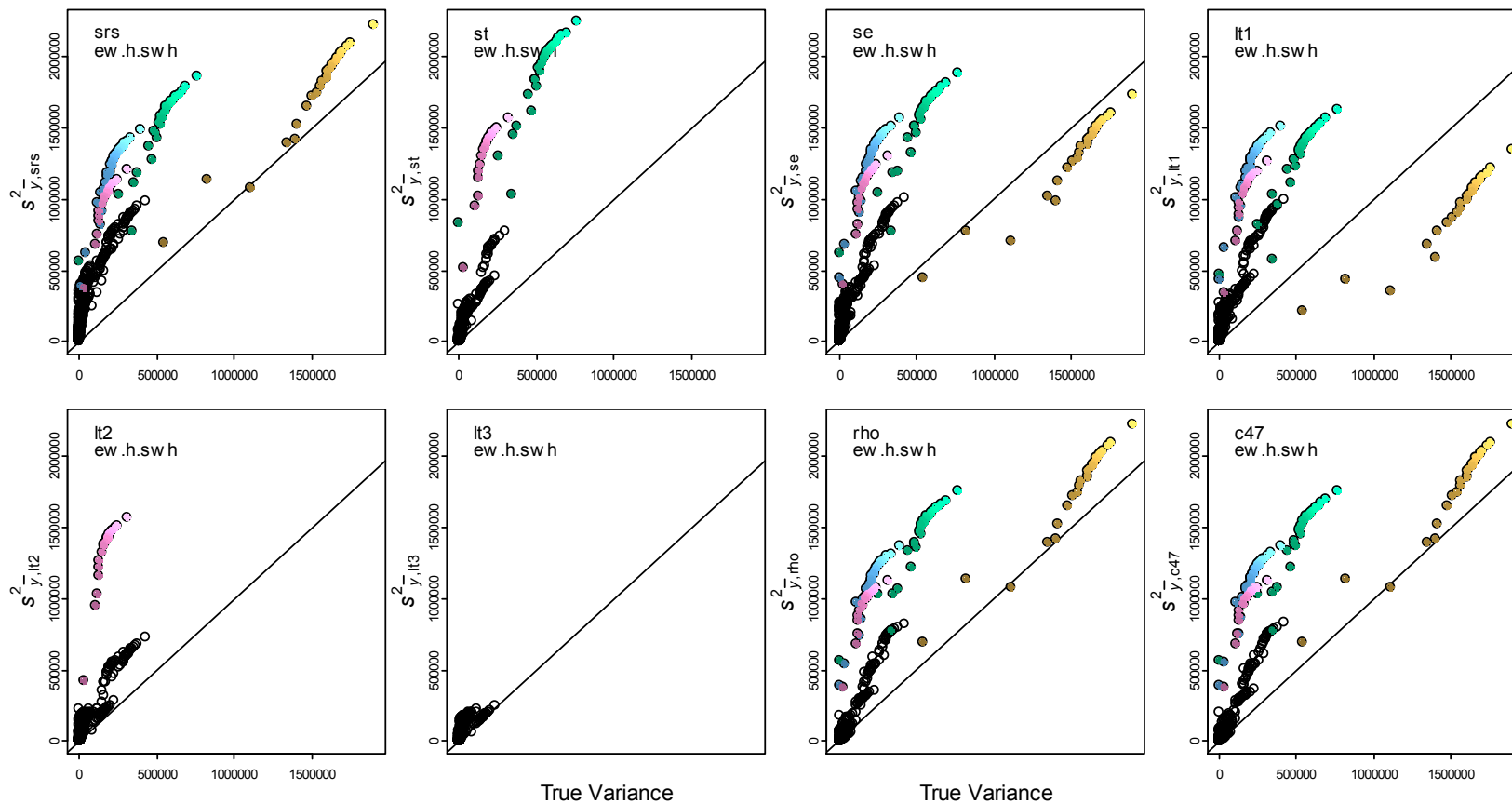
Estimated variance plotted on true SyS variance stem wood biomass per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
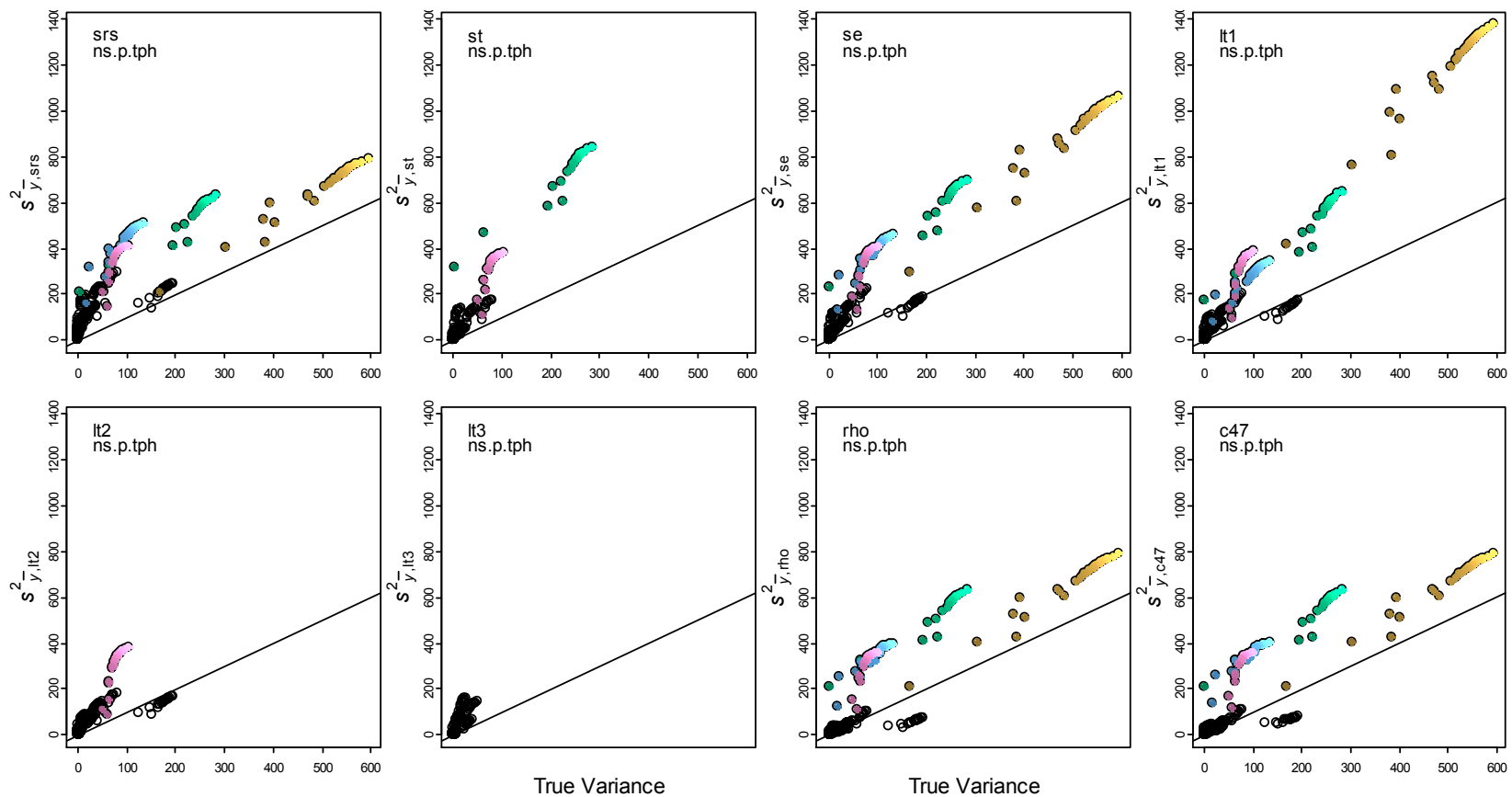
Estimated variance plotted on true SyS variance for number of hardwood trees per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance plotted on true SyS variance for basal area of hardwoods per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance plotted on true SyS variance for stem wood biomass of hardwoods per hectare on east-west swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance plotted on true SyS variance for number of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
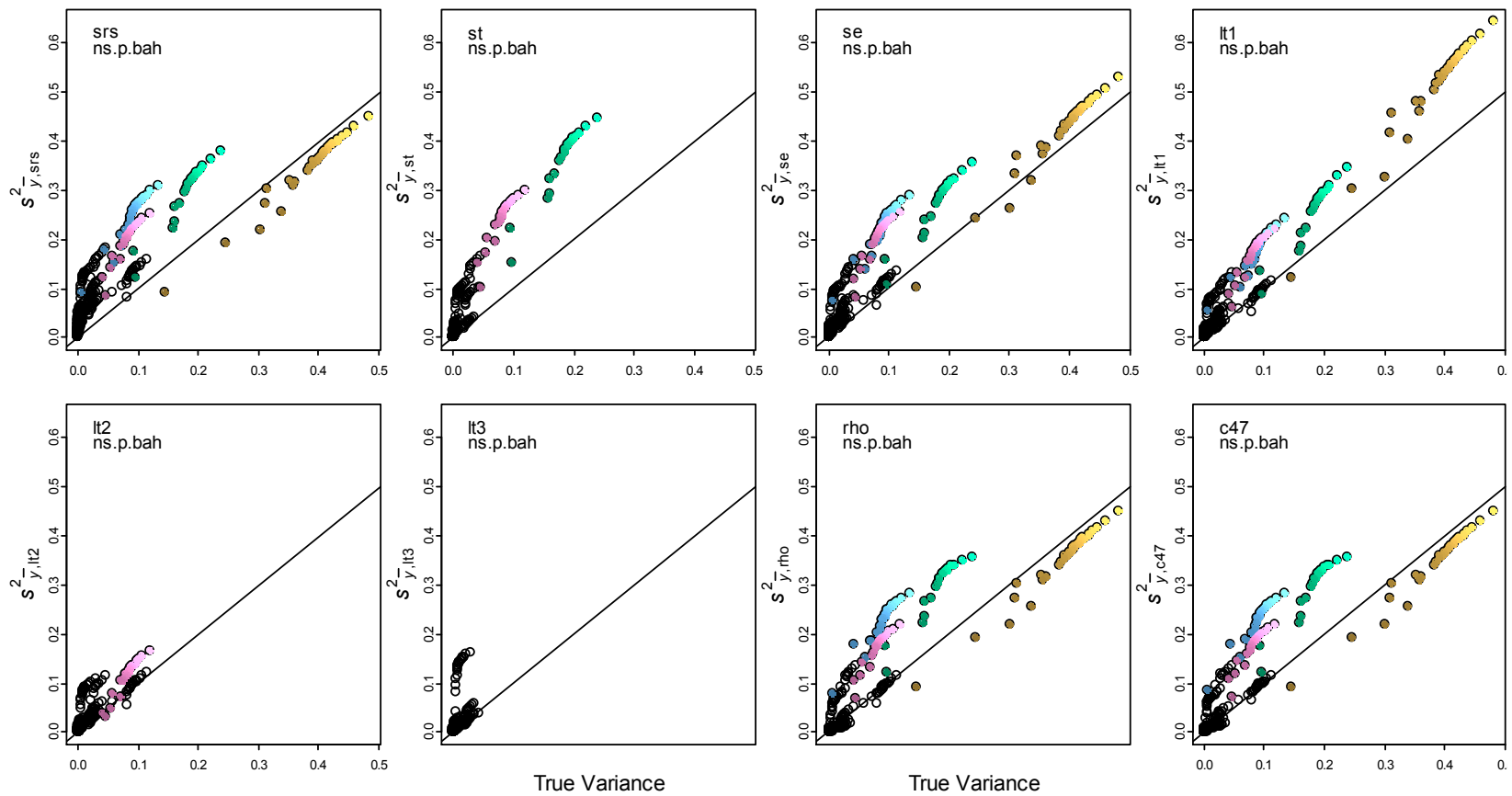
Estimated variance plotted on true SyS variance for basal area of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate n = 3 (gold), 4 (green), 5 (blue), and 6 (pink).
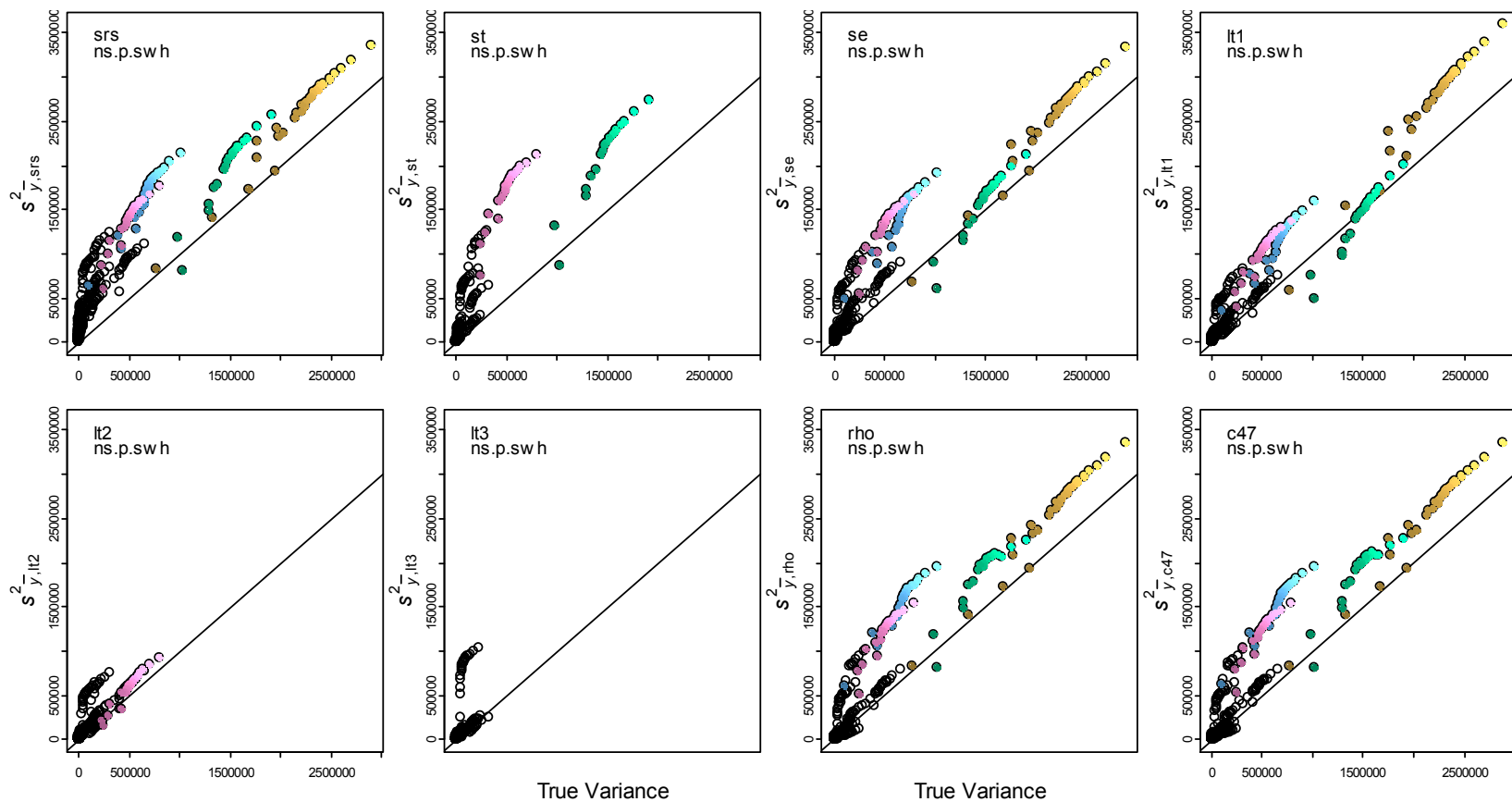
Estimated variance plotted on true SyS variance for stem wood biomass of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
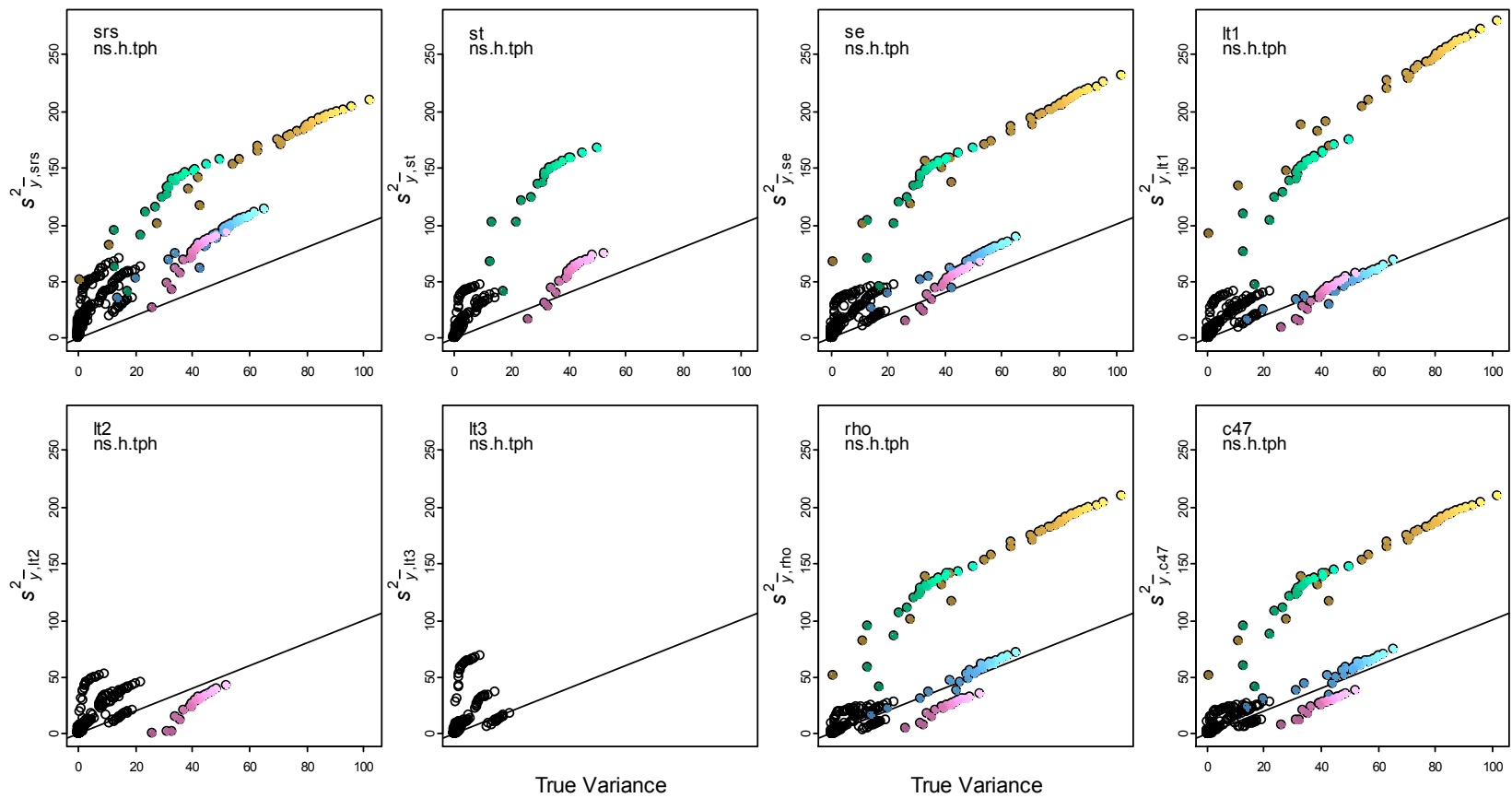
Estimated variance plotted on true SyS variance for number of hardwood trees per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
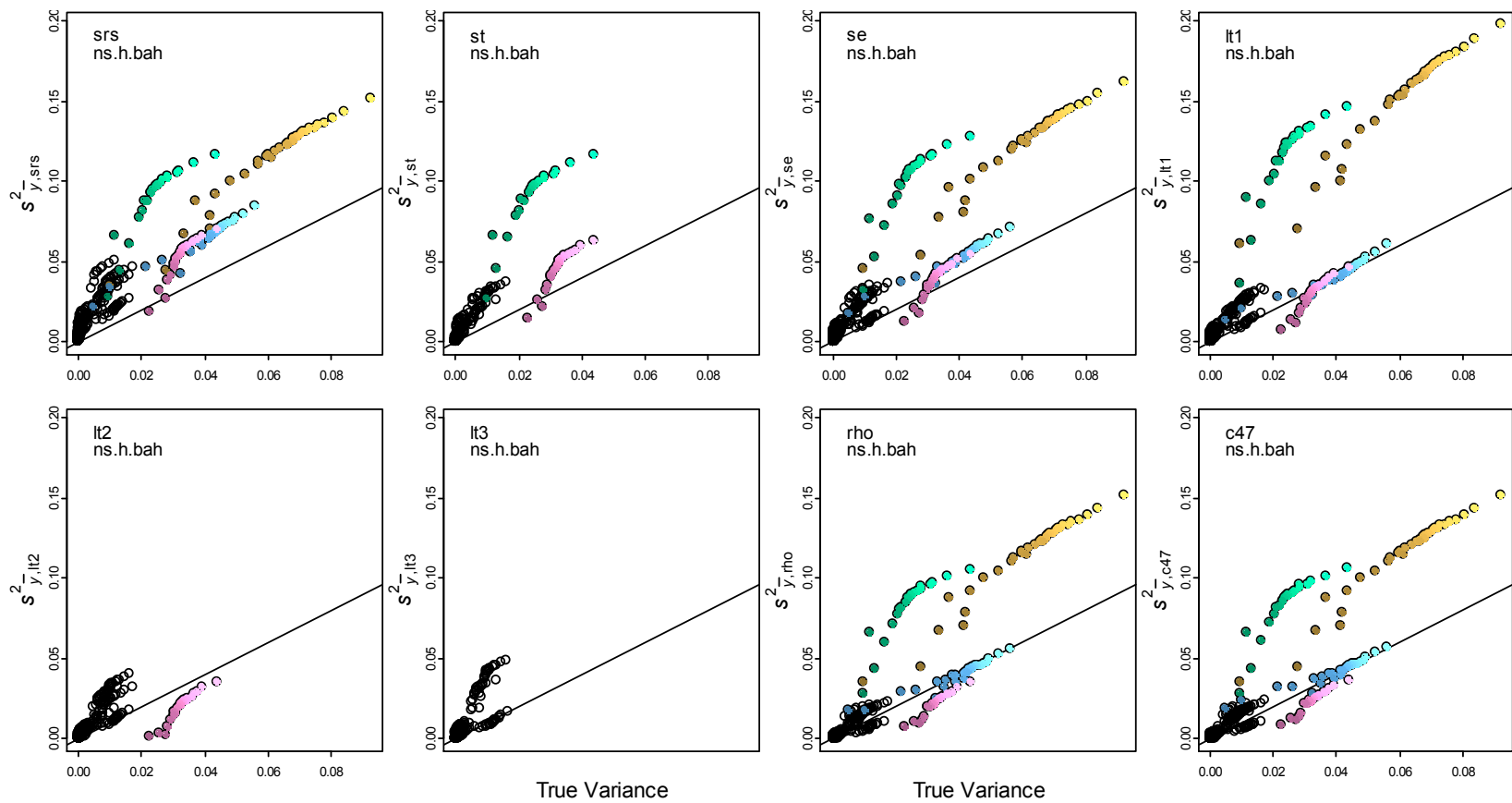
Estimated variance plotted on true SyS variance for basal area of hardwoods per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
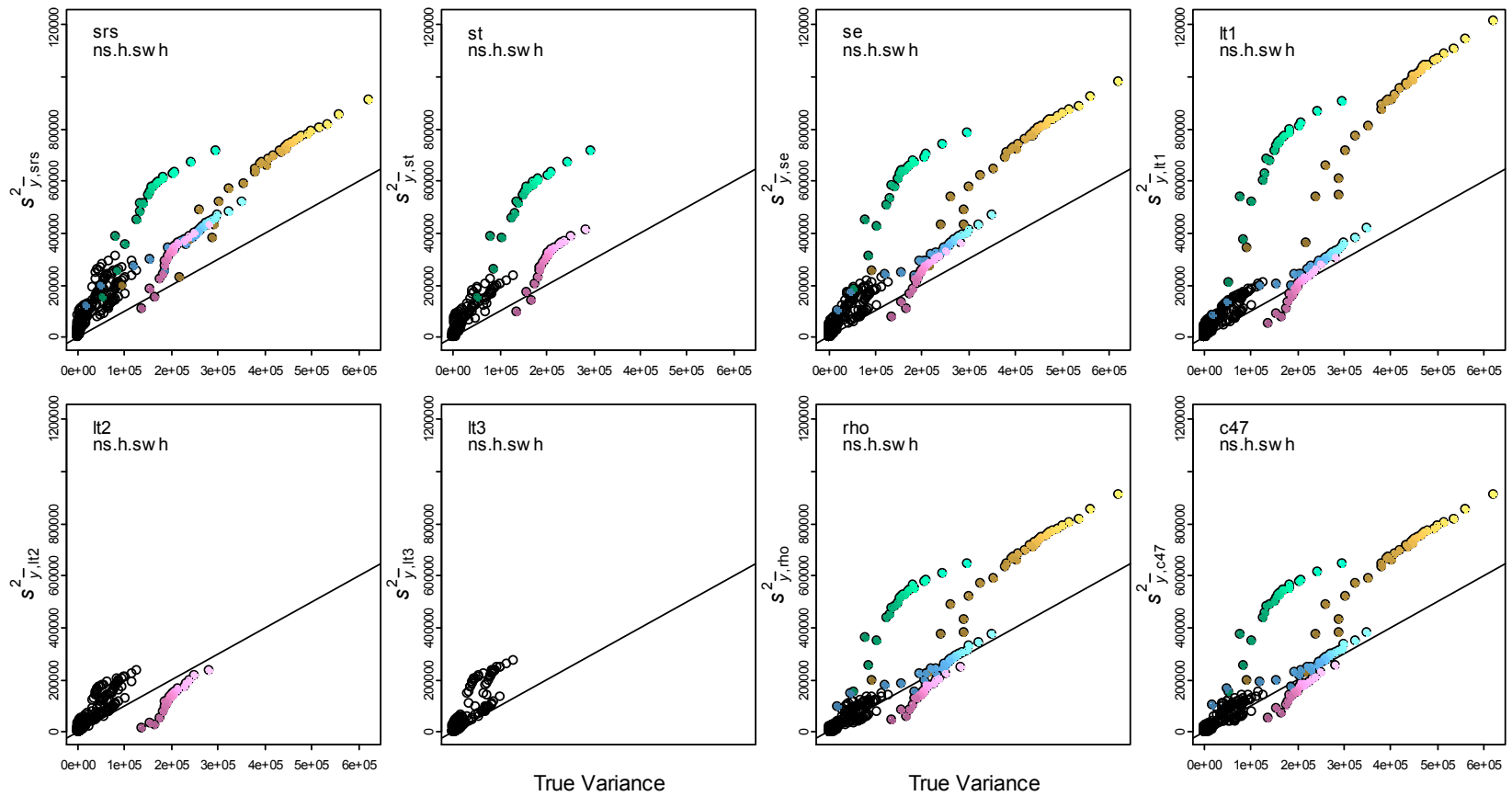
Estimated variance plotted on true SyS variance for stem wood biomass of hardwoods per hectare on north-south swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

APPENDIX D

Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for number of pines per hectare on east-west swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for basal area of pines per hectare on east-west swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for stem wood biomass of pines per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
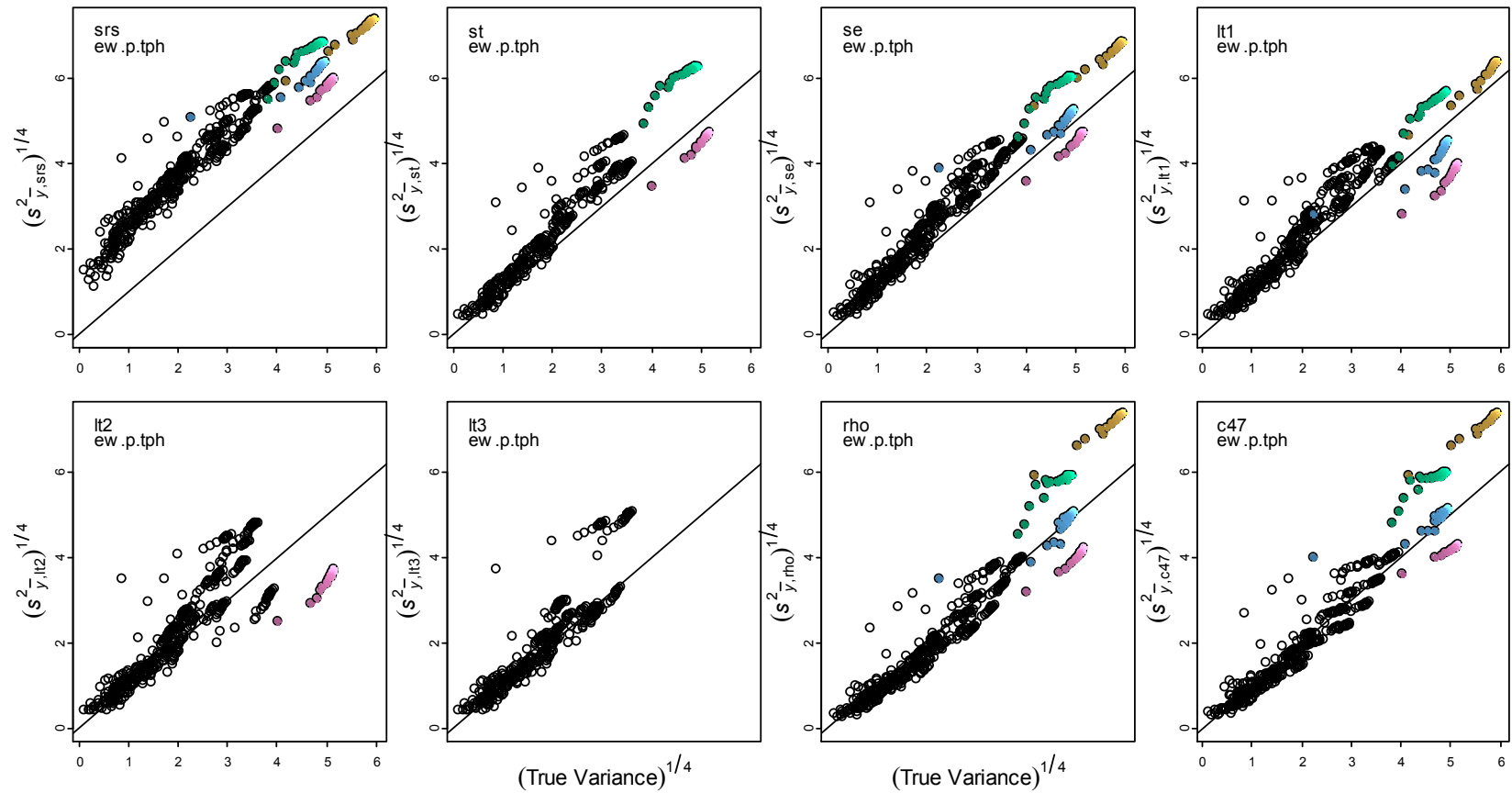
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for number of hardwood trees per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for basal area of hardwoods per hectare on east-west swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
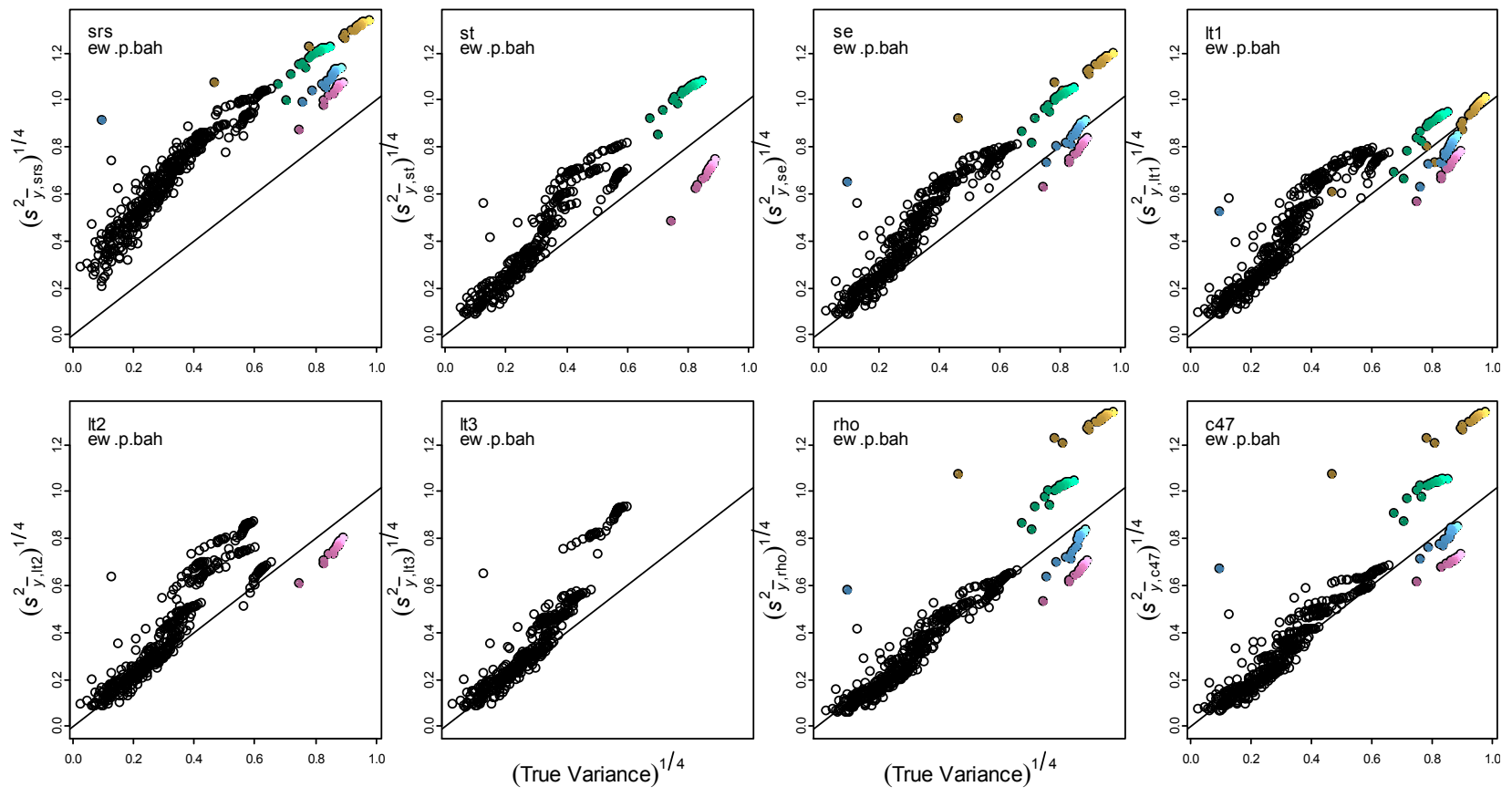
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for stem wood biomass of hardwoods per hectare on east-west swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
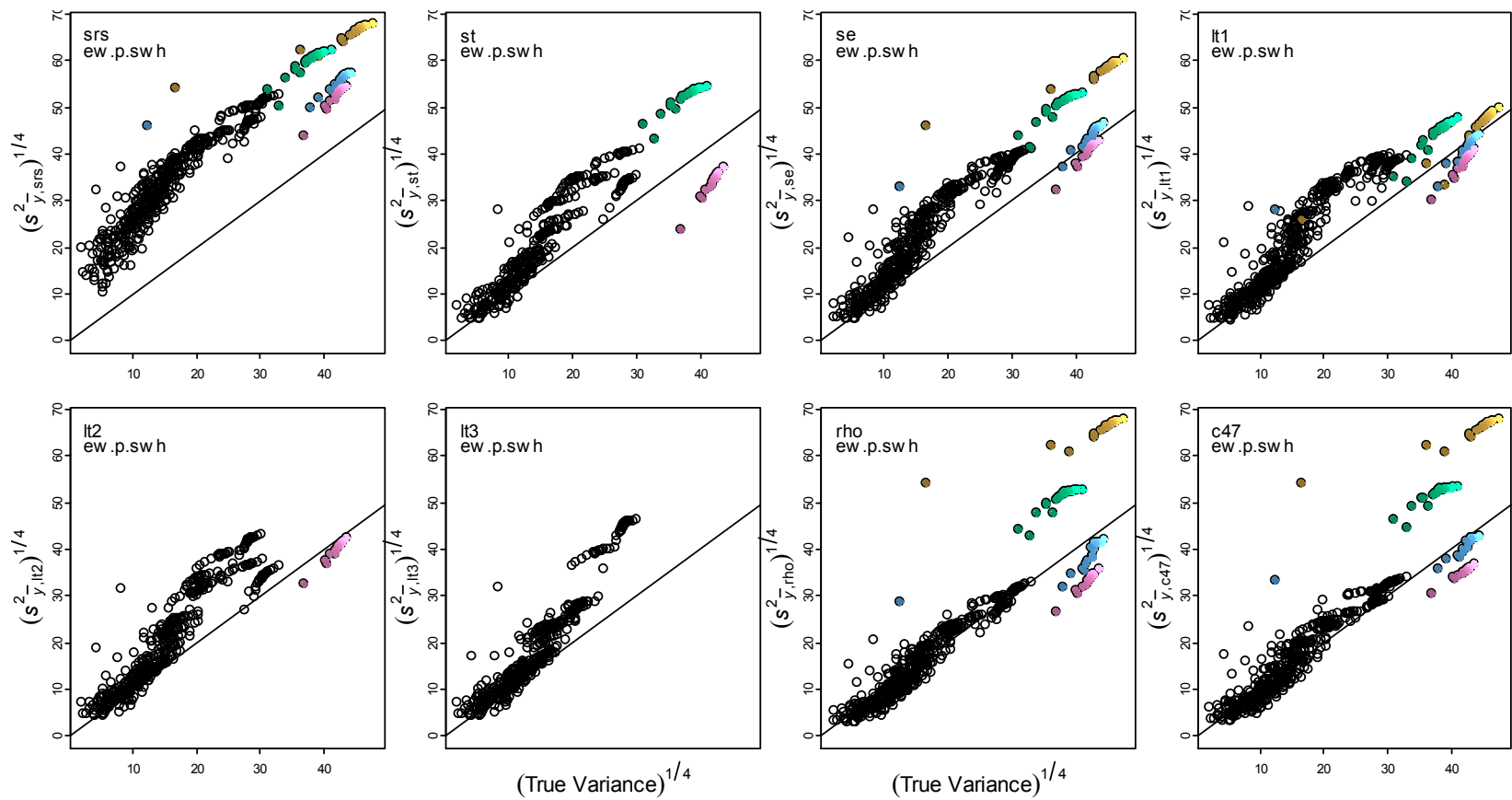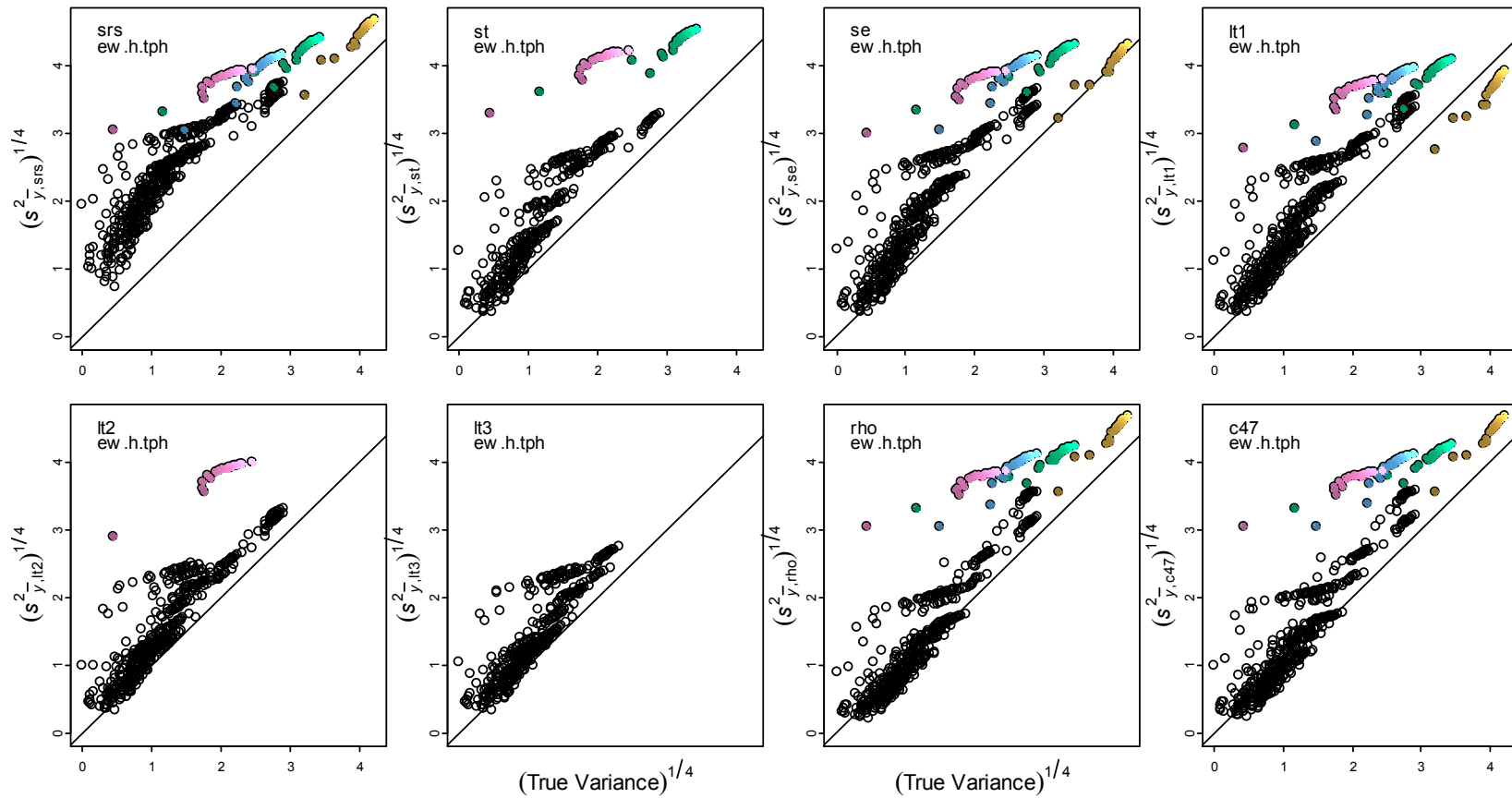
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for number of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
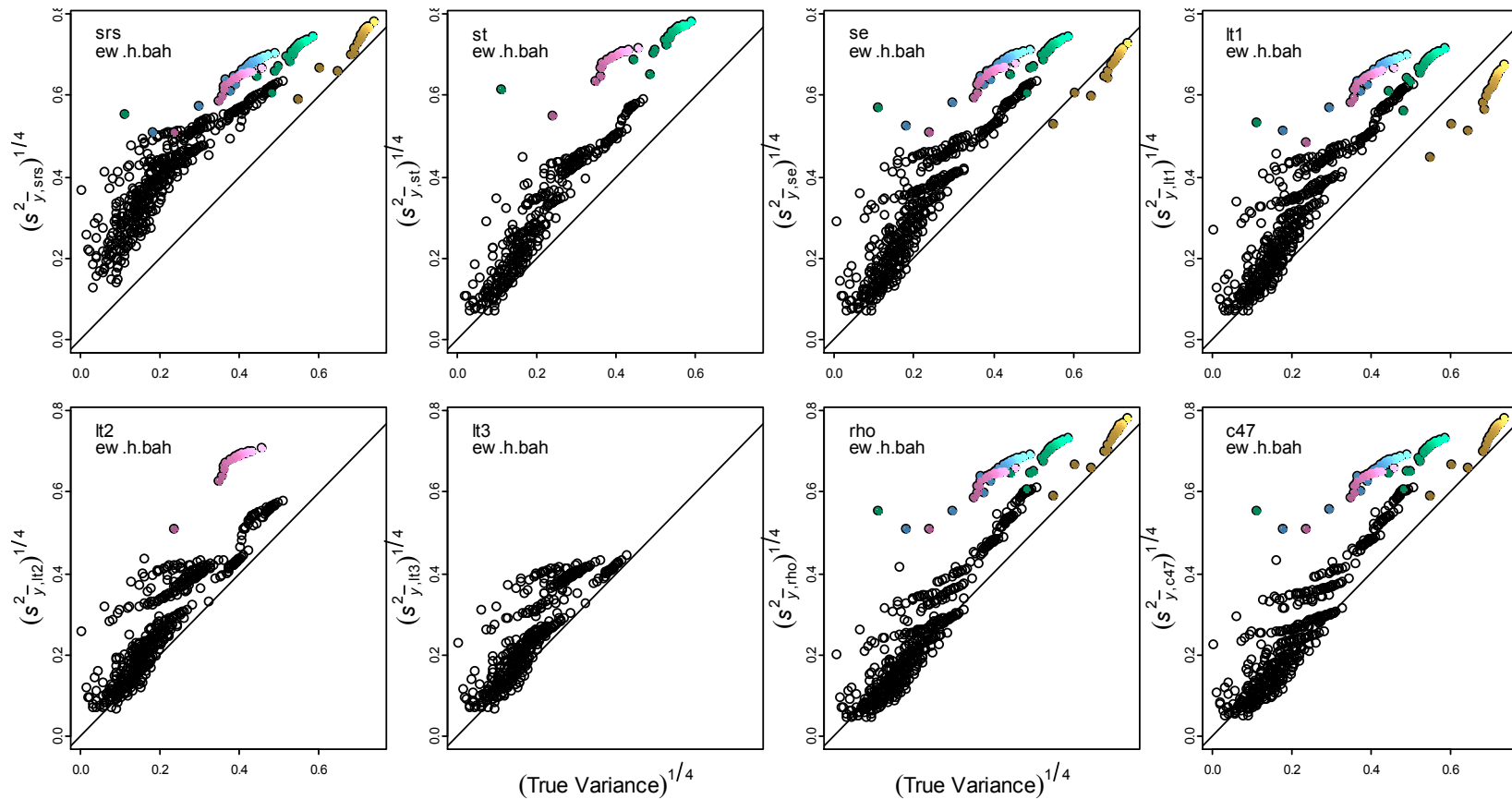
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for basal area of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
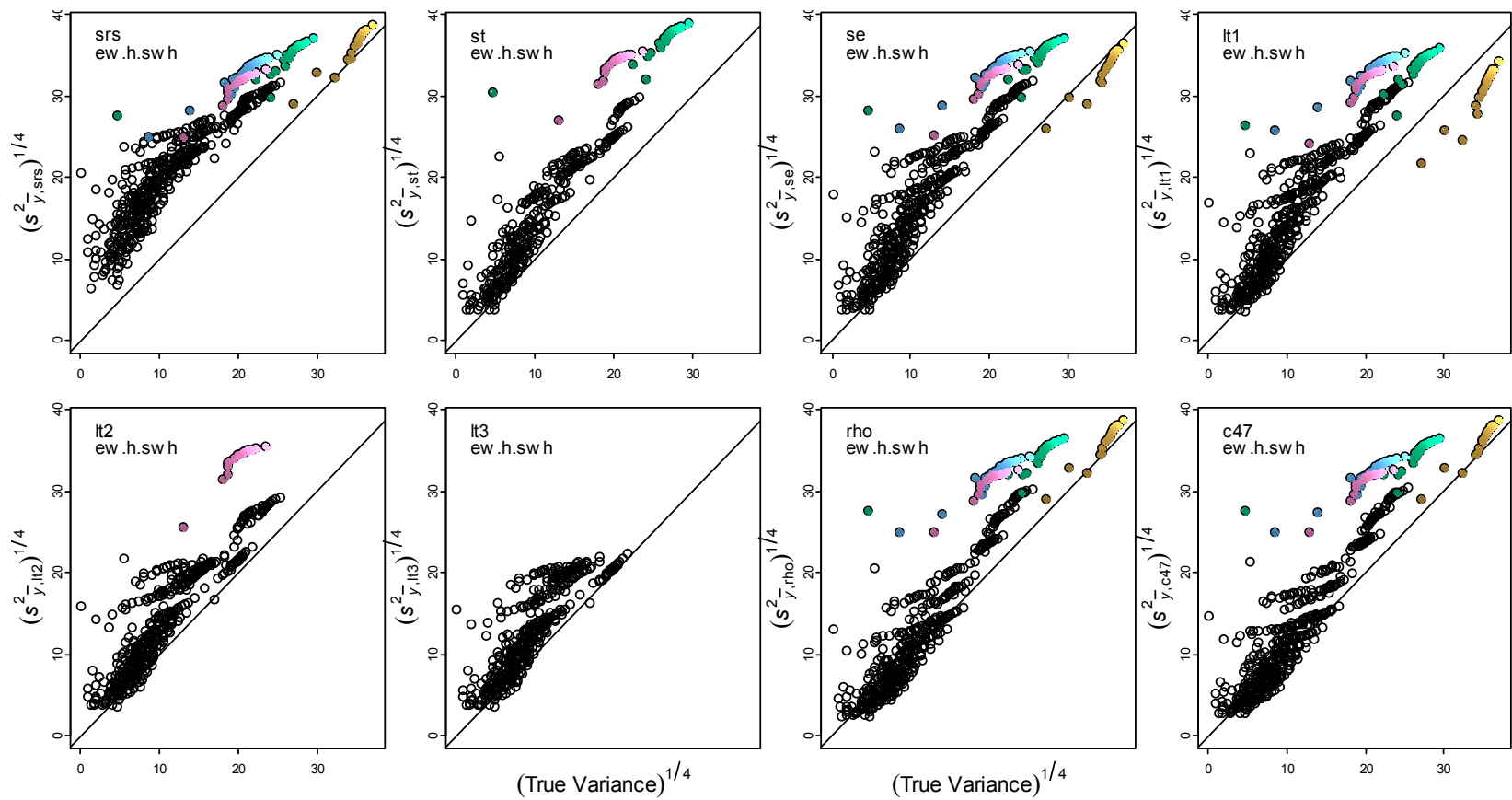
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for stem wood biomass of pines per hectare on north-south swaths. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
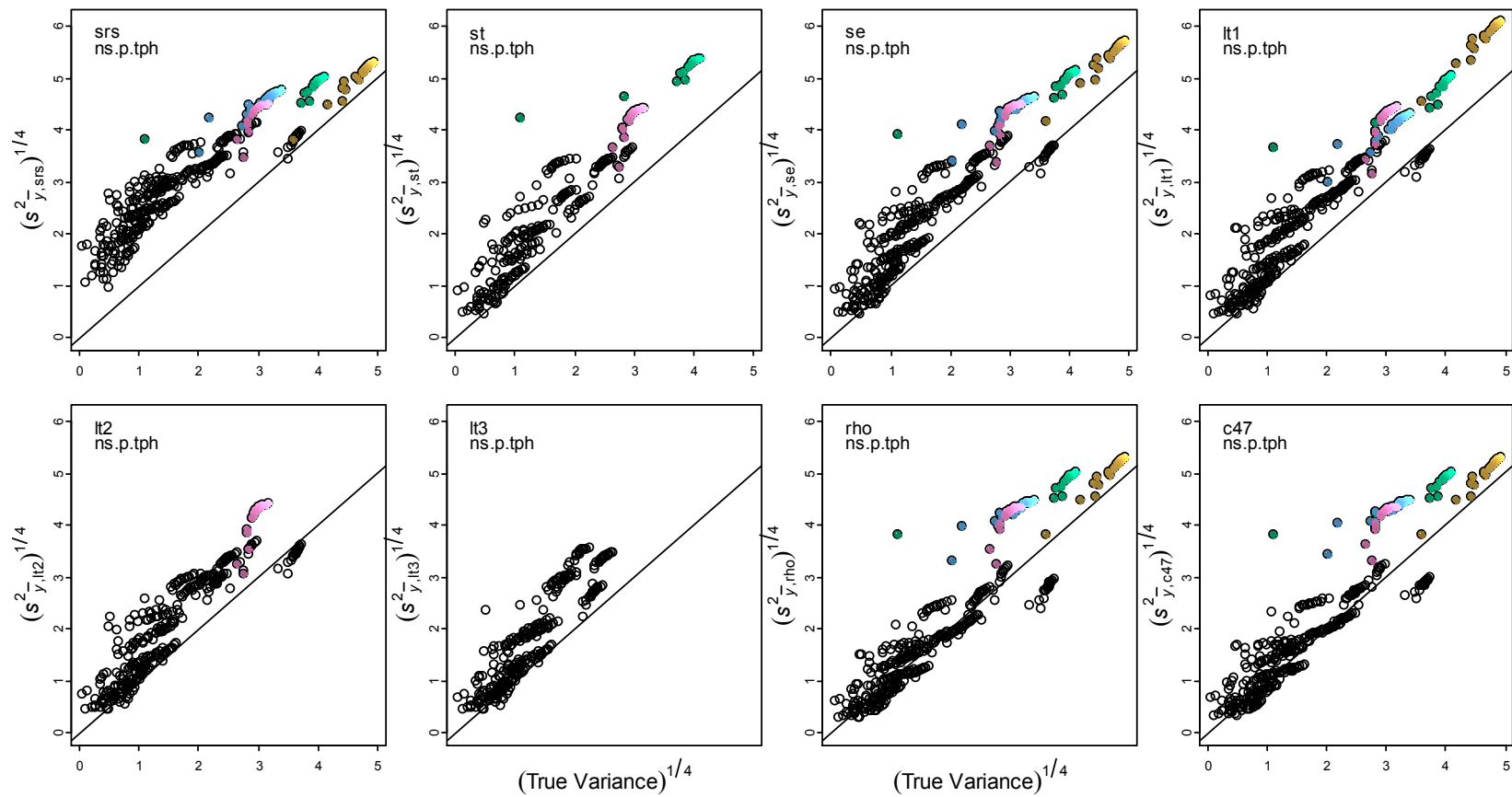
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for number of hardwood trees per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).
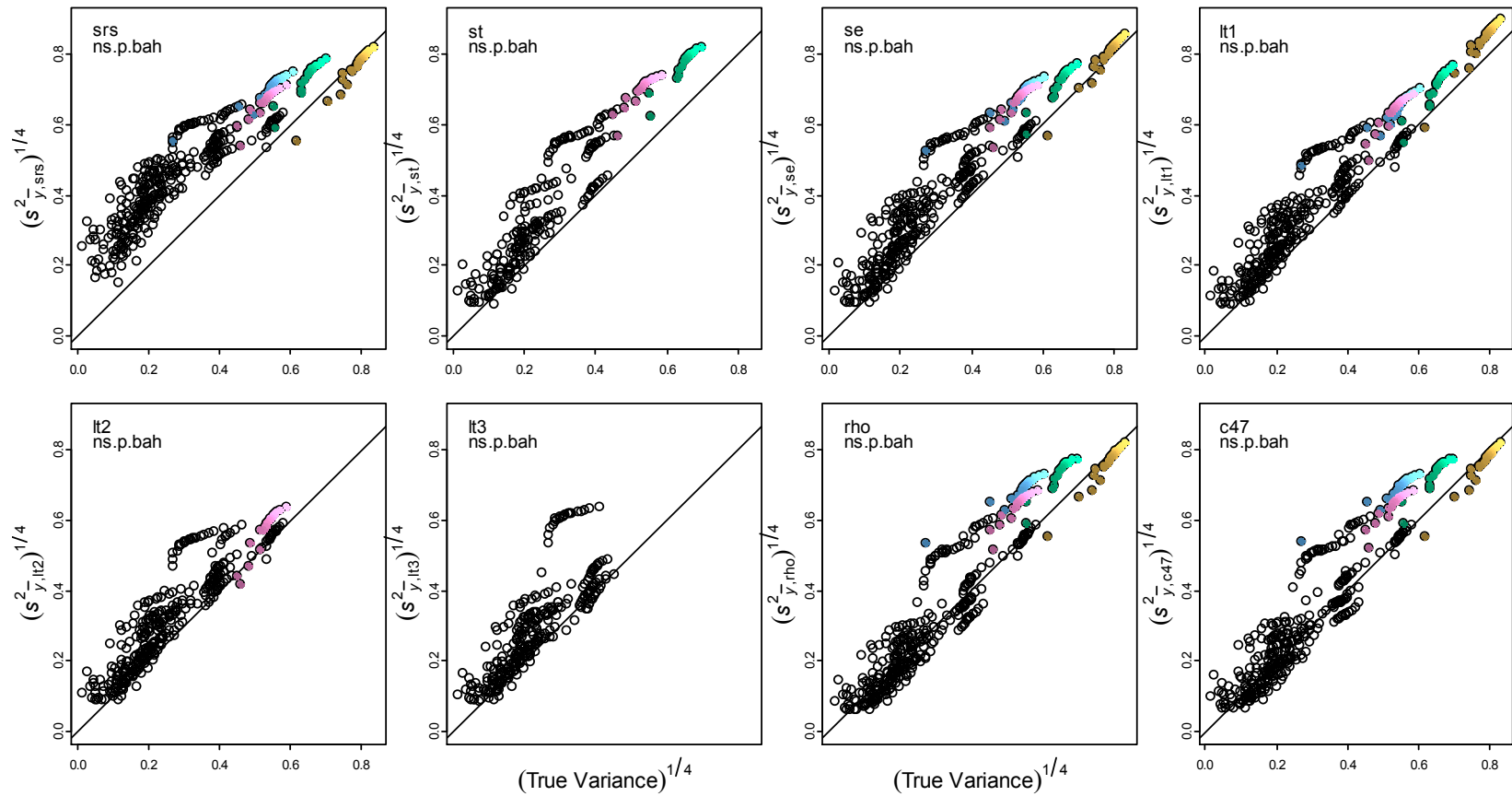
Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for basal area of hardwoods per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

Estimated variance transformed by $v^{1/4}$ plotted on true SyS variance for stem wood volume of hardwoods per hectare on north-south swaths. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$. Colors indicate $n = 3$ (gold), 4 (green), 5 (blue), and 6 (pink).

APPENDIX E

The above figures show Mean Square Error (MSE) for each of our eight estimators for Trees per Hectare (TPH) based on our population for east-west swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Basal Area per Hectare (BAH) based on our population for east-west swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Stem Wood Biomass per Hectare (SWH) based on our population for east-west swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Trees per Hectare (TPH) based on our population for east-west swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Basal Area per Hectare (BAH) based on our population for east-west swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Stem Wood Biomass per Hectare (SWH) based on our population for east-west swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Trees per Hectare (TPH) based on our population for north-south swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Basal Area per Hectare (BAH) based on our population for north-south swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.

The above figures show Mean Square Error (MSE) for each of our eight estimators for Stem Wood Biomass per Hectare (SWH) based on our population for north-south swaths of pine stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.
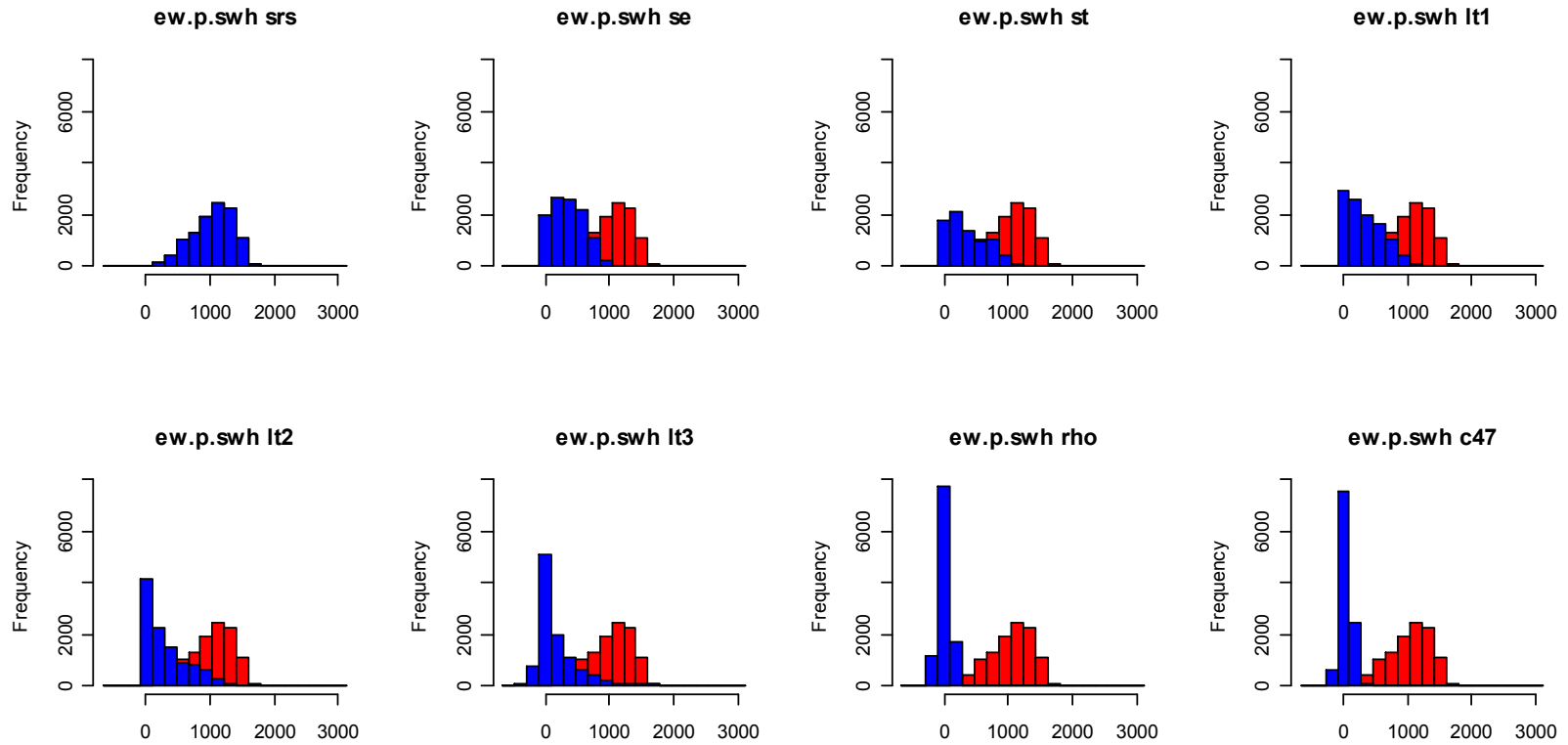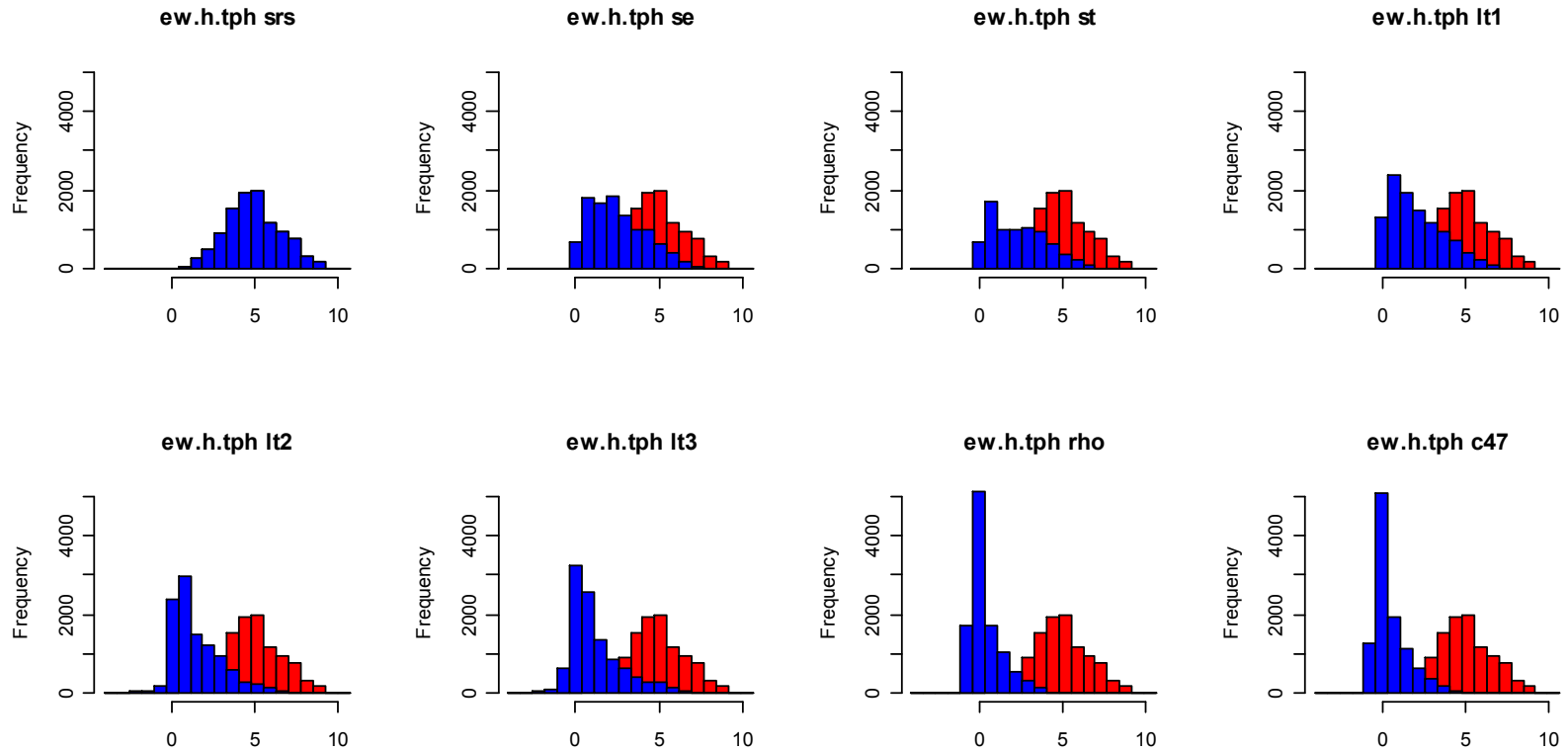
The above figures show Mean Square Error (MSE) for each of our eight estimators for Trees per Hectare (TPH) based on our population for north-south swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho 2$ and c47 refers to our $\rho 1$.
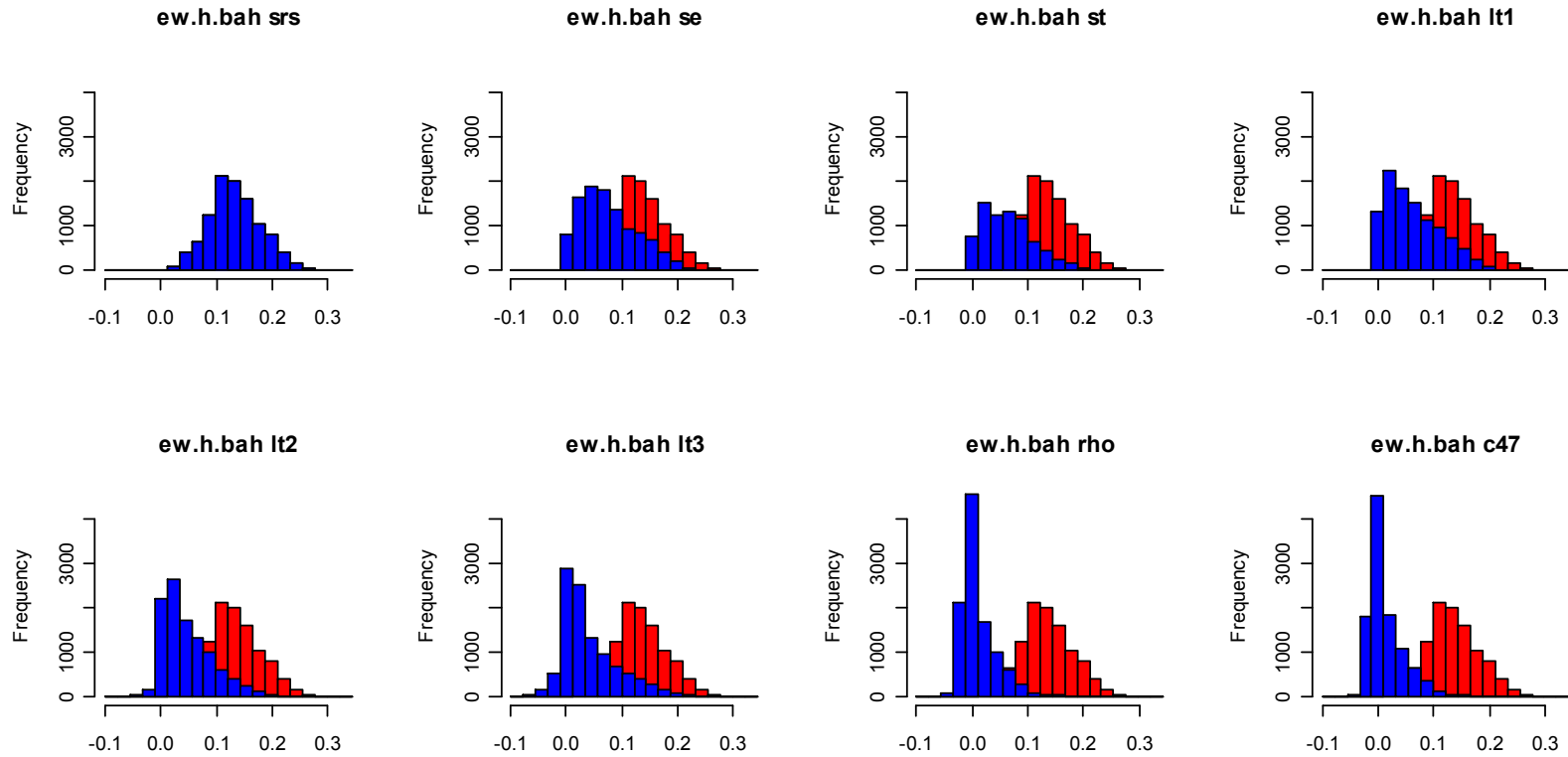
The above figures show Mean Square Error (MSE) for each of our eight estimators for Basal Area per Hectare (BAH) based on our population for north-south swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$.
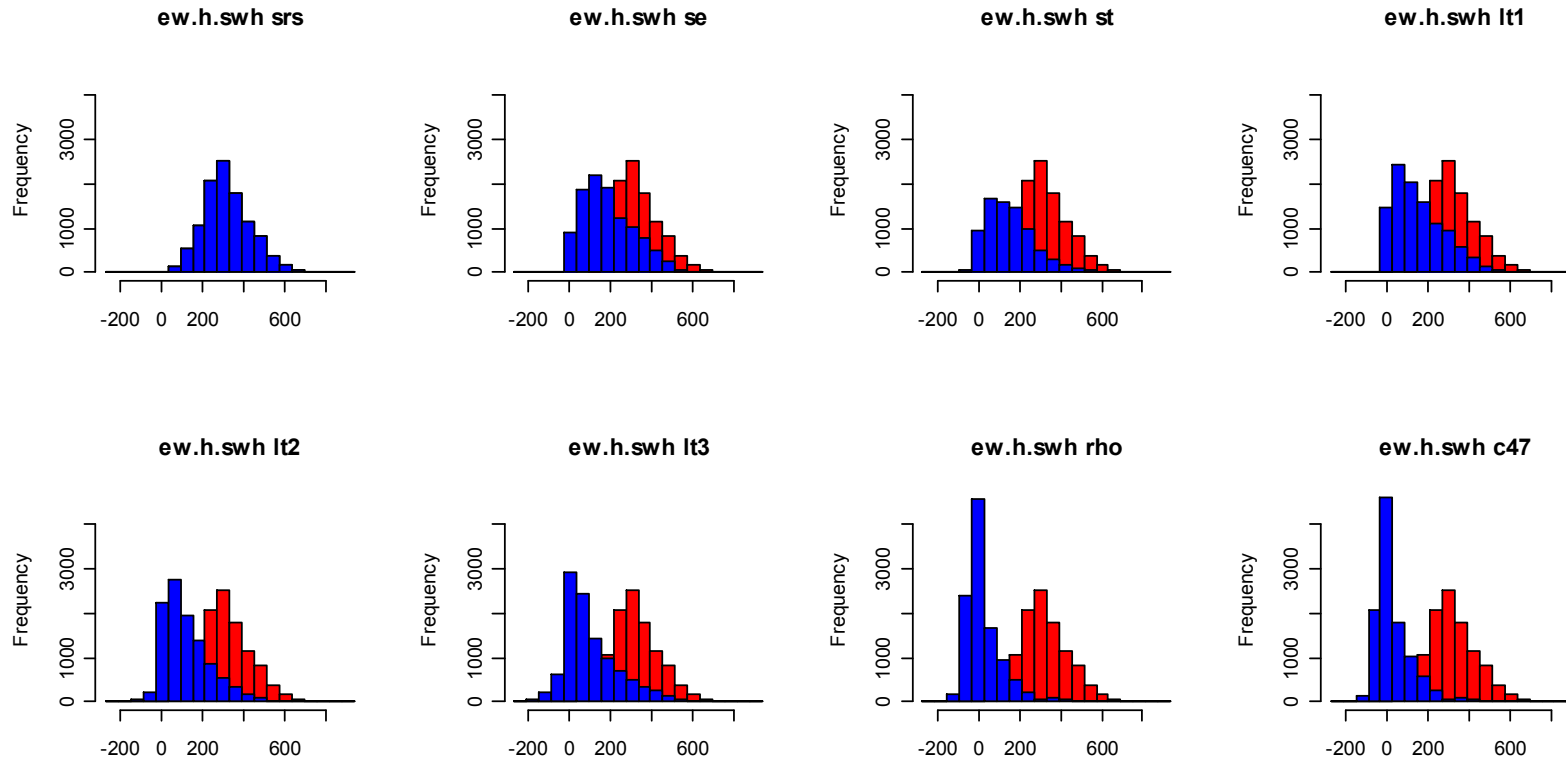
The above figures show Mean Square Error (MSE) for each of our eight estimators for Stem Wood Biomass per Hectare (SWH) based on our population for north-south swaths of hardwood stands. The upper-left figure shows the SRS estimator. This is repeated in red in the other figures for comparison. Note that estimator rho refers to our $\rho2$ and c47 refers to our $\rho1$.
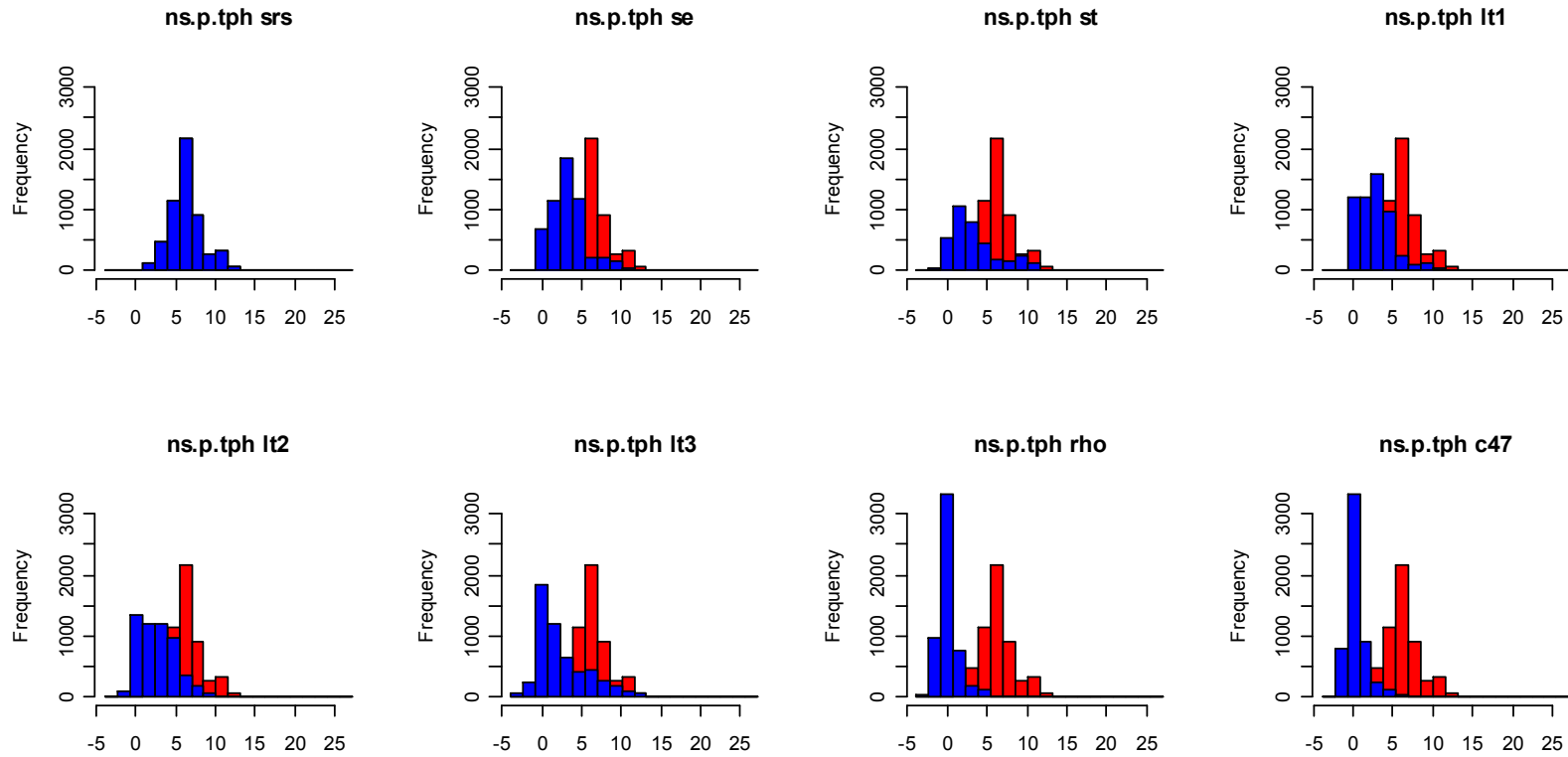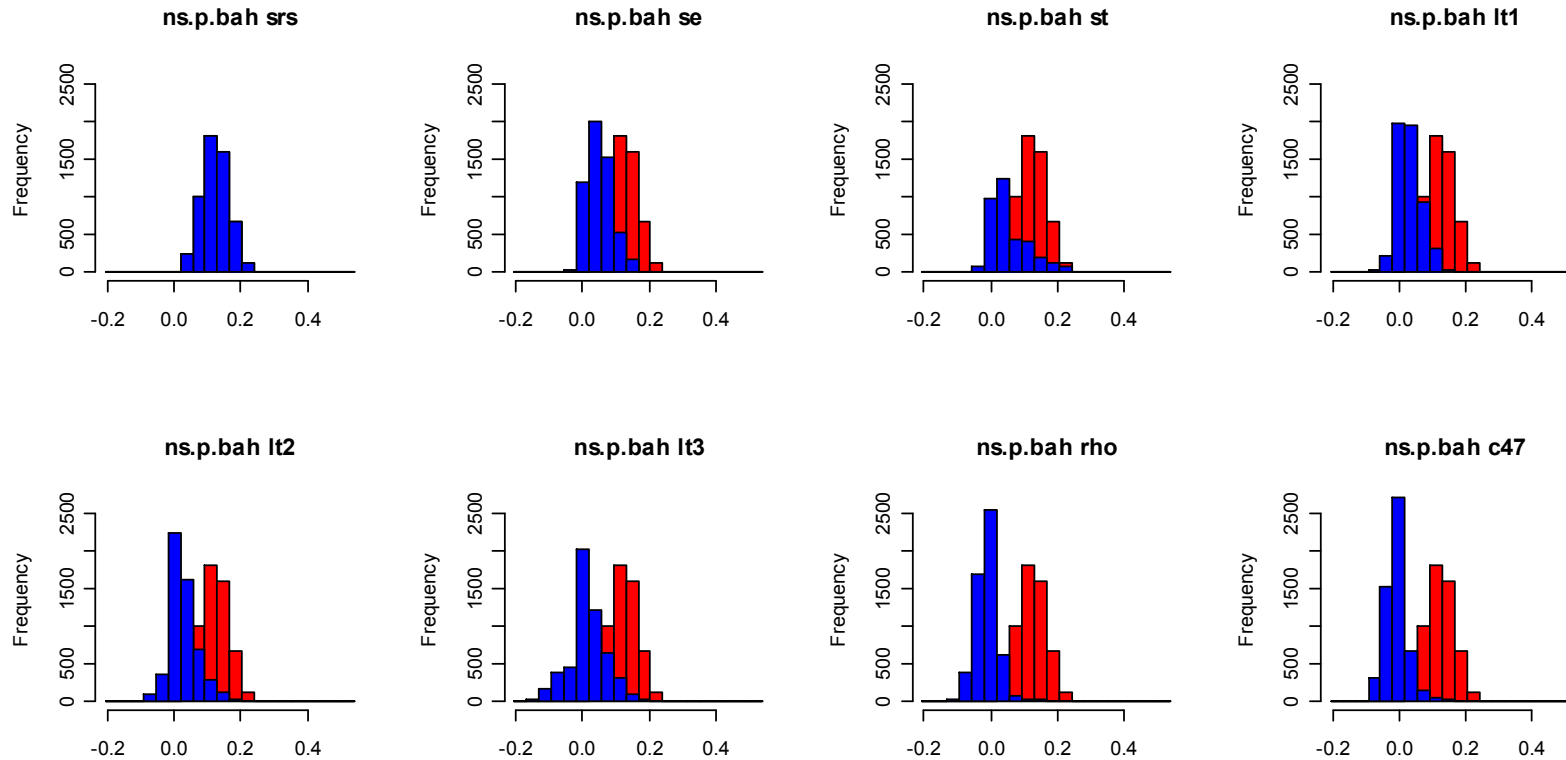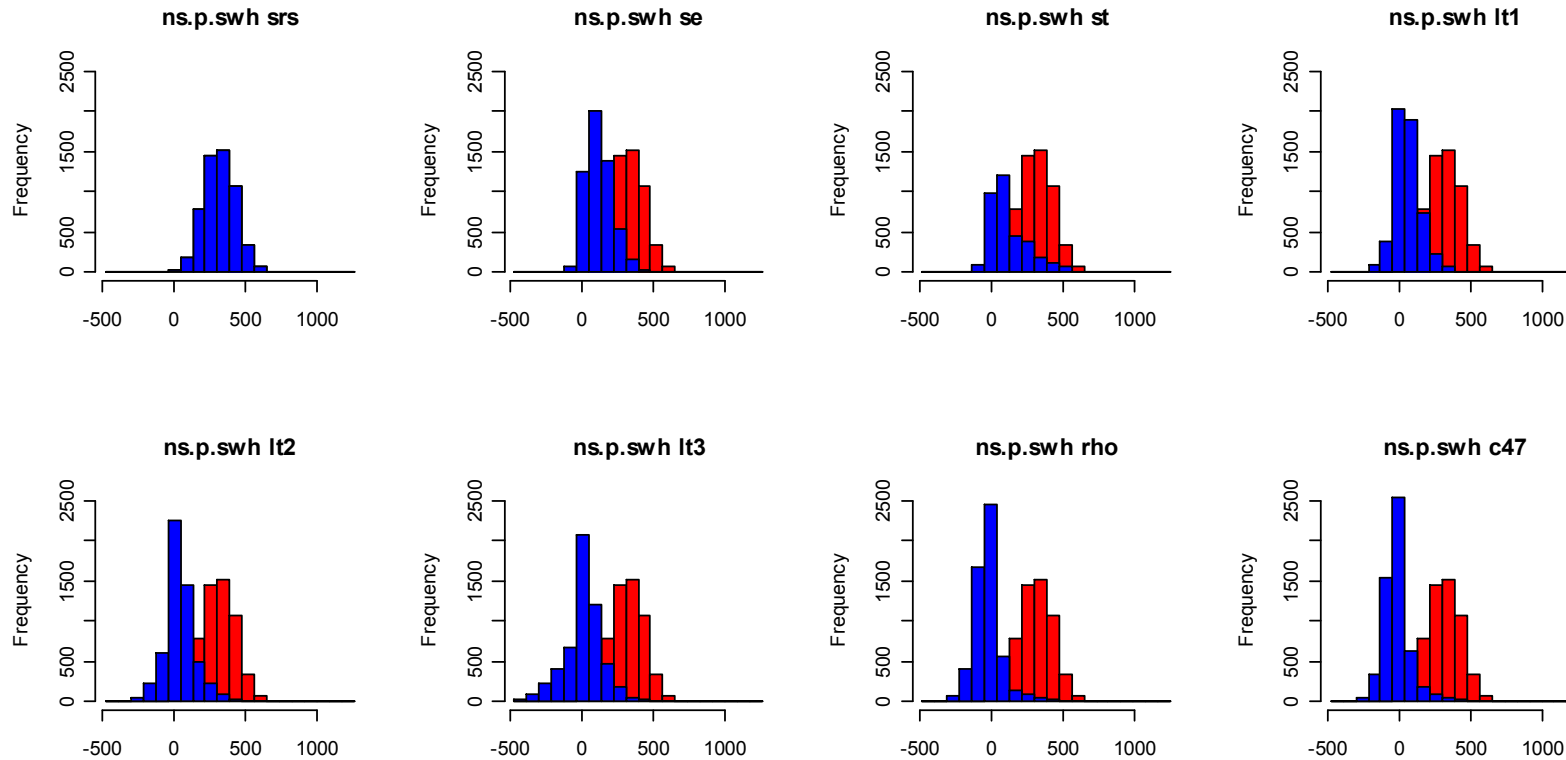
APPENDIX F

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	409.38	152.06	47.32	105228.57	4623.02	2314.65
240	70	35	99.11	12.29	4.30	8586.76	40.69	6.12
280	60	12	485.91	198.37	64.12	228411.13	43031.70	1536.59
280	60	15	357.42	120.70	37.66	94856.61	5875.69	225.94
280	60	20	234.23	56.98	18.36	35577.93	315.81	784.52
280	60	30	115.68	17.09	5.83	10685.60	41.29	42.18
300	56	14	383.28	140.59	43.69	101522.02	5036.08	877.42
300	56	28	124.16	19.55	6.67	12933.59	73.78	19.63
336	50	25	138.20	23.89	8.15	17376.40	334.22	3.87
350	48	12	451.09	179.33	57.62	214197.28	38227.55	2261.55
350	48	16	293.36	90.39	28.13	68763.45	4319.19	43.85
350	48	24	144.13	26.41	8.91	20481.38	599.33	47.52
400	42	14	333.11	118.14	36.94	67035.85	1978.95	2085.50
400	42	21	163.71	33.71	11.32	18723.93	33.05	312.86
420	40	20	171.92	37.85	12.54	22831.98	398.39	74.46
560	30	15	225.83	68.18	21.48	35581.90	1712.79	312.88
600	28	14	246.20	82.41	25.74	53233.37	4106.82	41.35
700	24	12	286.02	94.40	31.50	93425.52	10633.93	1008.13

E-W Pine - Trees per Hectare

E-W Pine - Basal Area per Hectare								
linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	0.4193	0.1592	0.0407	0.16032771	0.02059308	0.00044095
240	70	35	0.1005	0.0071	0.0028	0.00991126	0.00003943	0.00000342
280	60	12	0.4952	0.2102	0.0558	0.21843966	0.03703199	0.00085117
280	60	15	0.3649	0.1294	0.0333	0.11439650	0.01318260	0.00017900
280	60	20	0.2396	0.0550	0.0151	0.05169281	0.00183000	0.00000948
280	60	30	0.1178	0.0116	0.0041	0.01204244	0.00001269	0.00001577
300	56	14	0.3923	0.1474	0.0375	0.14226421	0.01732184	0.00042118
300	56	28	0.1261	0.0142	0.0048	0.01471531	0.00007781	0.00000036
336	50	25	0.1409	0.0188	0.0061	0.01937185	0.00028874	0.00001716
350	48	12	0.4615	0.1939	0.0512	0.19368053	0.03264447	0.00088986
350	48	16	0.3016	0.0973	0.0250	0.08671334	0.00986072	0.00037284
350	48	24	0.1477	0.0229	0.0070	0.02167763	0.00050285	0.00004253
400	42	14	0.3437	0.1239	0.0317	0.10753035	0.01158263	0.00021949
400	42	21	0.1679	0.0313	0.0091	0.02773433	0.00087165	0.00005265
420	40	20	0.1778	0.0385	0.0107	0.02920641	0.00100405	0.00001465
560	30	15	0.2340	0.0716	0.0190	0.04806775	0.00491852	0.00010067
600	28	14	0.2549	0.0890	0.0229	0.05917917	0.00611683	0.00012313
700	24	12	0.2977	0.1117	0.0297	0.08929384	0.01243586	0.00087103

E-W Pine - Stem Wood Volume per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	2714768.00	953198.53	234239.79	6817910000000	746551300000	15249051039
240	70	35	650424.00	46331.40	16488.09	422178300000	2161380000	239444080
280	60	12	3207395.90	1257012.63	319759.83	9239375000000	1258981000000	23121904333
280	60	15	2364343.90	778900.88	192949.65	4959050000000	518723800000	7426330383
280	60	20	1552410.60	339844.19	89204.10	2294461000000	92745770000	2691597926
280	60	30	762763.30	75189.93	24305.72	518311500000	1046113000	343935685
300	56	14	2538121.00	880757.51	216335.32	5964890000000	602804800000	11490638579
300	56	28	816328.30	92801.44	28607.31	632237300000	4803177000	28227450
336	50	25	913648.30	121157.52	36410.57	833895000000	15239260000	1142341744
350	48	12	2993453.90	1163445.09	294740.10	8128122000000	1093524000000	21160329670
350	48	16	1957050.80	588269.41	146247.19	3736054000000	386714100000	16604309071
350	48	24	957652.40	148068.80	42011.49	895814900000	19061400000	975974432
400	42	14	2227514.80	739143.90	182023.14	4576475000000	415916200000	7555703511
400	42	21	1088287.10	193734.69	53550.86	1188021000000	37393830000	2829306374
420	40	20	1153481.90	240162.36	63579.52	1268485000000	45048790000	1267495261
560	30	15	1515070.10	426025.97	109292.09	2101151000000	201027900000	4686236248
600	28	14	1649185.20	531041.25	131617.35	2519235000000	221604800000	4869560748
700	24	12	1930417.70	687277.27	174534.15	3715256000000	466351400000	29253288671

E-W Hardwood - Trees per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	69.75	32.89	13.38	4904.06	1100.18	179.55
240	70	35	16.69	2.37	0.90	278.53	5.70	0.84
280	60	12	80.92	38.58	17.17	5475.04	993.66	113.44
280	60	15	60.00	27.95	10.61	3606.47	815.98	118.96
280	60	20	39.24	14.40	4.75	1424.24	165.29	9.32
280	60	30	19.31	3.26	1.24	369.94	10.16	1.36
300	56	14	64.28	30.04	12.12	4213.39	945.32	151.91
300	56	28	20.63	3.85	1.44	377.01	7.08	0.15
336	50	25	22.96	5.20	1.82	500.48	21.30	1.52
350	48	12	75.50	35.33	15.56	4490.81	717.07	54.19
350	48	16	49.62	22.36	8.02	2494.26	541.87	66.13
350	48	24	24.26	6.54	2.15	561.61	35.94	2.58
400	42	14	55.65	25.40	10.13	3112.04	645.49	98.75
400	42	21	27.10	7.95	2.72	621.44	33.26	0.25
420	40	20	28.74	9.66	3.27	828.61	97.04	10.80
560	30	15	37.22	14.93	5.90	1385.40	261.50	41.79
600	28	14	40.82	17.85	7.03	1850.61	390.96	68.68
700	24	12	46.11	19.48	8.65	1506.92	163.44	8.05

				E-W Hardwood - Basal Area per Hectare				
linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	0.0542	0.0293	0.0118	0.00279821	0.00091791	0.00013161
240	70	35	0.0130	0.0024	0.0008	0.00016640	0.00000540	0.00000045
280	60	12	0.0633	0.0354	0.0204	0.00361901	0.00101024	0.00051101
280	60	15	0.0468	0.0246	0.0088	0.00218339	0.00059546	0.00007078
280	60	20	0.0306	0.0135	0.0041	0.00082443	0.00014118	0.00000409
280	60	30	0.0151	0.0034	0.0011	0.00021582	0.00000903	0.00000049
300	56	14	0.0502	0.0267	0.0106	0.00242384	0.00076436	0.00010084
300	56	28	0.0162	0.0042	0.0013	0.00024875	0.00001430	0.00000081
336	50	25	0.0179	0.0053	0.0016	0.00032023	0.00003022	0.00000269
350	48	12	0.0589	0.0324	0.0204	0.00309951	0.00080418	0.00054474
350	48	16	0.0385	0.0195	0.0067	0.00141587	0.00034544	0.00003303
350	48	24	0.0189	0.0065	0.0019	0.00035627	0.00004401	0.00000355
400	42	14	0.0432	0.0227	0.0089	0.00192320	0.00064075	0.00009239
400	42	21	0.0211	0.0077	0.0023	0.00038141	0.00004161	0.00000076
420	40	20	0.0222	0.0090	0.0027	0.00050257	0.00008076	0.00000716
560	30	15	0.0288	0.0133	0.0048	0.00082764	0.00018190	0.00002324
600	28	14	0.0319	0.0156	0.0063	0.00109057	0.00037138	0.00005559
700	24	12	0.0356	0.0177	0.0075	0.00110361	0.00018304	0.00001198

E-W Hardwood - Stem Wood Volume per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	70	14	323569.78	178667.92	72153.89	92339055151	32139813526	4069685019
240	70	35	77778.66	16382.10	4915.36	5912575327	243497897	16149824
280	60	12	378530.55	218727.12	120617.89	133330731719	40414280882	18008838115
280	60	15	279973.41	149429.72	52862.83	78428544752	21755359854	2467098085
280	60	20	182880.58	85025.20	25058.03	28482751851	5516643134	137231628
280	60	30	90118.08	22800.09	6780.47	7650855249	410628433	17029605
300	56	14	299525.98	162848.31	64548.08	79072325040	26018558795	2831428201
300	56	28	96709.55	27979.62	8106.19	9041214235	695437622	42132507
336	50	25	106985.77	35384.42	10194.48	11403209506	1357959640	106125205
350	48	12	351687.72	199878.23	120095.69	115360677130	32423475153	18893542535
350	48	16	229958.26	119030.11	40564.10	48669123928	12011141798	992283340
350	48	24	112690.99	42037.71	11791.15	12766196567	1929074066	147901689
400	42	14	257766.38	138266.35	54239.63	64713151509	23472730528	2965949221
400	42	21	125755.04	49430.06	14512.58	13716205070	1854811543	43202757
420	40	20	132636.35	56241.18	16700.15	17922396738	3294438941	274751742
560	30	15	171511.25	80953.39	28447.06	29461318504	6629069346	810789703
600	28	14	189629.32	95001.02	38001.51	36517616383	13548963676	1818178816
700	24	12	210633.81	108341.26	45414.45	41942795780	7590389524	587813554

N-S Pine - Trees per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	91.09	42.90	12.43	5096.00	838.28	93.35
270	40	20	49.82	19.33	5.10	2475.36	452.93	31.27
300	36	12	111.73	51.35	17.61	12125.68	2852.92	237.87
300	36	18	54.08	21.19	5.67	2326.47	476.66	17.72
360	30	15	66.19	29.37	8.58	4222.31	1062.25	55.54
450	24	12	79.40	36.50	12.15	5584.00	1372.87	77.95

N-S Pine - Basal Area per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	0.0522	0.0193	0.0099	0.00150590	0.00005211	0.00001494
270	40	20	0.0283	0.0074	0.0035	0.00079197	0.00005630	0.00000957
300	36	12	0.0632	0.0272	0.0172	0.00305148	0.00037192	0.00008620
300	36	18	0.0307	0.0083	0.0043	0.00055543	0.00001622	0.00001340
360	30	15	0.0379	0.0125	0.0068	0.00126261	0.00010245	0.00001867
450	24	12	0.0453	0.0196	0.0119	0.00155307	0.00019957	0.00003097

N-S Pine - Stem Wood Volume per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	356953.10	109158.27	71361.30	83651245568	2363748620	11888345
270	40	20	193501.30	37530.57	23875.45	36626356442	1236419818	397894822
300	36	12	434766.00	165680.30	134905.30	154453967894	13812440182	8093659155
300	36	18	211865.60	44410.51	30503.62	30307086485	164067383	133301838
360	30	15	260464.90	70785.58	49353.69	62087473227	3410201188	1329800224
450	24	12	313684.40	117276.44	92566.30	81485454741	6839165011	3128601581

N-S Hardwood - Trees per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	19.28	10.01	4.07	113.02	16.71	52.61
270	40	20	10.79	4.19	1.56	103.09	10.76	0.43
300	36	12	23.94	13.69	9.87	387.00	126.08	143.43
300	36	18	11.74	4.35	1.72	124.70	17.01	1.44
360	30	15	14.02	6.81	2.68	167.73	20.15	0.12
450	24	12	15.79	8.84	2.96	207.70	86.37	6.29

N-S Hardwood - Basal Area per Hectare

linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	0.0136	0.0074	0.0034	0.00006134	0.00000509	0.00002100
270	40	20	0.0077	0.0032	0.0012	0.00003253	0.00000143	0.00000065
300	36	12	0.0166	0.0102	0.0072	0.00013602	0.00012207	0.00008827
300	36	18	0.0082	0.0033	0.0014	0.00006201	0.00001259	0.00000163
360	30	15	0.0096	0.0050	0.0021	0.00009735	0.00001852	0.00000186
450	24	12	0.0107	0.0059	0.0020	0.00006552	0.00000861	0.00000124

N-S Hardwood - Stem Wood Volume per Hectare								
linesPerSwath	N	n	mean.s2.ran	mean.s2.lt1	mean.s2.wv8	mse.s2.ran	mse.s2.lt1	mse.s2.wv8
240	45	15	80649.94	48840.92	34820.46	2302521173	260223715	1127489048
270	40	20	45246.48	21196.92	8239.59	954681781	50151427	42680823
300	36	12	97609.89	63876.96	43599.37	4289725675	5105017199	3069917693
300	36	18	48276.00	22355.76	9330.25	2161633733	557399539	83296908
360	30	15	56570.01	32890.43	14832.06	3439945656	923188581	163591653
450	24	12	62007.24	35556.85	12783.47	2092618382	346168764	26200539

VITA

Name: Wesley Tyler Marcell

Address: 1500 Research Parkway, Suite B217,
                 College Station, TX 77843-2120

Email Address: wmarcell@neo.tamu.edu

Education: B.S., Texas A&M University, 2004