

2005

## A form of two-phase sampling utilizing regression analysis

Michael Allen Fiery Jr.  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Fiery, Michael Allen Jr., "A form of two-phase sampling utilizing regression analysis" (2005). *Graduate Theses, Dissertations, and Problem Reports*. 2304.  
<https://researchrepository.wvu.edu/etd/2304>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# **A Form of Two-Phase Sampling Utilizing Regression Analysis**

**Michael Allen Fiery, Jr.**

**Thesis submitted to the  
Davis College of Agriculture, Forestry, and Consumer Sciences  
at West Virginia University  
in partial fulfillment of the requirements for the degree**

**Master of Science  
in  
Forestry**

**John R. Brooks, Ph.D., Chair  
Ray R. Hicks, PhD.  
Jingxin Wang, PhD.**

**Division of Forestry  
Morgantown, West Virginia**

**2005**

**Keywords: Forest Inventory, Two-Phase Sampling, Double Sampling, Regression Analysis.**

## **Abstract**

### A Form of Two-Phase Sampling Utilizing Regression Analysis

Michael Allen Fiery, Jr.

A two-phase sampling technique was introduced and tested where sawlog heights were measured on intensive (second phase) points and predicted on non-intensive (first phase) points. Regression analysis is used to predict heights on first phase points in order to achieve an estimate of volume for every point. Results indicate an improved estimate of both mean volume per acre and volume per acre by species when compared to traditional double sampling. An unbiased sampling error was also achieved in this process. Datasets consisted of mostly hardwood species from West Virginia and Maryland.

## **Acknowledgments**

I consider myself extremely lucky to have spent my graduate studies under Dr. John R. Brooks. His endless hours of patience and assistance throughout my education has provided me with skills I will carry my entire career. I also would like thank the WVU Division of Forestry and the Maryland DNR for allowing access to datasets essential to the completion of my research. Also, I would like to thank committee members Dr. Jingxin Wang and Dr. Ray R. Hicks for their teaching and guidance during my coursework and research time at WVU. Lastly, I would like to thank my parents for their love and support throughout the years.

## Table of Contents

Introduction	1
Chapter 1: Review of Literature	3
History of Forest Inventory in North America	3
Height Regression Sampling	4
Chapter 2: Methods	6
Description of Study Sites and Inventory Specifications	6
Tygart Tract	6
Coopers Rock State Forest and WVU Research Forest	7
Compartment 14	7
1967 Coopers Rock Inventory	8
Trout Pond	8
Savage River	8
Procedures for Data Analysis	8
Sampling Systems	9
Species Grouping for Height Prediction	11
Single Random Inventory	11
Analysis through Random Simulations	12
Chapter 3: Results	13
Single Random Inventory	13
Random Simulation of 500 Inventories	22
Chapter 4: Conclusions	30
Bibliography	32
Appendix	33

## Introduction

Forest inventory has undergone many changes over the past few decades. One of the major influences has been the desire to reduce field data collection time without sacrificing the accuracy of sample based estimates of trees per acre, basal area, weight, and volume. The reduction in fixed area plot size and the transition to horizontal point sampling has successfully reduced the number of in-trees per sampling unit. When double sampling with point sampling was first applied to forest inventory, a large time savings was realized since the variable of interest (usually volume) was only measured on a subsample of points. One drawback to double sampling is dbh and species are only recorded on second phase points, permitting no direct method of creating stand and stock tables, though procedures were developed for their estimates. Should an inventory require more accurate diameter distribution table estimates, other methods are available to complete this task without the use of intensive tree measurements on every sampling unit. An inventory system can be designed where dbh and species are tallied on all points, while more accurate measurements for volume can be conducted on fewer sampling units. This design would permit an accuracy equivalent to the intensive measurement of all sampling units for stems and basal area per acre. It also provides a means to disaggregate total volume to volume by species and size class. In areas where there are large variations in species value, it becomes increasingly important to be able to accurately estimate volume associated with each species. Under the proposed sampling system, dbh, species, and sawlog merchantable height would only be measured on second phase samples. On all non-intensive (first-phase) samples, only dbh and species would be recorded for each "in-tree." Through regression analysis, all heights necessary for volume estimation would be predicted for all non-intensive samples on a species or species group basis. Although this process is more time consuming than

the use of traditional double sampling, there are instances when the additional field time could be justified. Although this technique has been employed in other regions, no published record of the effects of height prediction on the accuracy and precision of typical volume estimates has been found. The research that follows will:

1. Outline an inventory system where boardfoot volume is known (measured) on intensive (second phase) points and estimated on non-intensive (first phase) points using height prediction based on regression analysis.
2. Evaluate the differences between stand and stock tables based on traditional double sampling with point sampling, the proposed height regression system and a traditional inventory where height and dbh are measured on every point.
3. Through computer simulation, provide an evaluation of the mean volume per acre and the sampling error associated with traditional double sampling using basal area per acre as the auxiliary variable and the proposed height regression system.

## **Chapter 1: Review of Literature**

### **History of Forest Inventory in North America**

As early as the 1920's sampling error was being calculated from strip cruises in Norway and Sweden (Spurr 1952). Soon after, line-plot sampling was adopted in the United States and systematic sampling became the norm for most inventory situations. Plot sizes were initially between at 0.20 and 0.25 acre (Johnson and Hixon 1952). By the 1970's smaller plot sizes were employed as they provided similar estimates and required less field measurement time per sample. Optimum plot size will vary depending on species, age, and spatial patterns but need not exceed 0.1 acre for most sawtimber inventories (Mesavage and Grosenbaugh 1956). The concept of point sampling was developed by Bitterlich, an Austrian Forester, and was introduced in the United States as a complete inventory system by Grosenbaugh in 1952. In 1965, Avery and Newton suggested that using a point sample basal area factor (BAF) of 10 for estimating volume on bottomland hardwoods would provide similar results to the use of 0.1 acre fixed area plots. Thus point sampling became more commonly used among foresters due mainly to the reduced field time required for an inventory. In the years to follow, larger BAF's of 20 or 40 were recommended for sawtimber cruises in Appalachia (Wiant et al. 1984). In addition to the use of smaller fixed area plot sizes and larger BAF's, the topic of two-phase sampling was also introduced as an efficient inventory technique. One of these two-phase sampling methods was double sampling, which although mentioned by Spurr (1952), was first outlined by Freese in 1962. With the introduction of double sampling, the cost of an inventory could be greatly reduced while providing an unbiased estimate of mean volume per acre and associated sampling



error of similar magnitude. When determining a ratio of intensive/non-intensive points, a ratio of 1 in 4 (25% intensive) has been accepted for most double sample inventories (Bailes 2004).

## Height Regression Sampling

From the standpoint of field measurement time, the use of height regressions to predict tree heights and thus volumes on a subset of the total number of sampling units require less measurement time than the measurement of volume on all sampling units. On the other hand, it is more time consuming than a double sample point sample inventory employing basal area per acre as the auxiliary variable, since diameter measurements are recorded for every “in-tree” on all sampling units. The additional time required to measure dbh on all “in-trees” can be offset by the increased diameter distribution information obtained. Although this technique of predicting heights on a subsample of points to obtain volume has been widely applied in the South, very little has been published about this technique. The overall effect on mean volume per acre when using some sampling units with measured heights and others with estimated heights is unknown. The variance of the mean is most likely reduced, as the natural variation in heights by diameter and thus on volume, has been removed through the prediction process. The magnitude of this effect is also unknown.

The following commonly used height model (Avery and Burkhart, 2002) has been used for a wide range of species and can be applied to both merchantable and total height:

$$\ln H_i = b_0 + b_1 D_i^{-1} + \ln \varepsilon_i \quad (1)$$

where:

H = total height of tree "i"

D = diameter at breast height for tree "i"

$b_0, b_1$  = constants to be determined from the data

$\varepsilon$  = random error term

This linear transformation of a typical sigmoidal relationship is employed in this study.

## Chapter 2: Methods

### Description of Study Sites and Inventory Specifications

Datasets from several areas in West Virginia and Maryland were available for analysis in this study. Each dataset included measurements of species, dbh, and sawlog height on every point which permitted the comparison of both two-phase sampling methods (double sampling and height regression sampling) to estimates where all “in-trees” were intensively measured on all sampling units.

#### Tygart Tract

The Tygart dataset was collected in the summer of 2004 from the Tygart Tract located in Dailey, WV. The tract is approximately 10 miles south of Elkins, WV and approximately 426 acres (Table 2) were inventoried. Primary species consisted of red maple (*Acer rubrum* L.), northern red oak (*Quercus rubra* L.), and chestnut oak (*Quercus prinus* L.). The original dataset consisted of 67, 1/5 acre circular plots where species, DBH (nearest 0.1 inch), sawlog height (0.1 foot) and total height (0.1 foot) were measured and recorded. Sawtimber boardfoot volumes (International 1/4 inch) were based on a 9-inch merchantable top diameter for hardwoods and an 8-inch merchantable top diameter for conifers, except where sawlog height was limited by stem form. Plots were allocated using a systematic 7 x 7 chain grid. All living trees were tallied if dbh was 3.6 inches or larger and sawlog height was measured on all sawtimber trees 11.6 inches dbh or larger. Distance from plot center to every “in-tree” was also collected to the nearest tenth of a foot. All measurements for height and distance from plot center were recording using an Impulse laser.

## Coopers Rock State Forest and WVU Research Forest

In the year 2000, an inventory was conducted at both Coopers Rock State Forest and the WV University Research Forest in Monongalia and Preston Counties, WV. Primary species found in this inventory included yellow-poplar (*Liriodendron tulipifera* L.), northern red oak, red maple, chestnut oak, and black cherry (*Prunus serotina* Ehrh.). These datasets were based on a BAF 20 point sampling inventory on a systematic grid. For each point, species, dbh (nearest 0.1 inch), and number of 16 foot logs (nearest ½ log) were tallied for every living tree. Sawtimber boardfoot volumes (International ¼ inch) were based on a 10-inch merchantable top diameter for hardwoods and an 8-inch merchantable top diameter for conifers, except where sawlog height was limited by stem form. Diameters were measured using a diameter tape and heights were recorded using a Spiegel relascope.

The data was separated into three different administrative units; WVU Research Forest, Coopers Rock State Forest, and Coopers Rock State Forest Annex. The WVU Research Forest portion contained 2,013 points while the Coopers Rock State Forest contained 1,081 points and Coopers Rock Annex portion contained 98 points (Table 2).

### Compartment 14

The Compartment 14 dataset comes from a 138 acre compartment of the WVU Research Forest located in Monongalia County, WV, inventoried in 2004 (Table 2). Primary species consisted of chestnut oak, white oak (*Quercus alba* L.), and yellow-poplar. It consists of BAF 20 point sample data from 52 points where species, dbh (nearest inch), and number of 16 foot logs (whole and ½ logs) were tallied for every living “in-tree” that was at least 11.6 inches dbh.

### **1967 Coopers Rock Inventory**

This inventory was conducted in 1967 on a portion of the WVU Research Forest totaling 3,500 acres and consisted of 384 BAF 10 points where dbh and number of 16-foot logs (nearest  $\frac{1}{2}$  log) were collected on every point. Species data was not available for this dataset. Only sawtimber size trees 11.6 inches or larger were inventoried.

### **Trout Pond**

This inventory was conducted near the trout pond adjacent to the WVU Research Forest. Primary species consisted of northern red oak, yellow-poplar, and black cherry. The dataset consists of 30 square 0.4 acre plots where species, dbh (nearest 0.1 inch), number of 16-foot logs (nearest  $\frac{1}{2}$  log), and distance from plot center to every “in tree” was recorded to the nearest 0.1 foot.

### **Savage River**

The Savage River dataset comes from the Savage River State Forest in Garrett County, MD. Primary species consisted of red oak, red maple, and chestnut oak. This dataset consisted of 214,  $\frac{1}{5}$  acre circular plots which formed the basis of a continuous forest inventory system located throughout the 53,473 acres of forest. At each plot, species, dbh (nearest 0.1 inch), and number of 8-foot logs were tallied for every “in-tree”. Distance from plot center to every “in-tree” was also recorded to the nearest foot.

### **Procedures for Data Analysis**

This study will focus on inventory procedures utilizing horizontal point sampling. Each dataset used in this study is based on point sampling using a BAF of 10 or 20 or in three cases a BAF 20 inventory was created based on the selection of appropriate “in trees” at each point

utilizing the measured distance to every “in tree” and the limiting distance from the diameter measurement. Each dataset, regardless of the data collection procedures, will be referred to as a point sampling inventory.

### **Sampling Systems**

Since all “in-trees” on all points were measured in each dataset, trees per acre, basal area per acre, mean boardfoot volume per acre, and the standard error for boardfoot volume were calculated based on procedures for horizontal point sampling using simple random sampling (SRS) statistics as described by Shiver and Borders (1996).

Procedures for double sampling were conducted for point sampling as described by Shiver and Borders (1996) with basal area as the auxiliary variable. A ratio of means estimator was used to establish mean boardfoot volume per acre and the standard error of the mean. A 1:4 ratio was used where one out of every four points was intensively measured.

Height regression sampling (HRS) was conducted on each dataset using the same 1:4 ratio used in the double sampling inventory. Procedures for sampling with height regressions require a dataset where species and dbh were recorded for every “in-tree” on every point while merchantable sawlog height was only measured on intensively measured points. The following linear height model was used to obtain parameter estimates by species or species group:

$$\ln(MHT) = \beta_0 + \beta_1 \left( \frac{1}{DBH} \right) \quad (2)$$

With parameter estimates obtained for each species group, heights can be predicted on both intensive and/or non-intensive points.

Two different methods for data analysis were conducted and evaluated for height regression sampling each producing a different set of results. Sampling error for Method 1 is

statistically unsound while Method 2 provides a more adequate sampling error. Method 1 (srs method) for height regression sampling calculates volume estimates using simple random sampling statistics. The assumption is made that, the error associated with predicted heights on non-intensive points is minimal and that the estimates of volume per acre and the associated standard error, can be estimated using simple random sampling statistics.

Method 2 (ratio method) for height regression sampling uses an estimate of volume (using predicted heights for all trees) for each point. The actual volume on just the intensively measured points is based on measured tree heights. At this point the dataset can be treated as a double sample where actual volume is the variable of interest and estimated volume is the auxiliary variable. The mean volume per acre and standard error can be calculated using equations 3 and 6 respectively. A ratio estimator was used to find the mean volume per acre:

$$\bar{y}_{hrs} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \times \frac{\sum_{i=1}^{n'} x_i}{n'} = \hat{R} \bar{x}' \quad (3)$$

where:

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$$

$\bar{y}_{hrs}$  = average volume per acre

$y_i$  = actual volume per acre on intensive point 'i'

$x_i$  = predicted volume per acre on point 'i'

$n$  = number of intensive points

$n'$  = total number of points

The variance of the ratio can be calculated from the equation:

$$S_{\hat{R}}^2 = \frac{\sum_{i=1}^n y_i^2 - 2\hat{R}\sum_{i=1}^n x_i y_i + \hat{R}^2 \sum_{i=1}^n x_i^2}{n-1} \quad (4)$$

The overall variance can be calculated by:

$$S_{\bar{y}_{hrs}}^2 = \frac{S_y^2}{n'} + \frac{S_{\hat{R}}^2}{n} \left( \frac{n' - n}{n'} \right) \quad (5)$$

The standard error can be found by taking the square root of the overall variance:

$$S_{\bar{y}_{hrs}} = \sqrt{S_{\bar{y}_{hrs}}^2} \quad (6)$$

### Species Grouping for Height Prediction

From a biological standpoint, height prediction can be improved when sawlog heights are estimated on a species or species group basis. When there were an ample number of sample trees for a particular species, a separate set of parameters were determined. In this analysis, a minimum of five observations was required to predict species specific parameter estimates. For any species having four or fewer measured heights, height predictions were based on parameters established from the “other” species group.

### Single Random Inventory

A single systematic inventory was conducted on each dataset for simple random sampling, double sampling, and height regression sampling (methods 1 and 2). A random number generator was employed for double sample and height regression sampling to select 25 percent of the points as second phase samples while the balance were considered as first phase samples. Measurements taken and available on non-intensive points not required by the two-phase sampling systems were subsequently ignored. In cases where the top diameter was unknown, a 9-inch dob merchantable top was assumed for all hardwoods and a 7-inch dob top



was used for softwoods. Boardfoot volume was calculated on an individual tree basis using the equation published by Scott (1979) for International ¼ inch log rule.

### **Analysis through Random Simulations**

Since each dataset includes dbh and height measurements on every tree, simple random sampling was conducted for a traditional point sampling inventory where every “in-tree” was intensively measured (species, dbh, and sawlog height). The results of this simple random sample were assumed to be the actual values and the results from the other inventory systems were compared to this estimate. In order to quantify the variation in the mean and standard error for double sampling and height regression sampling, a Visual Basic 6.0 program was written to conduct simulations where different points were selected at random (using a random number generator without replacement) as second phase samples for each simulation. Once again a 1:4 ratio was used where 25% of points were second phase. Simulations were conducted 500 times for both double sampling (basal area as auxiliary variable) and height regression sampling. Each simulation provided an estimate of trees per acre, basal area per acre, boardfoot volume per acre, and the standard error for boardfoot volume per acre. Although 3,000 simulations is more commonly used, 500 simulations for each dataset provided adequate results since additional simulations resulted in no change in the distribution of mean or standard error.

## Chapter 3: Results

### Single Random Inventory

Predicted and actual heights were documented for each dataset in order to test the accuracy of height prediction. The results of the regression process indicate a slight increase in variance with increasing tree size although the distribution of merchantable height errors appeared random (Figure 1). Additional inventories provided similar results (Figures 9-15, Appendix). Height prediction for the most common species on the Tygart Tract (white oak) is shown in Figure 2. Results from the other inventories tested were similar (Figures 16-22, Appendix). In most cases, predicted heights were within ten feet of the actual height at least 60% of the time (Table 1).

Volume was compared on intensive points to see how close the relationship was between actual and predicted volume. Actual volume was plotted over predicted volume for intensive points (Figure 3) which shows a slight increase in variance with increasing volume. The correlation coefficient was 0.96 or higher for all inventories (Figures 23-29, Appendix). With a high correlation between predicted volume (auxiliary variable) and actual volume (variable of interest) the ratio estimator used in HRS method 2 for the mean and standard error should be appropriate. At the same time, with such a similarity between predicted and actual volume, HRS method 1 should provide an appropriate estimate of the mean.

Results for mean boardfoot volume per acre were summarized for the datasets with respect to each sampling system conducted. The results are listed in Table 2 along with their difference from the SRS mean. In seven of the eight inventories, the HRS method 2 mean was

closer to the SRS mean than double sampling. Five of the eight inventory means for HRS method 1 provided an estimate closer to SRS.

The mean boardfoot volume per acre for each inventory (Table 2) tends to display a less biased mean with height regression sampling than with double sampling. The sampling error for these inventories was calculated and listed in Table 3. In each of the eight of the inventories, the height regression sampling (methods 1 and 2) standard error is lower than that for double sampling. The standard error was also lower than SRS in all eight inventories for HRS method 1 and lower in two of the eight inventories for HRS method 2. When boardfoot volume for the Tygart Tract was evaluated at the species group level, height regression sampling provides a much more accurate estimate of volume (Table 4). Results for all other inventories were similar (Tables 6-11, Appendix).

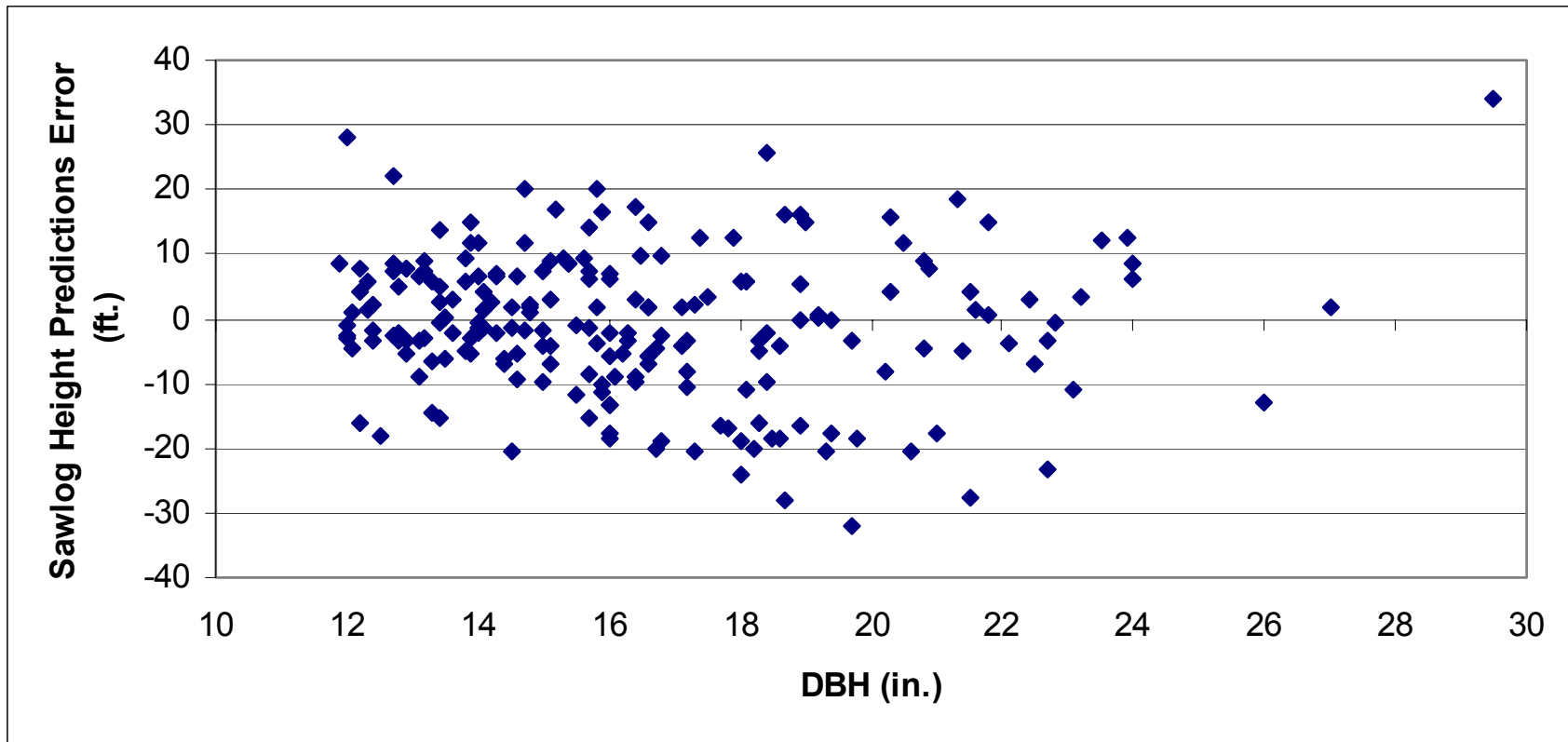


Figure 1. Height prediction error in feet for all species by DBH with height regression sampling of the Tygart Tract. (n=209)

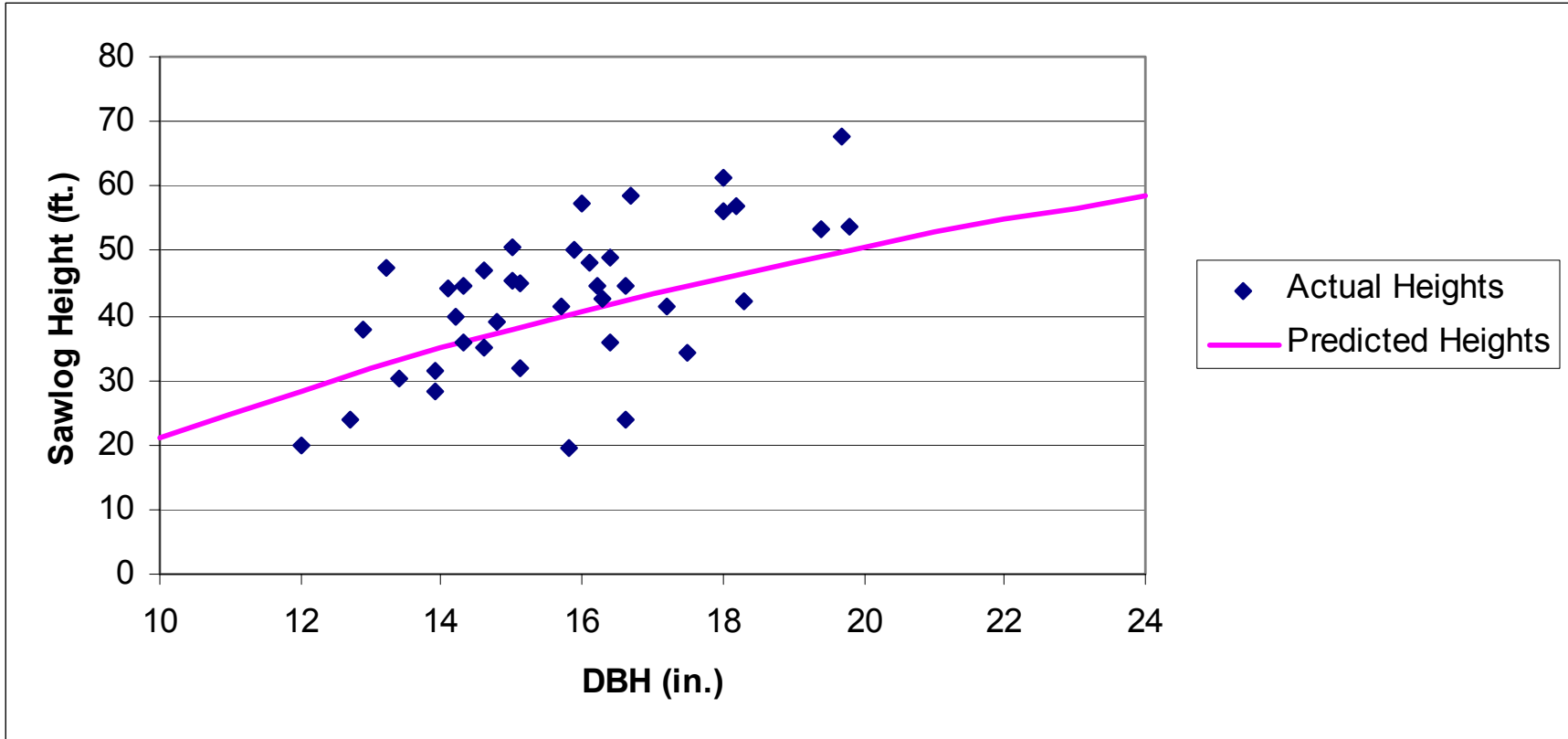
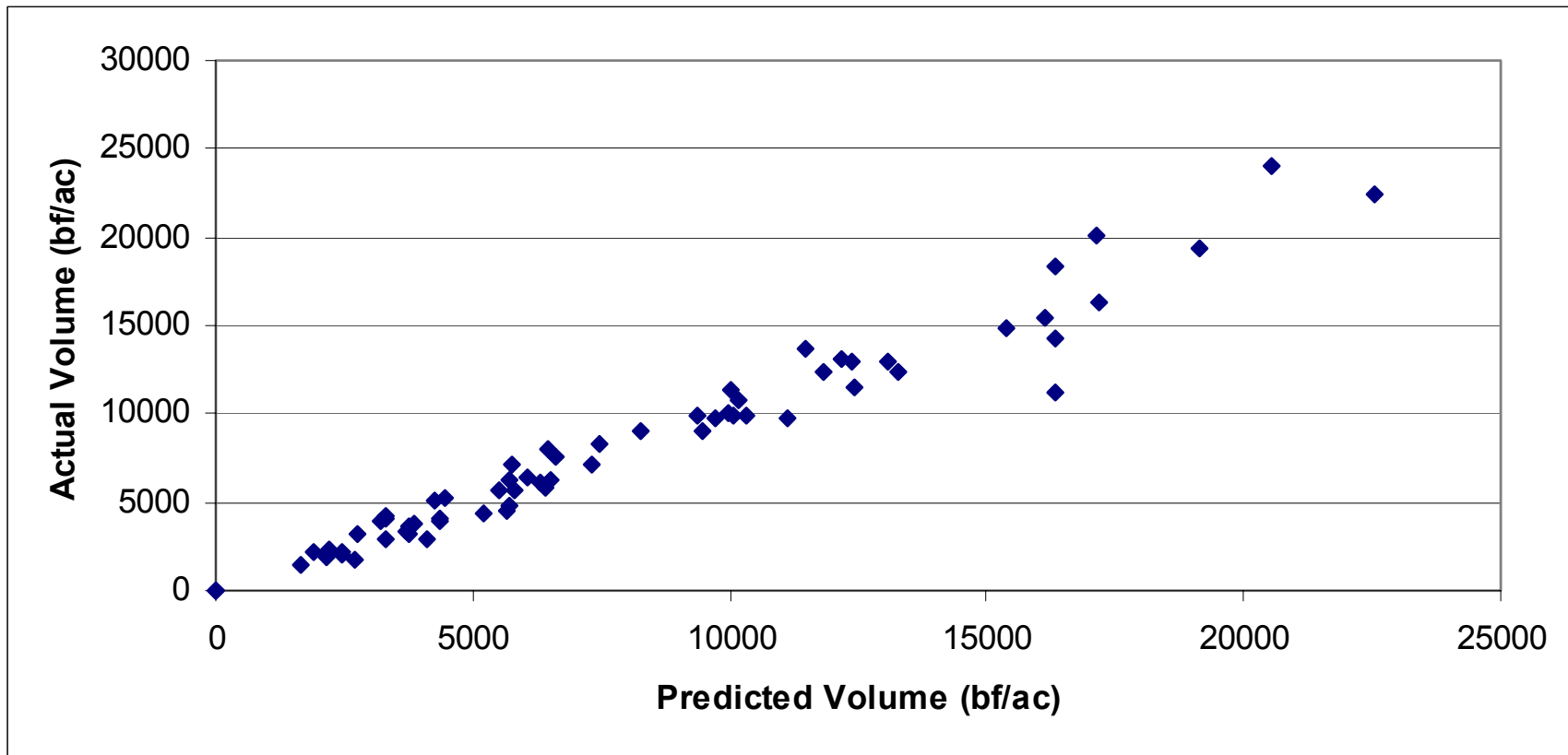


Figure 2. Actual and predicted sawlog heights in feet for white oak (*Quercus alba* L.) with height regression sampling of the Tygart Tract. (n=39)

**Table 1. Percentage of predicted heights within 10 feet of actual measured heights.**

<b>Tract</b>	<b>n</b>	<b>within 10 feet</b>	<b>% within 10 feet</b>
Tygart	210	144	68.57
WVU Reseach Forest	8,052	5,331	66.21
Coopers Rock State Forest	4,892	2,158	44.11
Coopers Rock Annex	392	240	61.22
Compartment 14	254	158	62.2
1967 single species	1,253	966	77.09
Trout Pond	178	132	74.16
Savage River	932	630	67.60



**Figure 3. Relationship between actual and predicted volumes for all points on the Tygart Tract (Correlation Coefficient = 0.979).**

**Table 2. Mean boardfoot volume per acre (International ¼ inch) for simple random sampling, double sampling, and height regression sampling (methods 1 and 2).**

Tract	Acres	Points	Mean Boardfoot Volume Per Acre				Difference from SRS		
			SRS	DS	HRS (srs)	HRS (ratio)	DS	HRS (srs)	HRS (ratio)
Tygart	426	67	7,883.4	7,935.5	7,834.8	7,840.7	52.1	-48.6	-42.7
WVU Research Forest	7,594	2013	9,676.1	9,730.9	9,520.8	9,640.9	54.7	155.4	-35.3
Coopers Rock State Forest	4037	1081	10,643.7	10,426.8	10,589.0	10,743.4	-217.0	-54.8	99.7
Coopers Rock Annex	364	98	10,341.9	9,945.2	10,511.0	10,418.4	-396.8	169.0	76.5
Compartment 14	138	52	11,076.6	10,909.2	11,209.8	11,053.8	-167.4	133.2	-22.7
1967 single species	3,500	384	2,539.4	2,531.3	2,473.0	2,546.5	-8.0	-66.3	7.1
Trout Pond	N/A	30	14,591.6	14,988.3	13,899.0	13,718.9	396.5	-692.9	-872.9
Savage River	53,473	214	10,363.2	8,359.0	10,551.4	10,732.6	-2,004.2	188.2	369.4



**Table 3. Standard error associated with boardfoot volume per acre for simple random sampling, double sampling, and height regression sampling (methods 1 and 2).**

Tract	Standard Error (bf/ac)				Standard Error as percentage of the mean			
	SRS	DS	HRS (srs)	HRS (ratio)	SRS (%)	DS (%)	HRS (srs) (%)	HRS (ratio) (%)
Tygart	683.8	1,055.2	659.2	662.0	8.67	13.30	8.45	8.44
WVU Research Forest	153.9	230.6	145.5	159.7	1.59	2.37	1.68	1.66
Coopers Rock State Forest	200.1	343.4	192.4	214.6	1.88	3.29	2.03	2.00
Coopers Rock Annex	653.2	911.9	626.2	670.5	6.32	9.17	6.38	6.44
Compartment 14	916.3	1,534.8	894.6	890.7	8.27	14.07	7.95	8.06
1967 single species	130.9	152.3	119.9	146.7	5.16	6.02	5.93	5.76
Trout Pond	1,254.8	1,717.8	1,190.8	1,354.4	8.60	11.46	9.74	9.87
Savage River	481.0	1,840.8	480.4	502.2	4.64	22.02	4.76	4.68

**Table 4. Boardfoot volume per acre by species group for each sampling method (Tygart Tract).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Mixed Oak	2,357.9	2,421.0	2,513.8	2,511.9	2.68	6.61	6.53
Other Conifers	235.9	130.4	229.1	229.0	-44.72	-2.88	-2.92
Other Hardwoods	1,040.9	1,036.6	1,058.0	1057.3	-0.41	1.64	1.58
Red Maple	432.4	271.6	422.1	421.8	-37.19	-2.38	-2.45
Red Oak	838.6	866.2	893.1	892.4	3.29	6.50	6.42
Sugar Maple	544.5	1,321.6	537.3	536.9	142.72	-1.32	-1.40
White Oak	1,272.5	1,260.3	1,206.2	1205.3	-0.96	-5.21	-5.28
Yellow-poplar	1,160.8	627.8	986.9	986.2	-45.92	-14.98	-15.04

## Random Simulation of 500 Inventories

Although a single inventory conducted on each of these nine datasets helps display trends and tendencies for each of the sampling types, further investigation can be conducted with repeated simulations of these datasets for both double sampling and height regression sampling. Predicted height curves for the first 20 of the 500 simulations were documented for the dominant species which for the Tygart Tract is white oak (Figure 4). Additional inventories provided similar results (Figures 30-36, Appendix). The majority of those curves appear reasonable, however, in rare cases a negative height curve can be obtained where sawlog height decreases with increasing dbh. In this instance, the mean sawlog height for all sample trees for that species group was used.

For the 500 unique inventories, the assumption was made that the SRS mean represents the average value. Average bias and root mean squared error for mean boardfoot volume were documented for each dataset with regards to double sampling and height regression sampling (Table 5, Figures 5-6). Average bias is lower than double sampling in only two of the eight datasets for HRS method 1 while it is lower for five of the eight inventories with HRS method 2. Root mean squared error is lower than double sampling in all eight of the inventories for both height regression sampling methods. Also note that root mean squared error is slightly lower for HRS method 2 than HRS method 1 in all eight inventories (Figure 6).

The variation in mean boardfoot volume for Tygart Tract is much higher when double sampling is conducted than with height regression sampling methods 1 and 2 (Figure 7). Both methods 1 and 2 for height regression sampling provided a similar range in means for the Tygart Tract and most of the other inventories tested (Figures 37-43, Appendix). However, in the two largest inventories (Figures 40 and 41, Appendix), the 500 means are centered on the actual SRS

mean in method 2, while they are noticeably below it in method 1. In general, the range in distribution of mean volume per acre decreased as number of inventory points increased (DS and both HRS methods). A Dunnett's test was also conducted to determine mean separation between the SRS mean and other sampling methods. In four of the datasets the F-value was found to be not significant and in the other four datasets, each sampling method tested (DS, HRS methods 1 and 2), were found to have no significance difference from the SRS mean.

The variation in standard error for the Tygart Tract using the same 500 simulations was also noted (Figure 8). All 500 inventories of this dataset provide a standard error with double sampling that is higher than the SRS standard error. The same general trend occurs for the other datasets (Figures 44-50, Appendix). The standard error for both HRS methods tends to be distributed above and below the actual standard error for simple random sampling. In HRS method 1 the standard error distribution covers a much smaller range than with HRS method 2. The same differences between method 1 and method 2 occur for all other datasets. Also note that in many instances the calculated standard error is lower with height regression sampling (methods 1 and 2) than with simple random sampling.

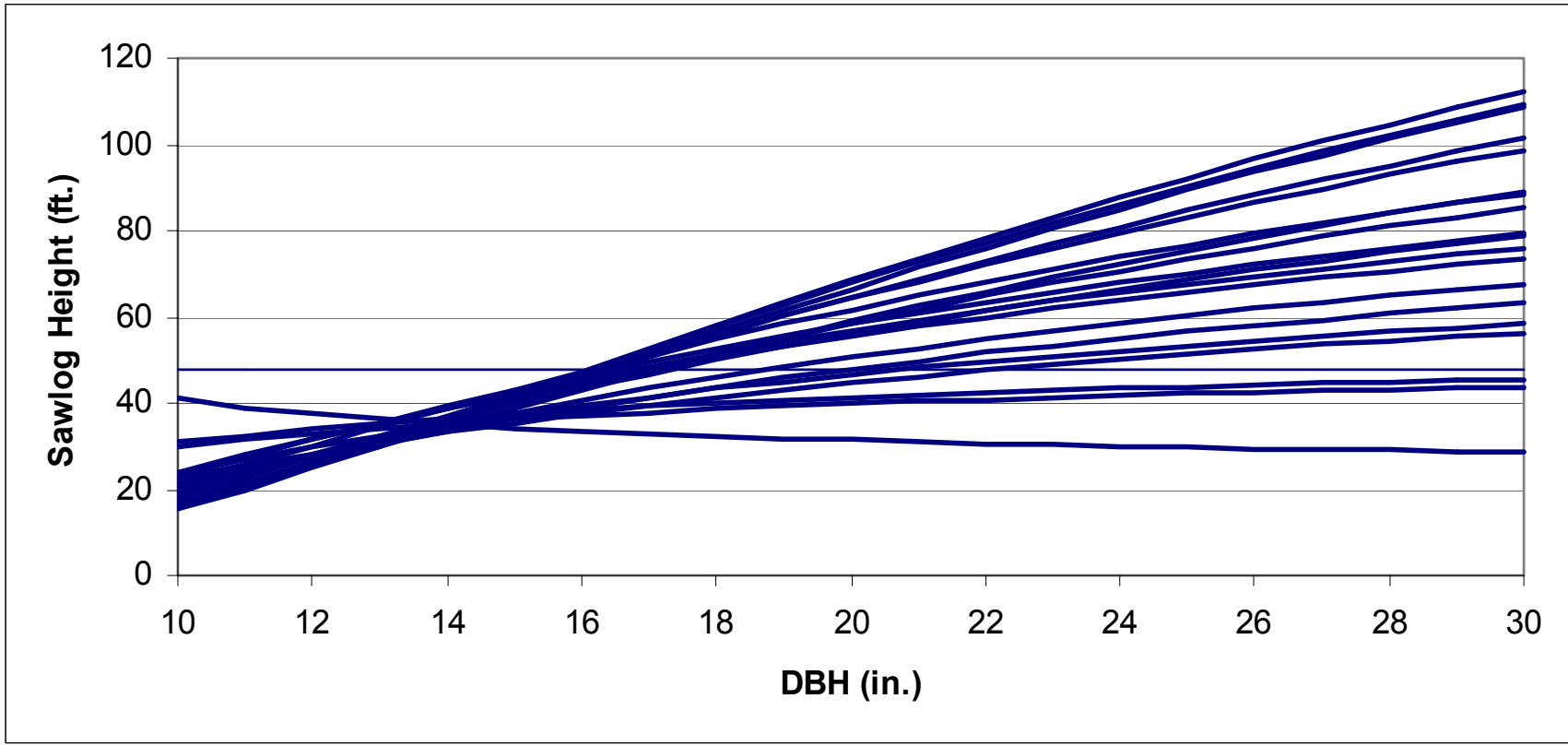


Figure 4. Predicted height curves for white oak (*Quercus alba* L.) in the first 20 of the 500 simulations of Tygart Tract. Number of sample trees ranges from 5 to 17.

**Table 5. Average bias and root mean squared error for mean boardfoot volume per acre with double sampling and height regression sampling with 500 simulations of each dataset.**

Tract	Acres	Points	Average Bias (bf/ac)			RMSE (bf/ac)		
			DS	HRS (srs)	HRS (ratio)	DS	HRS (srs)	HRS (ratio)
Tygart	426	67	82.4	69.5	33.5	952.9	239.8	223.7
WVU Research Forest	7,594	2013	8.3	-125.0	3.0	197.1	137.8	57.0
Coopers Rock State Forest	4037	1081	-15.6	-146.0	10.8	266.0	168.8	82.0
Coopers Rock Annex	364	98	39.0	163.7	74.5	787.4	326.3	286.0
Compartment 14	138	52	75.3	189.5	198.3	1,413.3	494.1	471.0
1967 single species	3,500	384	-3.4	-72.7	-3.3	75.0	90.0	52.5
Trout Pond	N/A	30	-330.7	324.0	158.6	1,290.8	613.1	524.2
Savage River	53,473	214	20.8	-192.9	-55.9	1,283.3	301.4	237.8

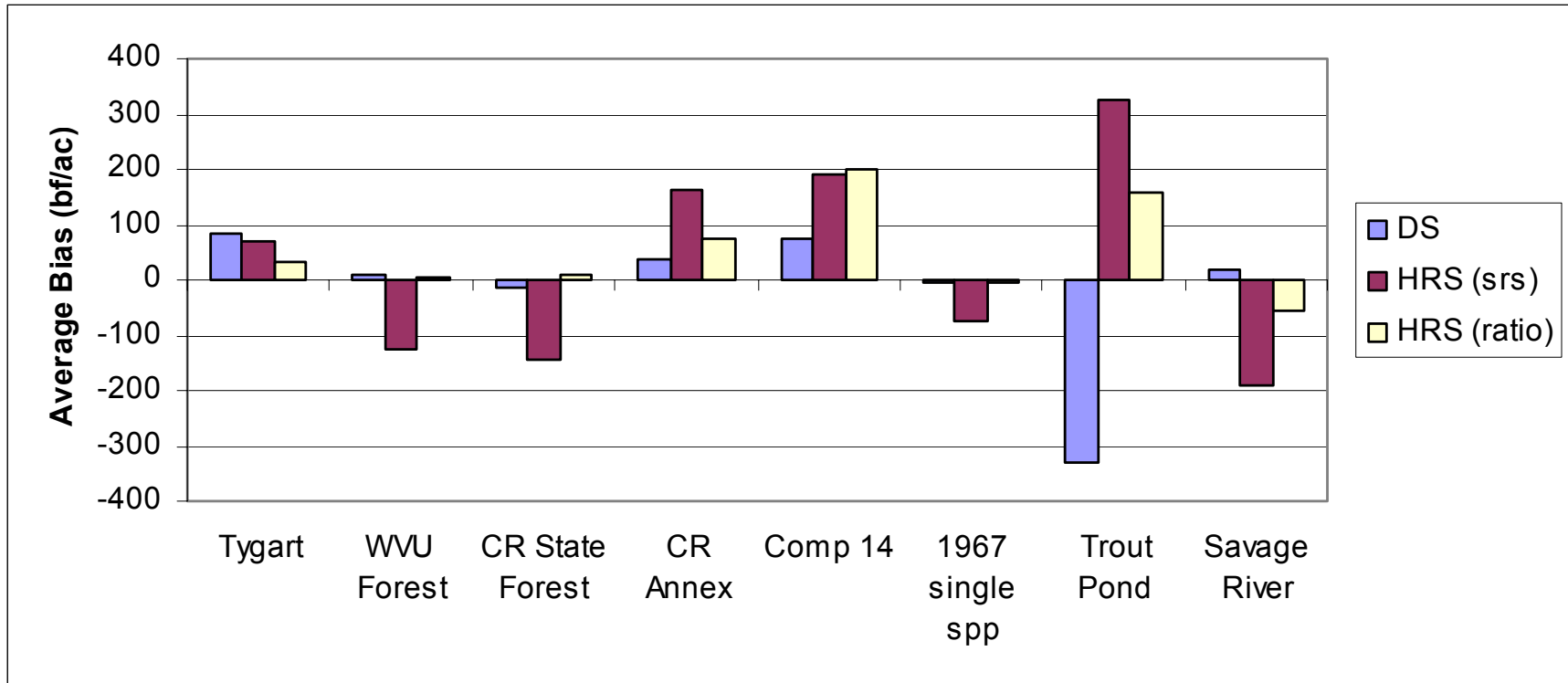


Figure 5. Average bias in boardfoot volume per acre for double sampling and height regression sampling with 500 simulations of each dataset

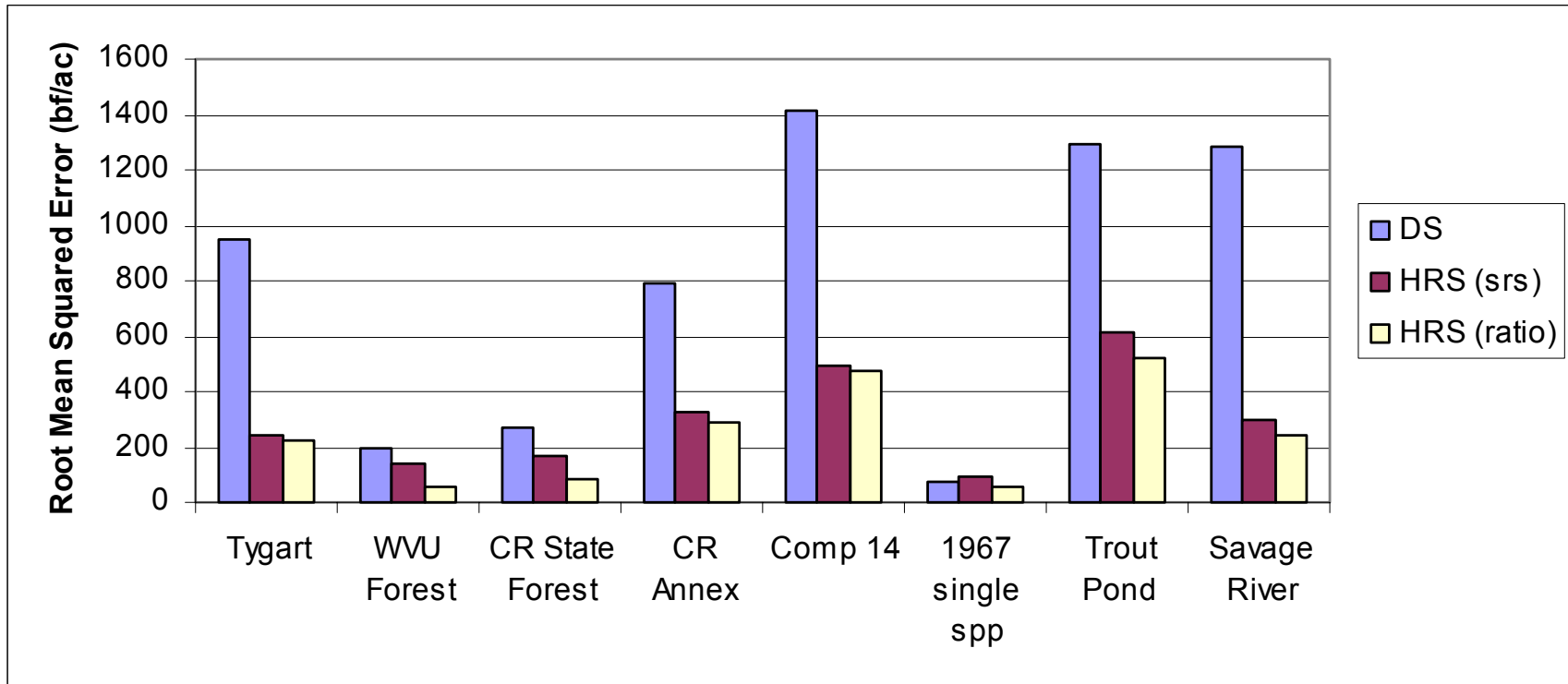


Figure 6. Root mean squared error in boardfoot volume per acre for double sampling and height regression sampling with 500 simulations of each dataset.



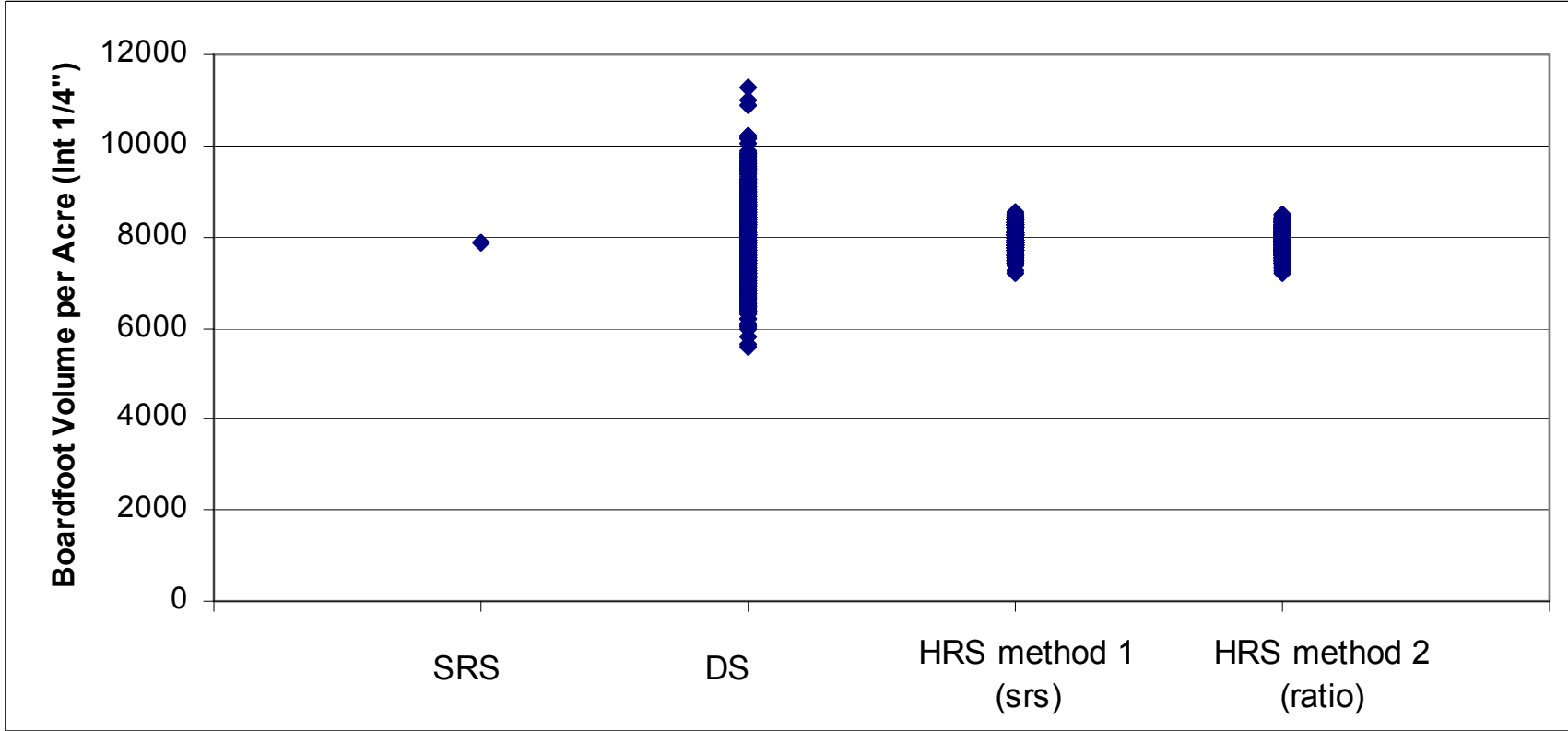


Figure 7. Variation in mean boardfoot volume per acre for Tygart Tract after 500 simulations by sampling type.

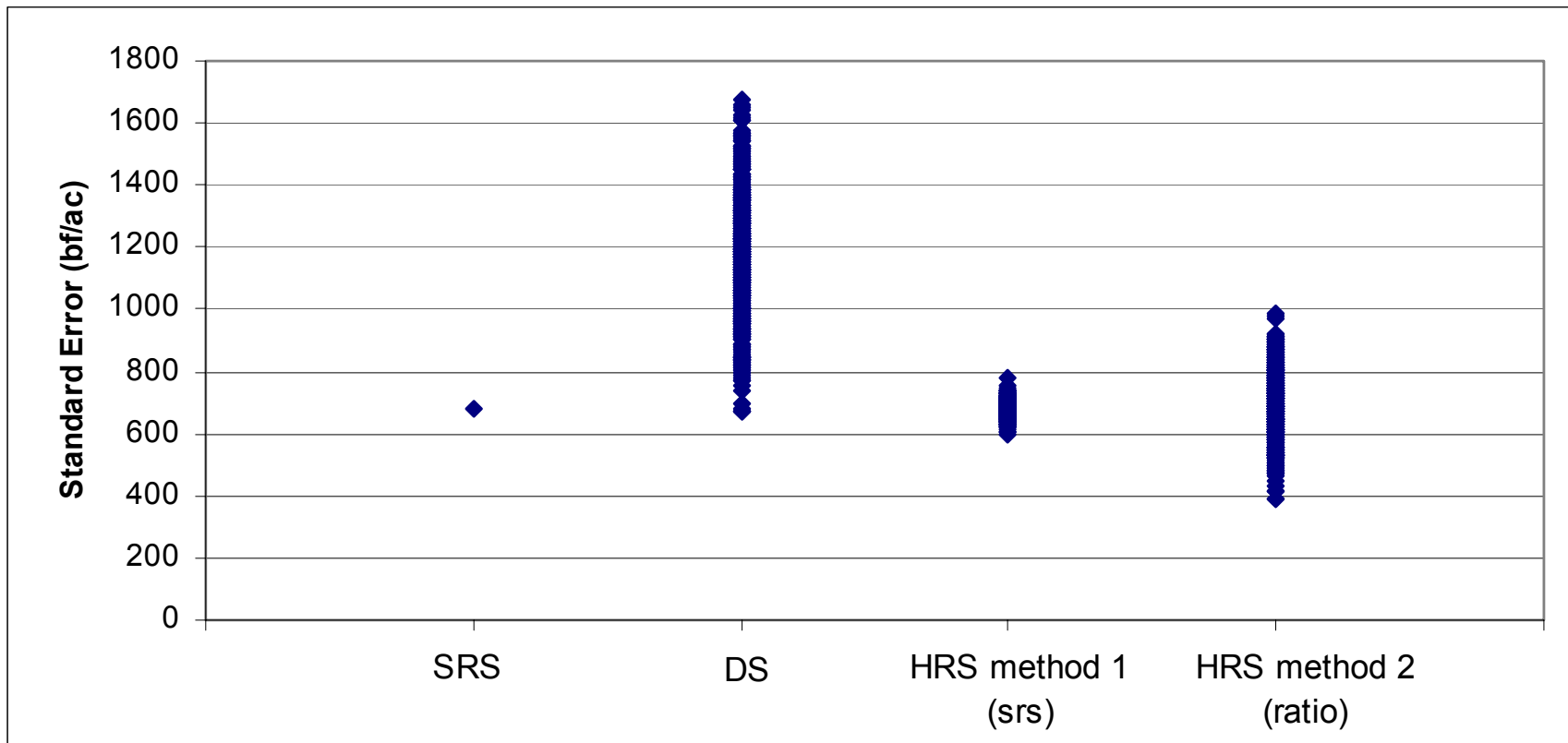


Figure 8. Standard error for Tygart Tract after 500 simulations by sampling type

## Chapter 4: Conclusions

The use of regression analysis to predict merchantable heights and obtain volume estimates on a subset of sampling units has provided some positive results. Mean volume estimation was usually more accurate and precise than the traditional double sampling procedures. The variance of the mean is reduced in part since the natural variation in heights by diameter and thus on volume has been removed through the prediction process. However, this reduction in variation seems to provide unbiased results. Another advantage to using HRS over double sampling is the improved estimate of volume by species, an important feature in areas where there is a large variation in value by species.

Results suggest that HRS method 2 (ratio estimator) provides slightly better results than HRS method 1 (srs estimator) which is magnified on larger inventories. Although data analysis procedures are more complex with method 2, the same field procedures are required regardless of HRS method. Therefore, the cost associated with the field inventory component would be no greater with HRS method 2. Overall differences between these two methods become very small when compared to the added bias associated with traditional double sampling.

Both height regression sampling methods used in obtaining a mean volume per acre and the associated sampling error provide reasonable results and a smaller standard error than with traditional double sampling. The standard error for HRS method 2 is statistically sound but it appears that the estimated standard error for HRS method 1 is adequate.

The application of the HRS procedure is not warranted in many cases. The measurement of dbh on every sample tree may require more time resources than a single inventory can justify. However, in those situations where the additional diameter and species distribution data is desired, this procedure requires less field collection time than the complete enumeration of every

tree and provides reasonably accurate estimates of mean volume per acre even in variable hardwood populations.

## Bibliography

- Avery, T. E. and H. E. Burkhart. 2002. Forest Measurements. McGraw-Hill Companies Inc. New York, pp.185-186.
- Avery, G., and R. Newton. 1965. Plot sizes for cruising timber in Georgia. *J. For.* 63:930-932.
- Bailes, W. W. 2004. A Comparison of Basal Area and Merchantable Height as Auxiliary Variables for Double Sampling with Point Sampling. M.S. Thesis, W.Va. Univ.
- Freese, F. 1962. Elementary Forest Sampling. USDA For. Ser. Handbook No. 232.
- Grosenbaugh, L.R. 1952. Plotless Timber Estimates, New, Fast, Easy. *Journal of Forestry*. 50: 32-37.
- Johnson, F. A. and H. F. Hixon. 1952. The most efficient size and shape of plot to use for cruising in old growth Douglas-fir timber. *J. For.* 50: 17-20.
- Mesavage, C. and L. R. Grosenbaugh. 1956. Efficiency of several cruising designs on small tracts in north Arkansas. *J. For.* 54: 569-576.
- Scott, C. T. 1979. Northeastern Forest Survey Board-Foot Volume Equations. Northeastern Forest Experiment Station. Forest Service Research Note NE-271.
- Shiver, B. D. and B. E. Borders. 1996. Sampling Techniques for Forest Resource Inventory. John Wiley & Sons Inc. New York.
- Spurr, S. H. 1952. Forest Inventory. The Ronald Press Company., New York, New York.
- Wiant, H. V., Jr., D. O. Yandle, and R. Andres. 1984. Is BAF 10 a good choice for point sampling? *North J. Appl. For.* 1:23-24.

## **Appendix**

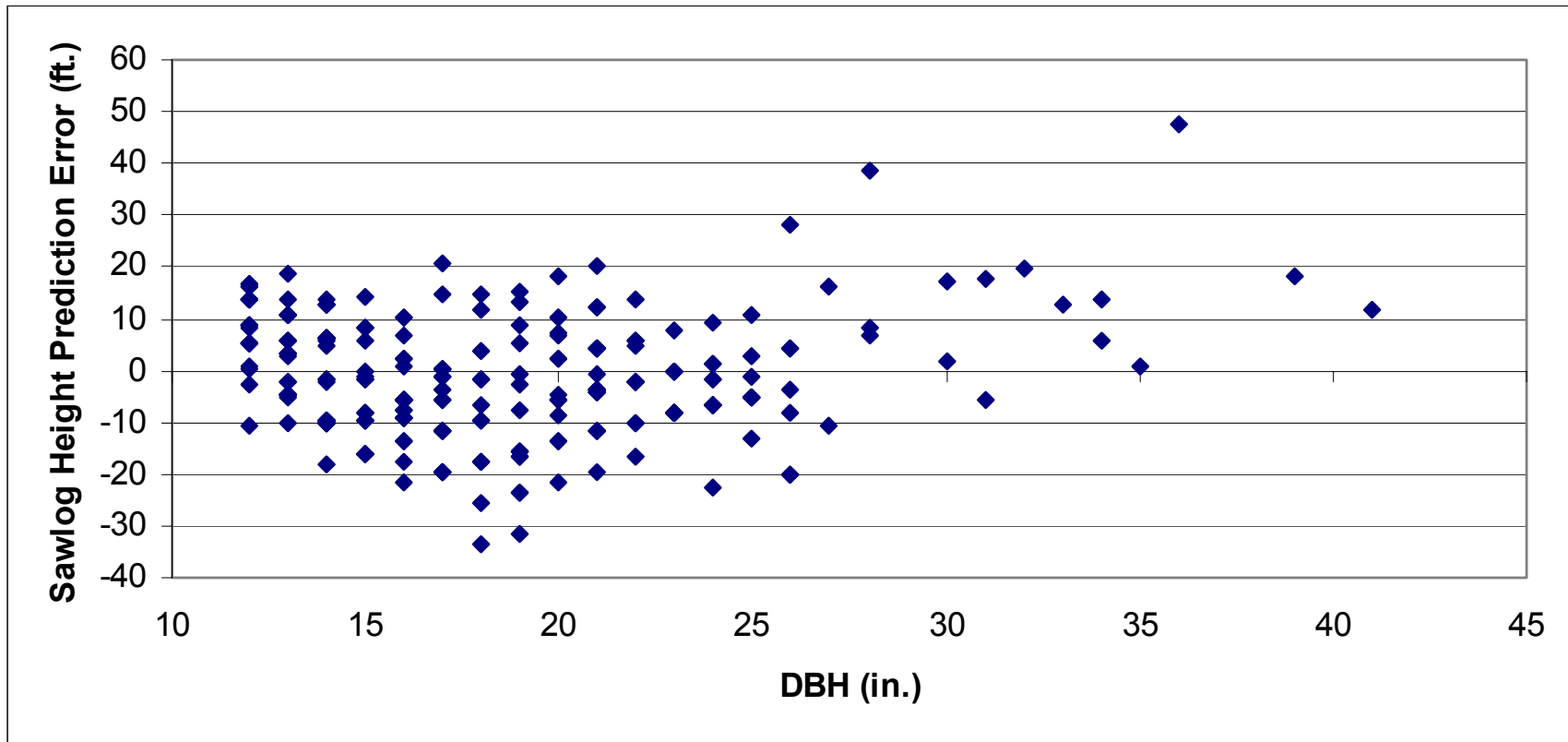


Figure 9. Height prediction error in feet for all species by DBH with height regression sampling of Compartment 14. (n=254)

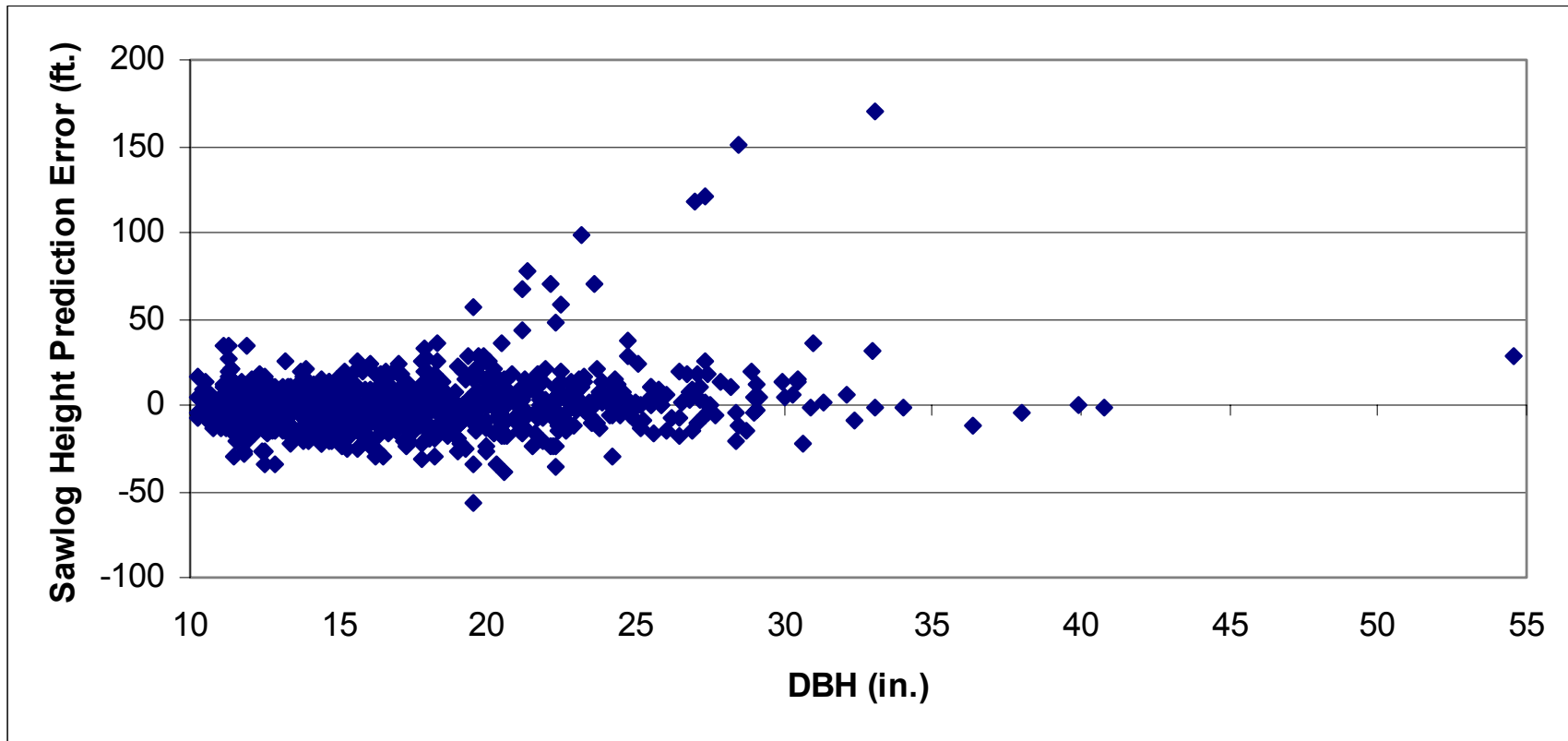


Figure 10. Height prediction error in feet for all species by DBH with height regression sampling of Savage River. (n=982)



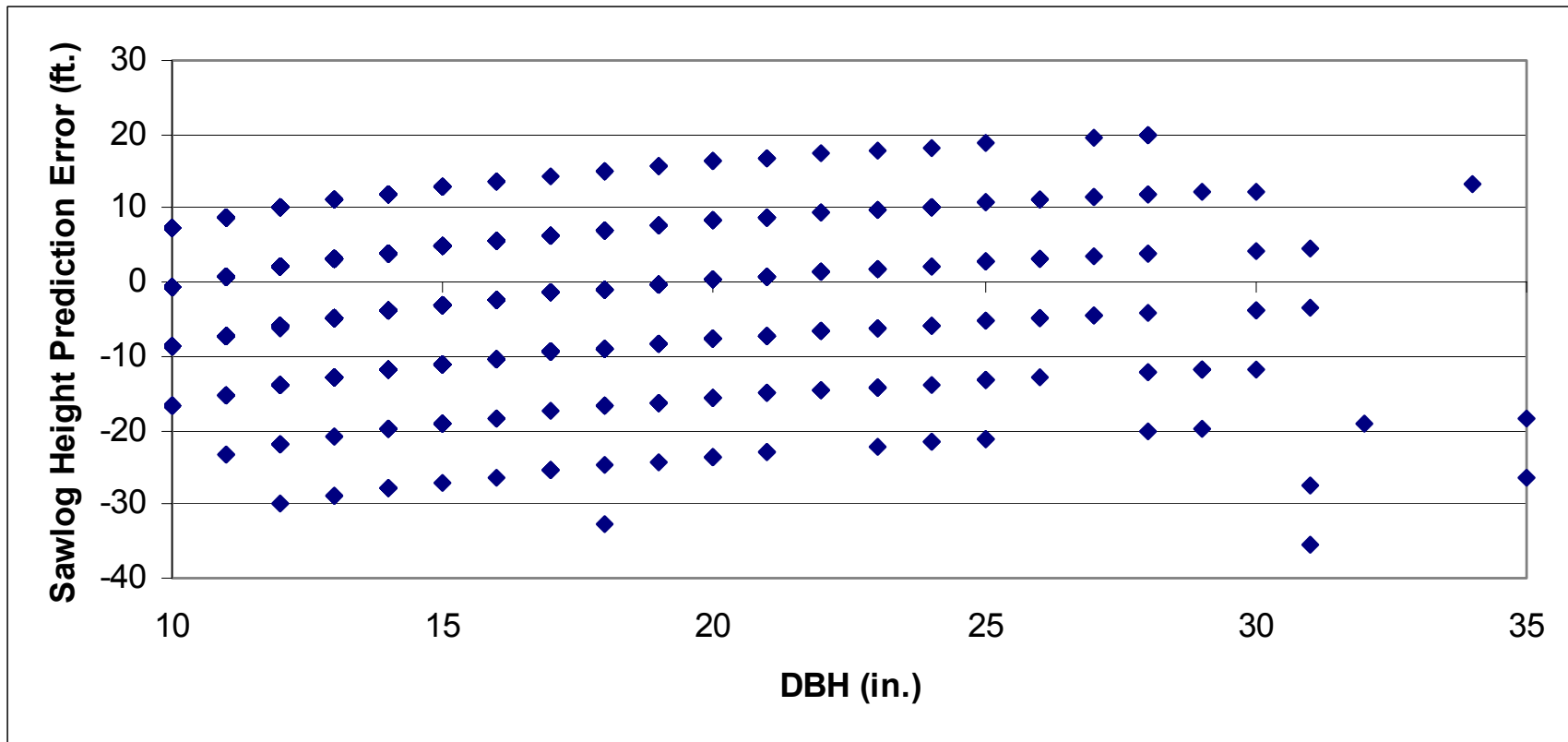


Figure 11. Height prediction error in feet for all species by DBH with height regression sampling of the 1967 single species dataset. (n=1,253)

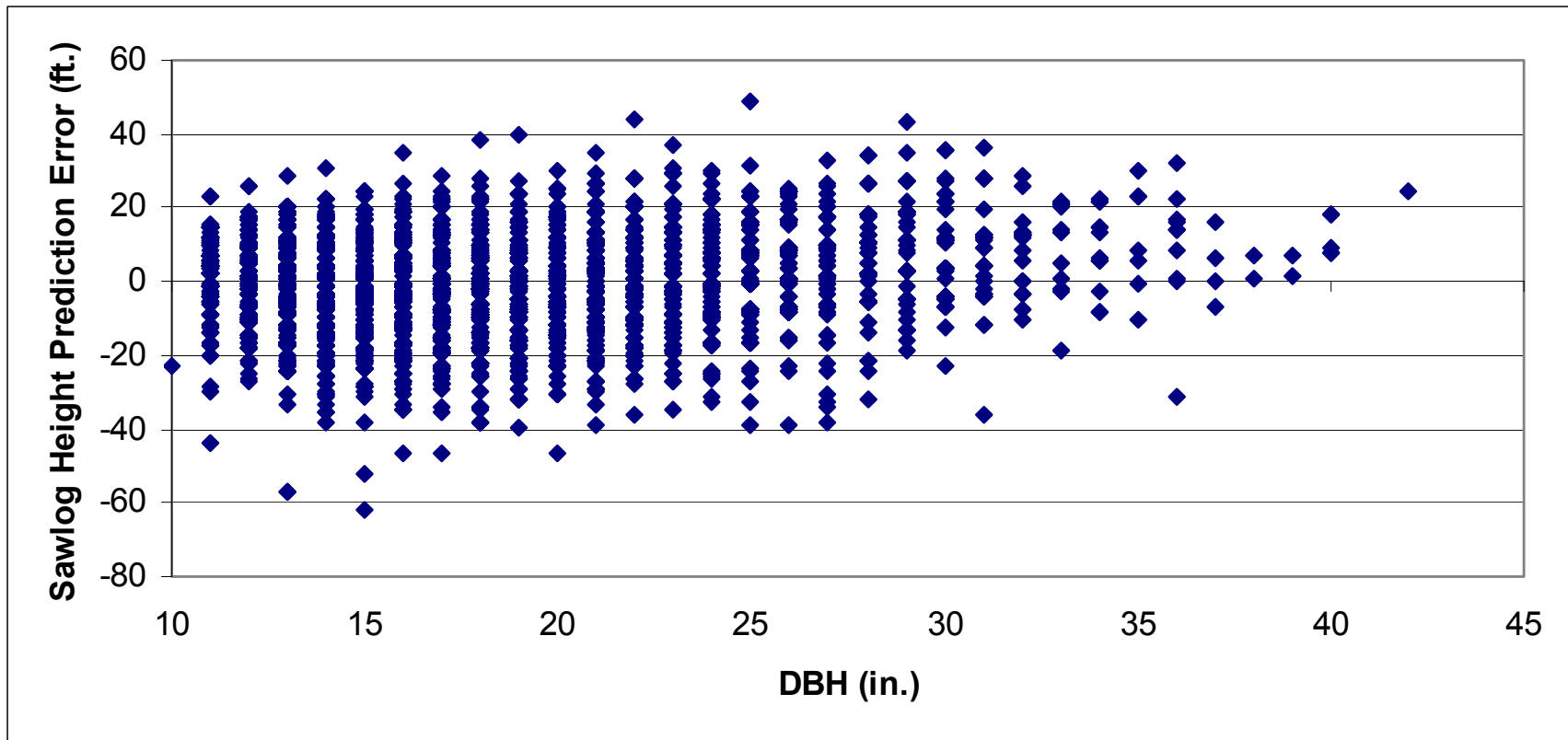


Figure 12. Height prediction error in feet for all species by DBH with height regression sampling of the WVU Research Forest. (n=8,052)

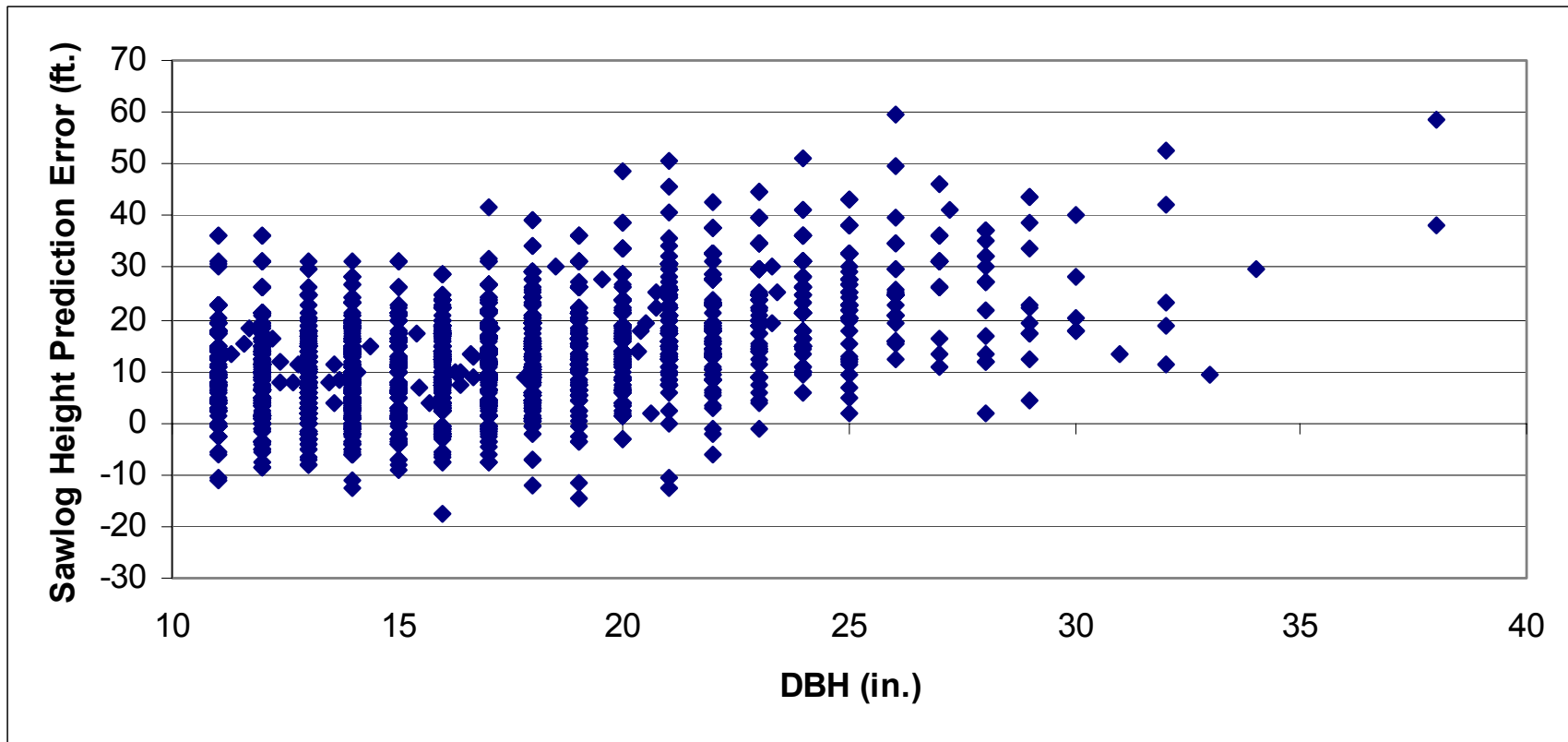


Figure 13. Height prediction error in feet for all species by DBH with height regression sampling of the Coopers Rock State Forest. (n=4,892)

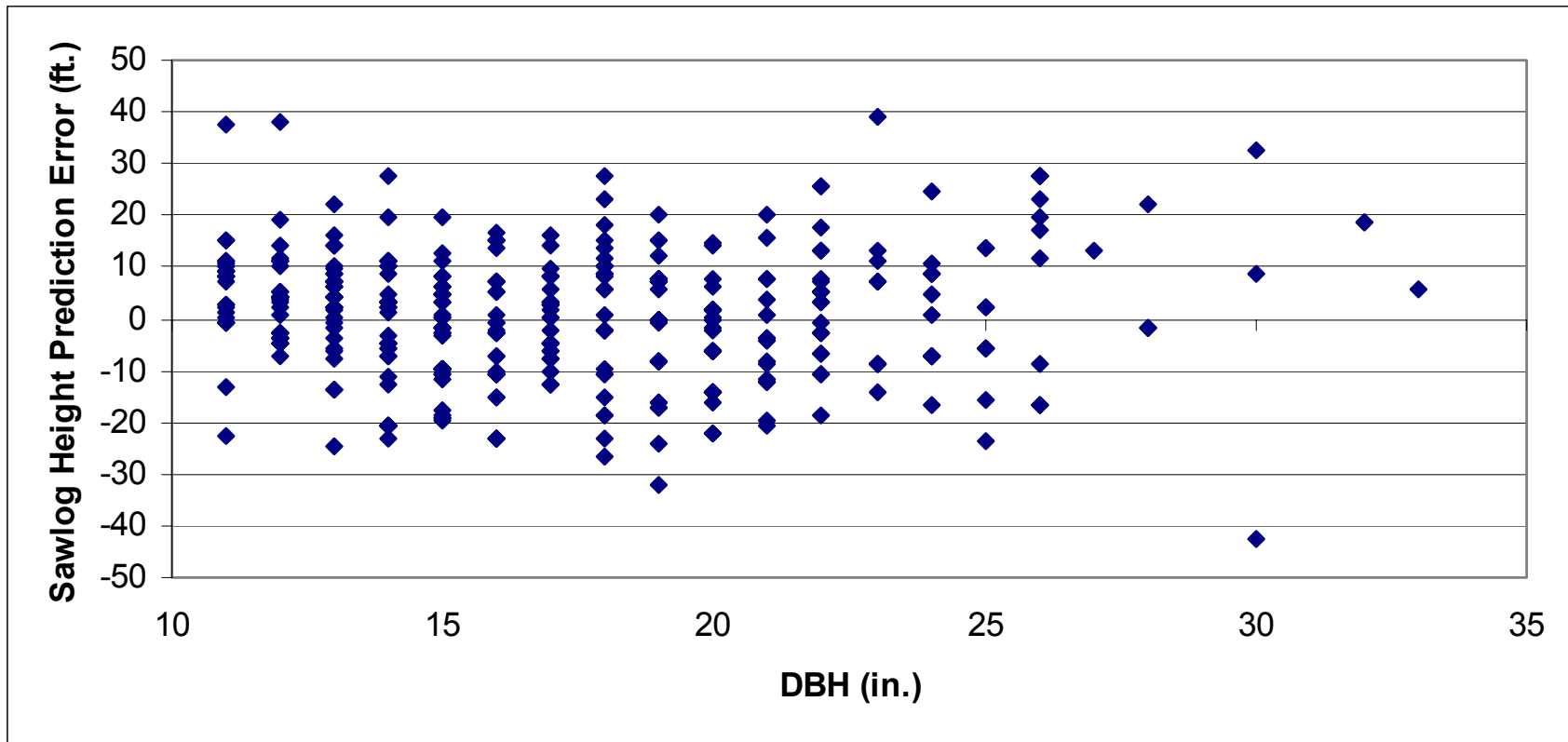


Figure 14. Height prediction error in feet for all species by DBH with height regression sampling of the Coopers Rock Annex. (n=392)

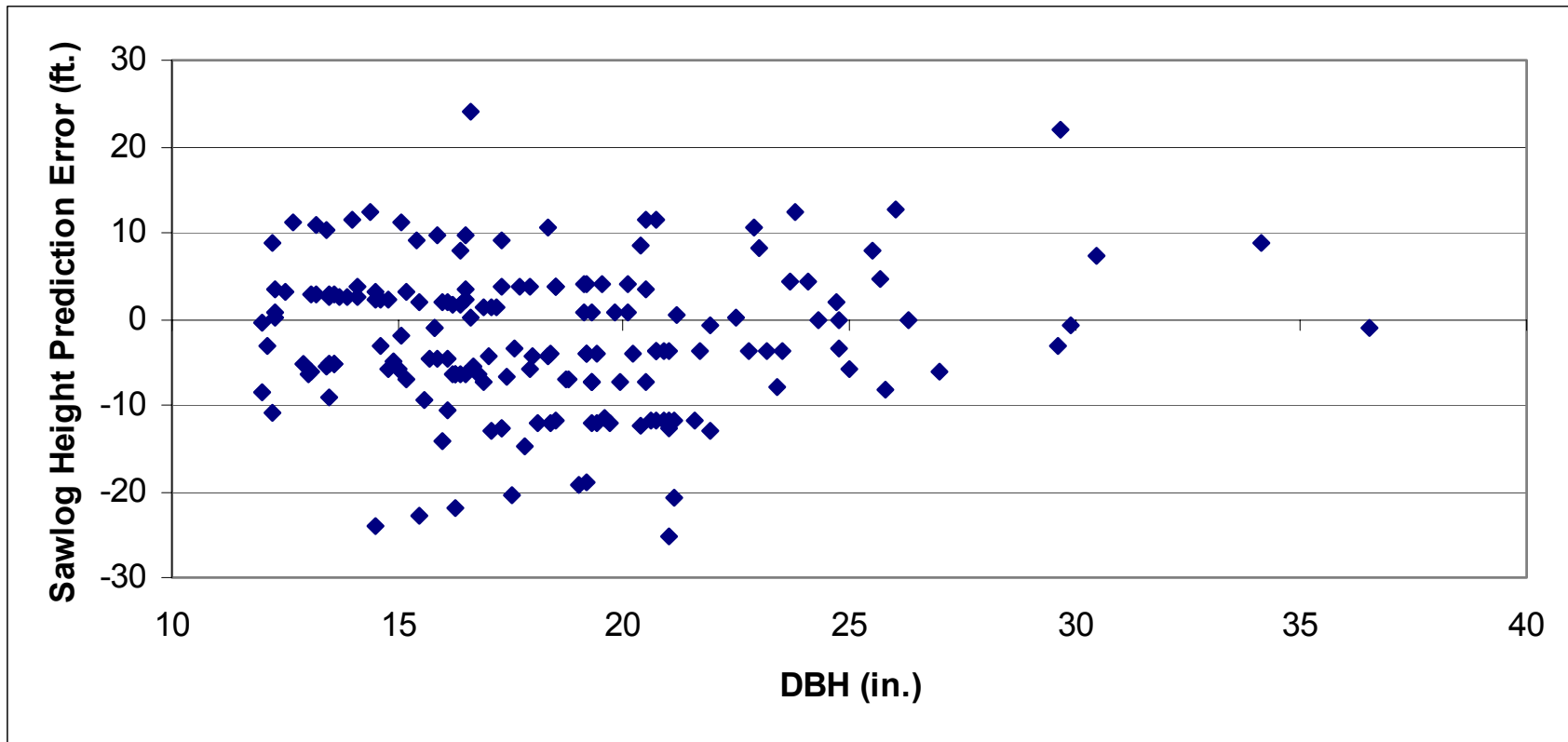


Figure 15. Height prediction error in feet for all species by DBH with height regression sampling of the Trout Pond dataset. (n=177)

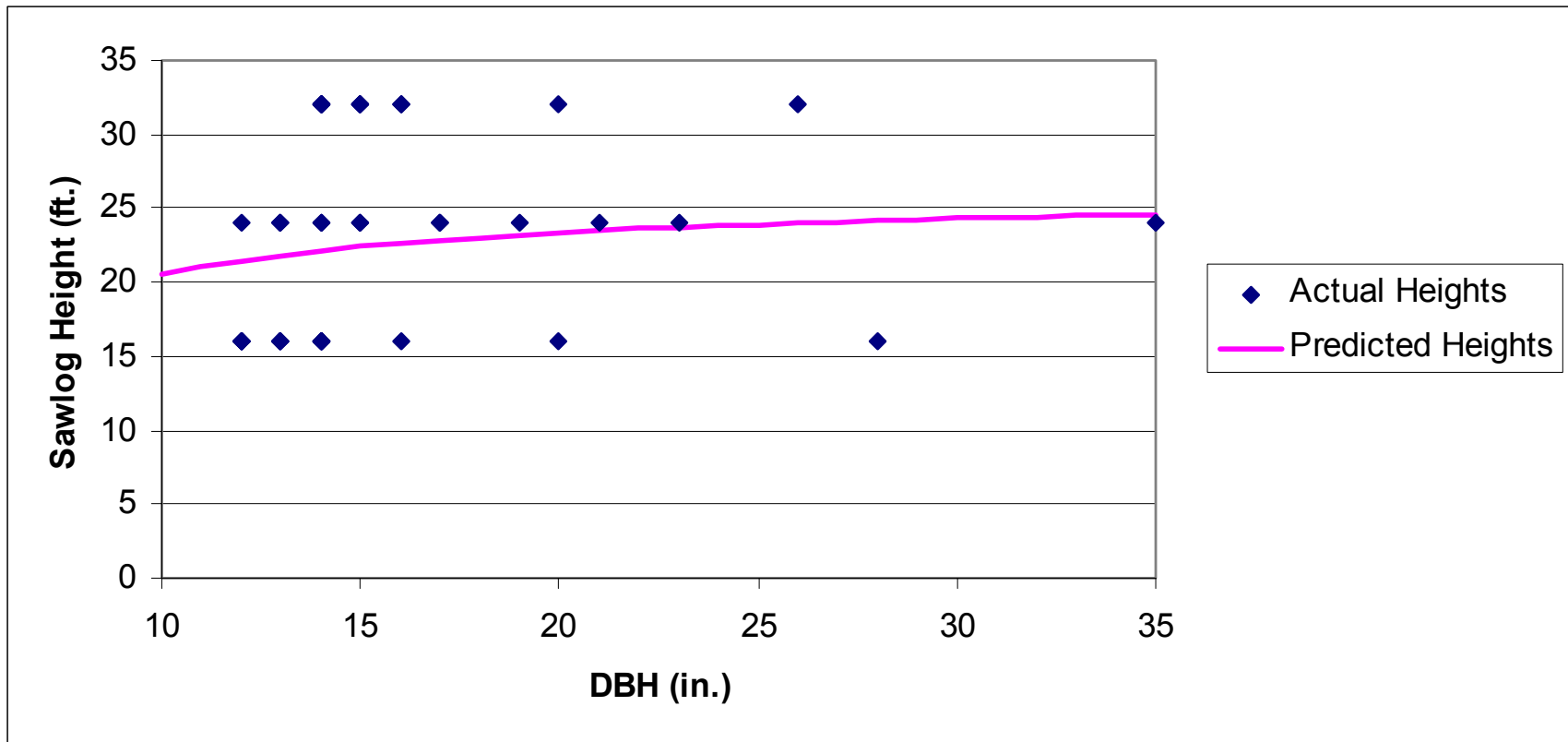


Figure 16. Actual and predicted sawlog heights in feet for chestnut oak with height regression sampling of Compartment 14. (n=39)

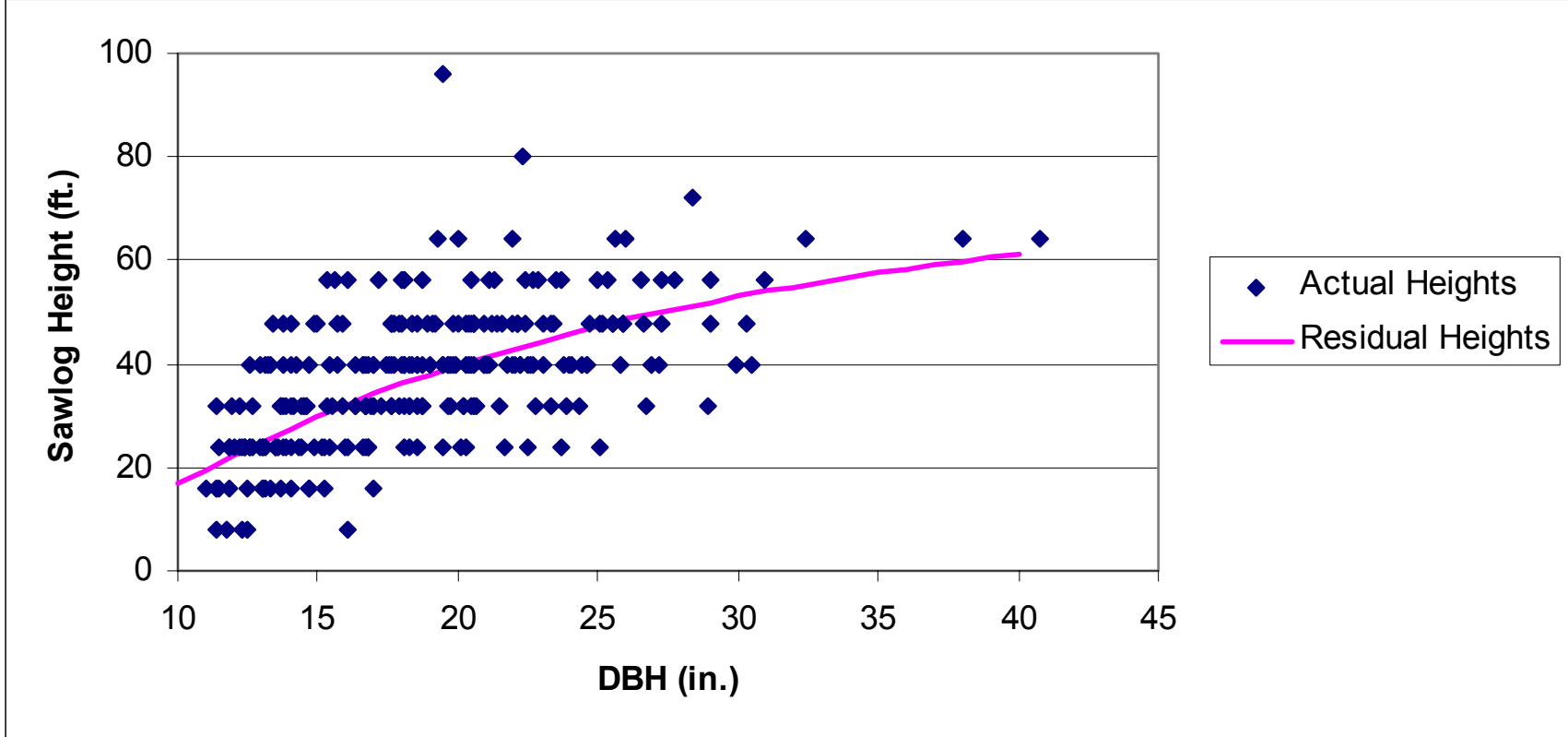


Figure 17. Actual and predicted heights in feet for red oak with height regression sampling of Savage River. (n=267)

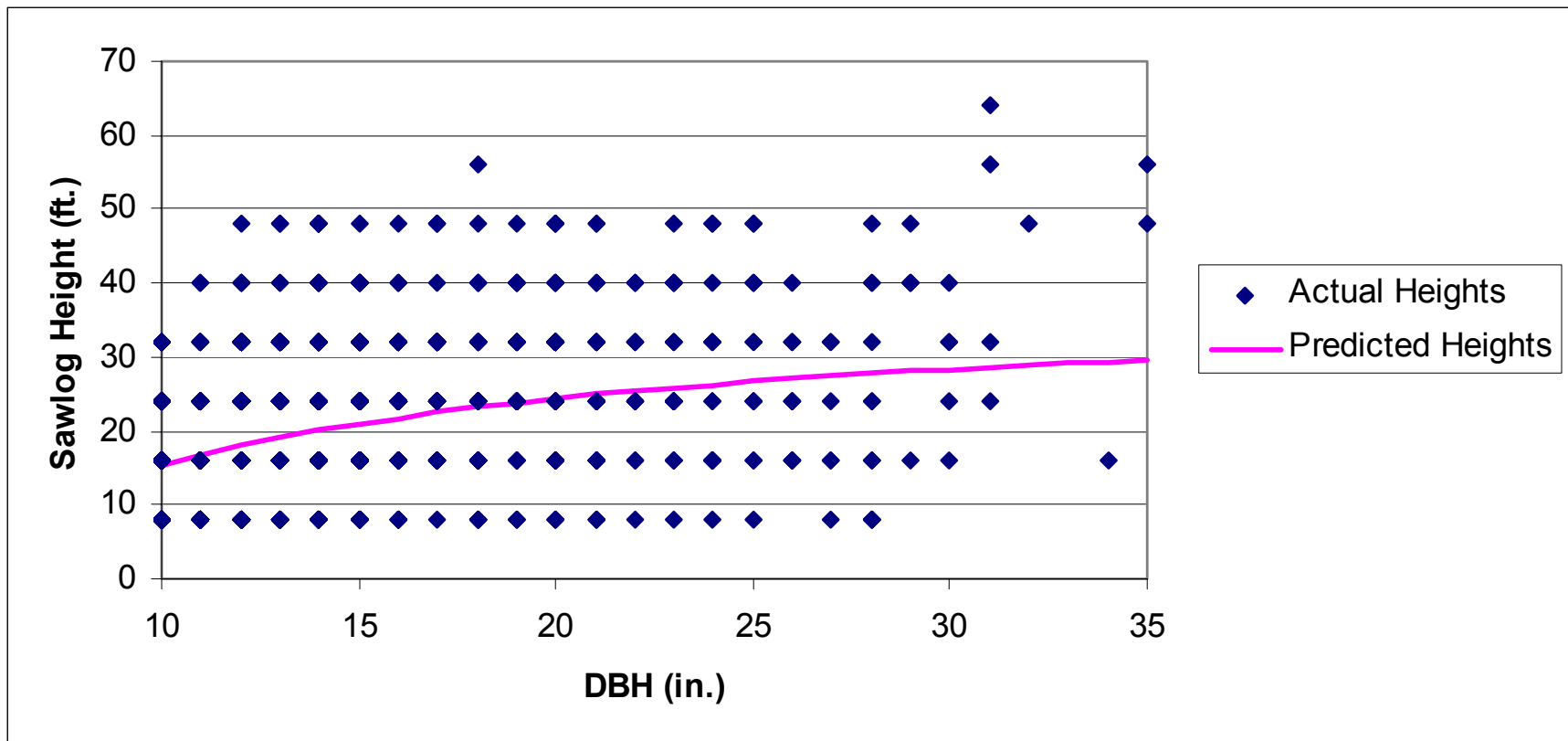


Figure 18. Actual and predicted heights in feet for all species with height regression sampling of the 1967 single species dataset. (n=1,253)



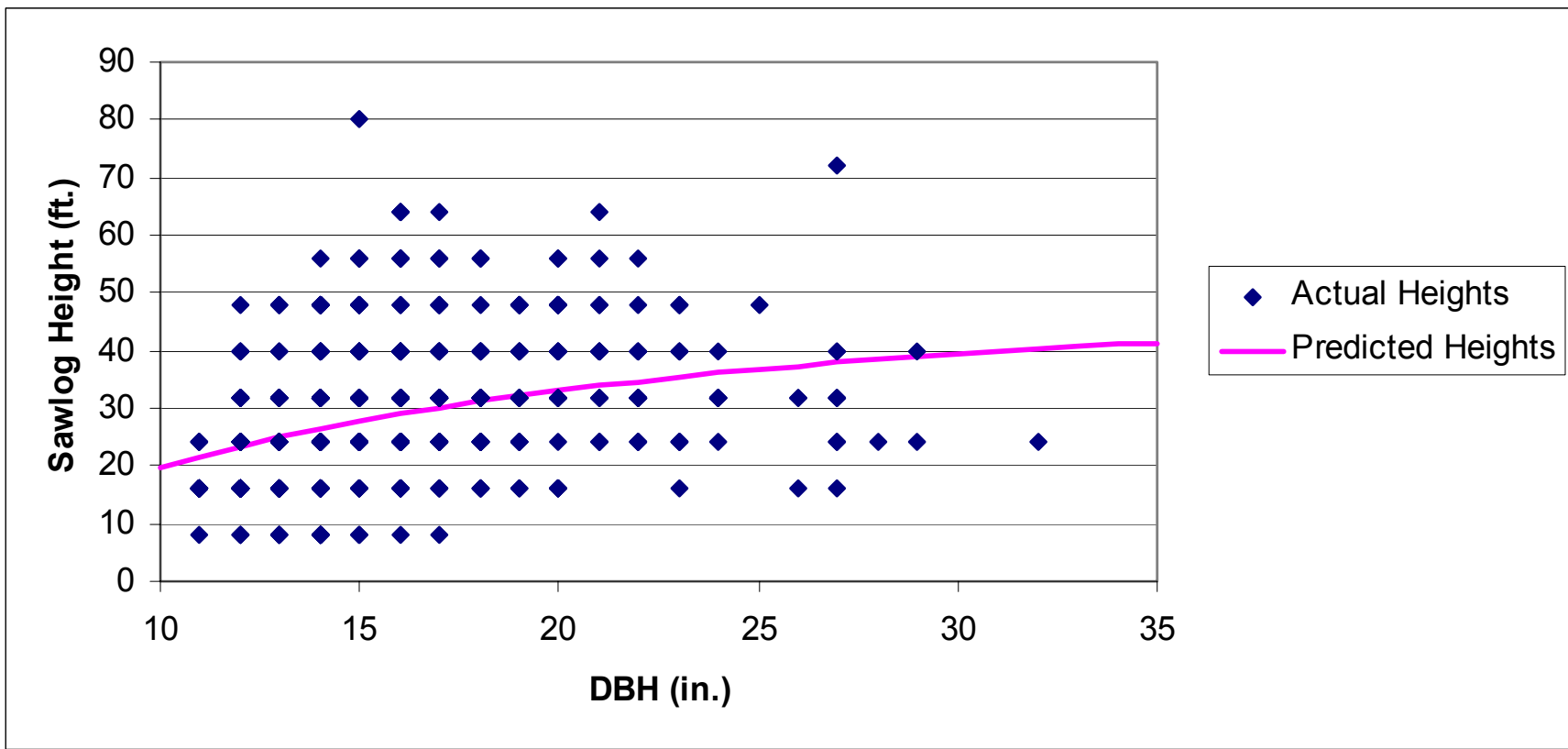


Figure 19. Actual and predicted heights in feet for black cherry with height regression sampling of the WVU Research Forest. (n=476)

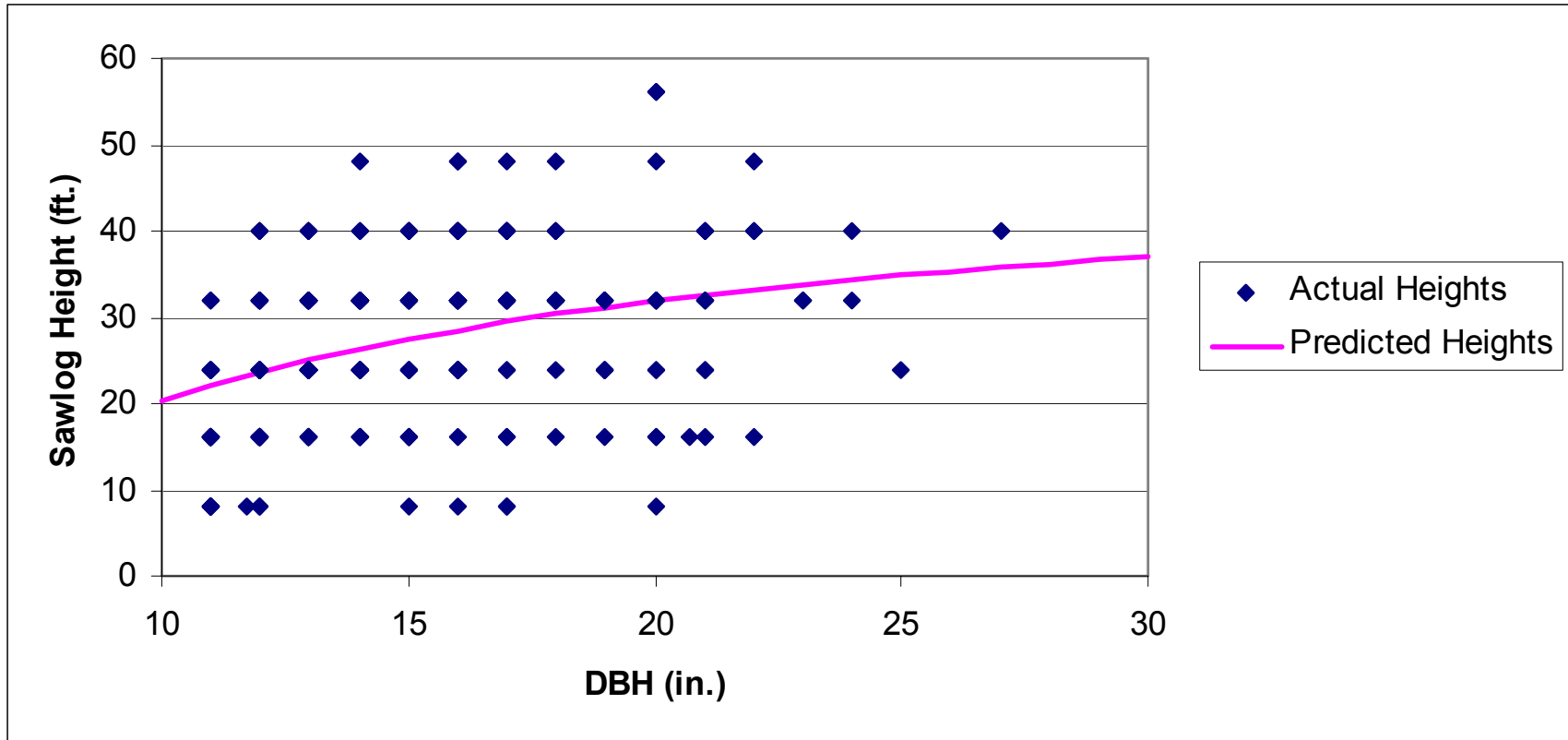


Figure 20. Actual and predicted heights in feet for white oak with height regression sampling of Coopers Rock State Forest. (n=248)

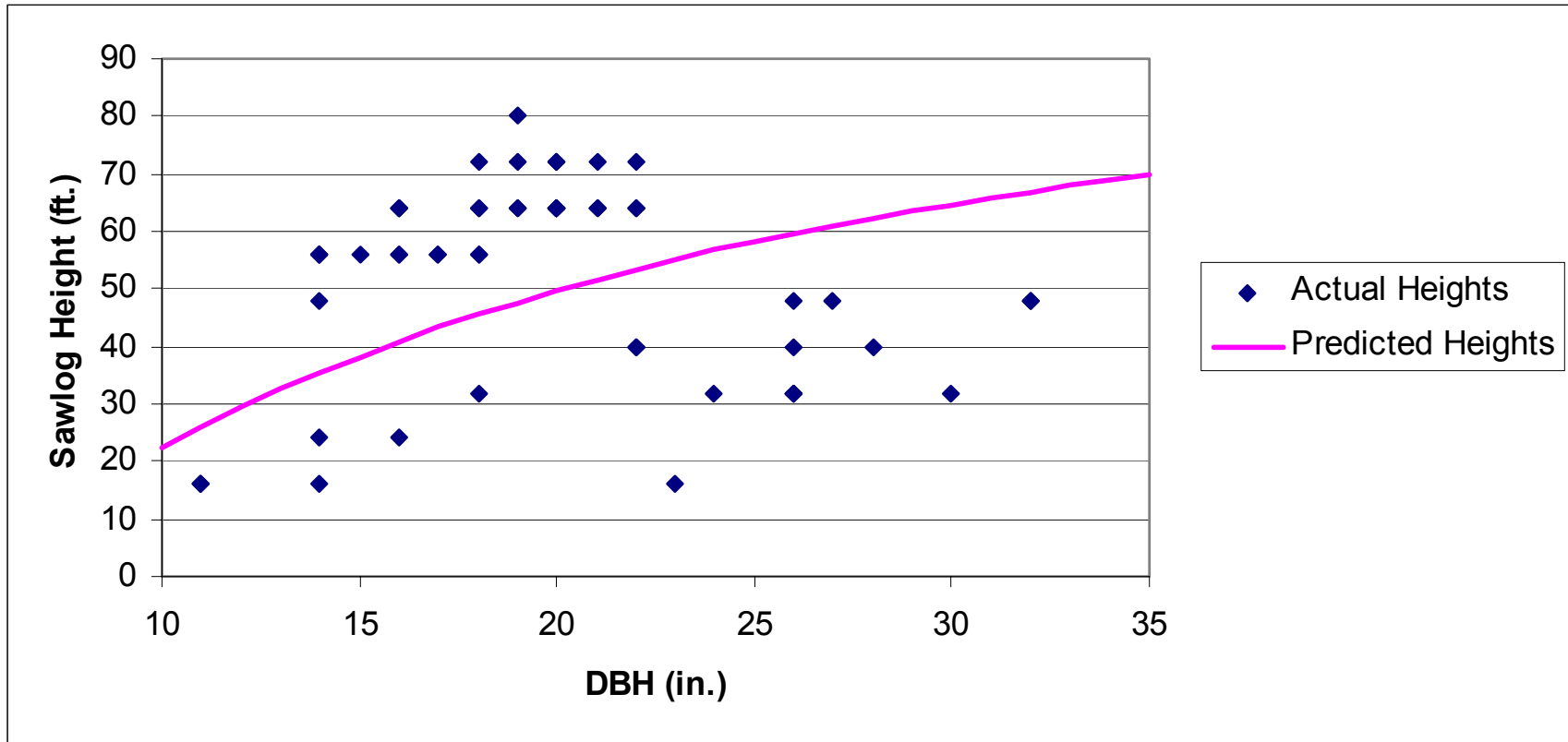


Figure 21. Actual and predicted heights in feet for yellow-poplar with height regression sampling of Coopers Rock Annex. (n=55)

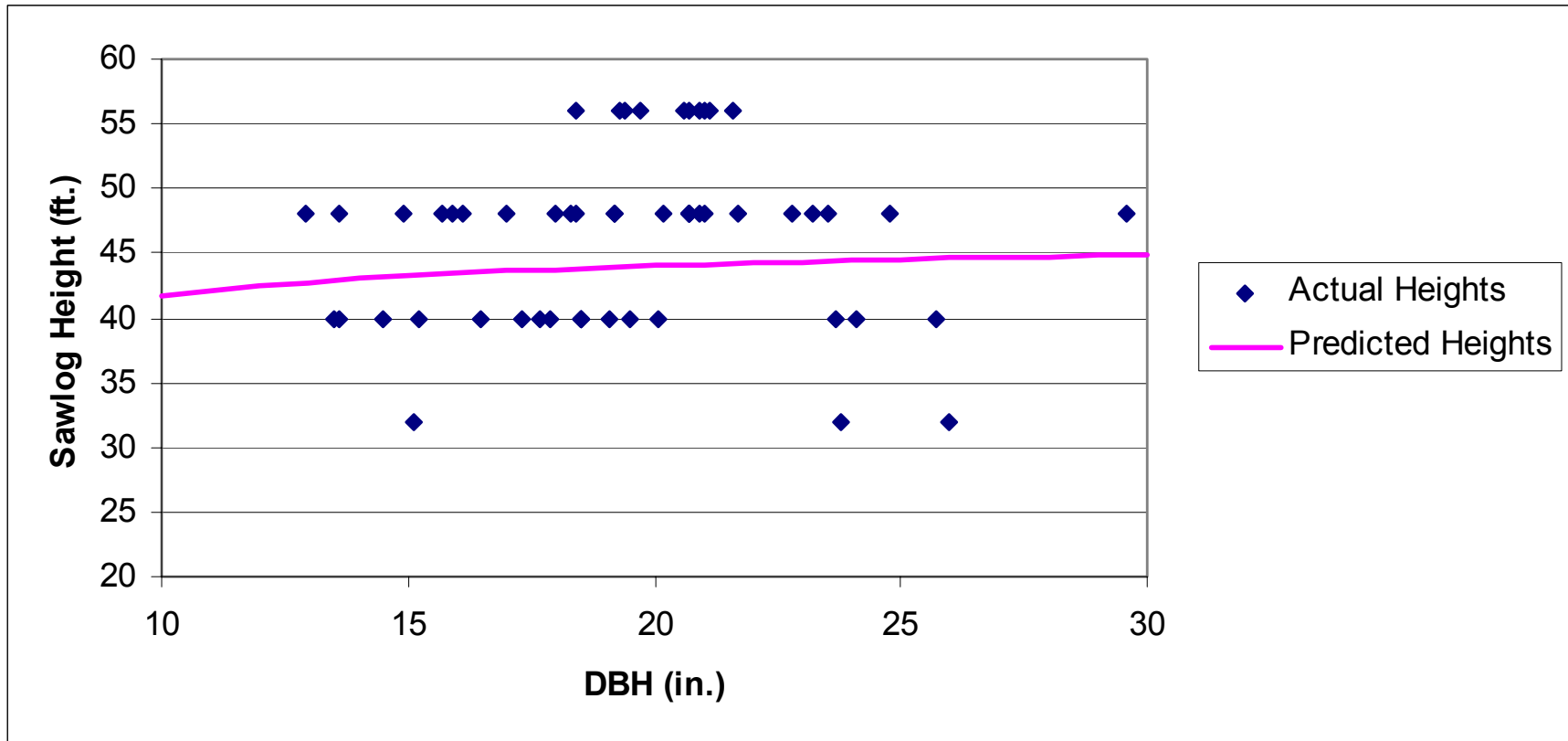


Figure 22. Actual and predicted heights in feet for yellow-poplar with height regression sampling of the Trout Pond dataset. (n=54)

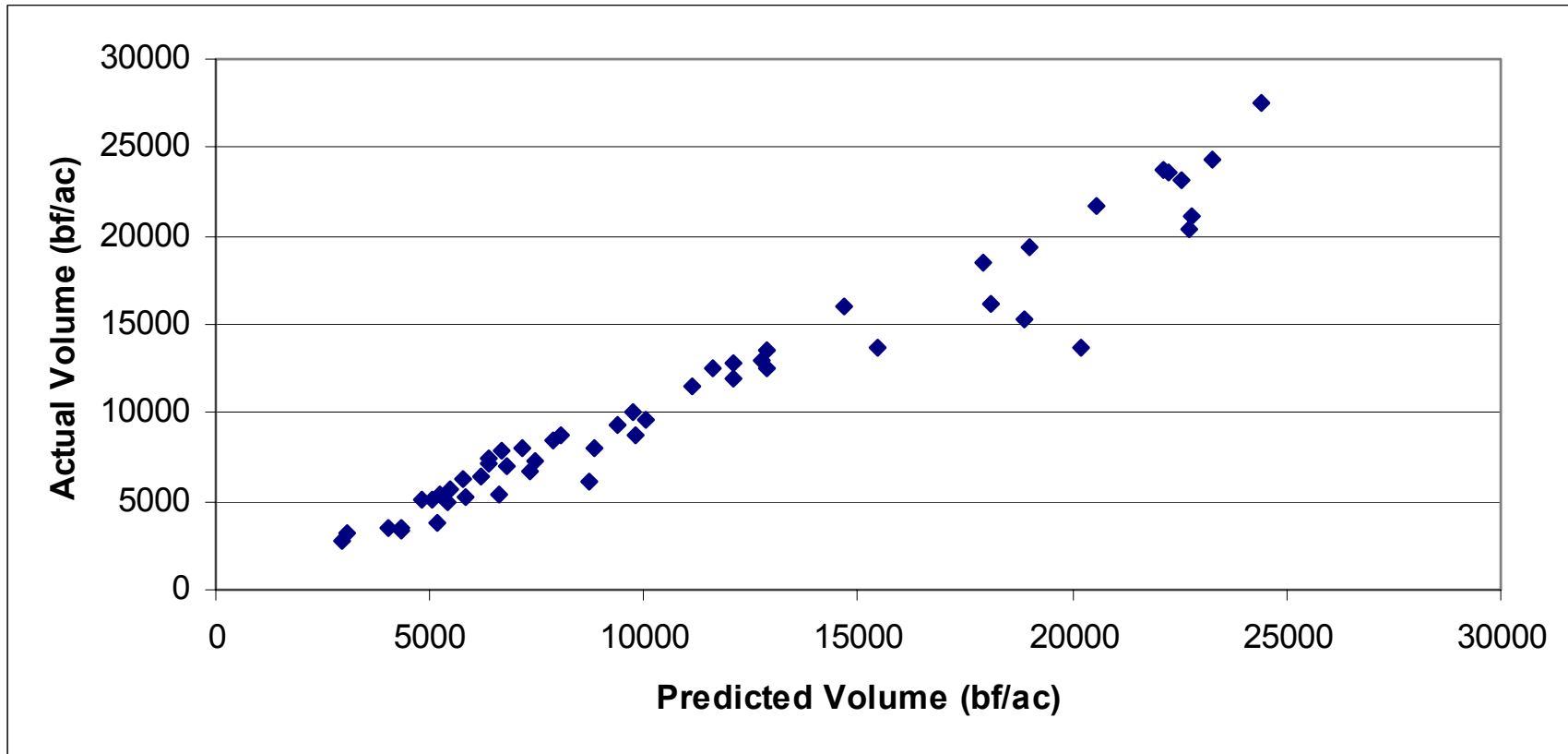
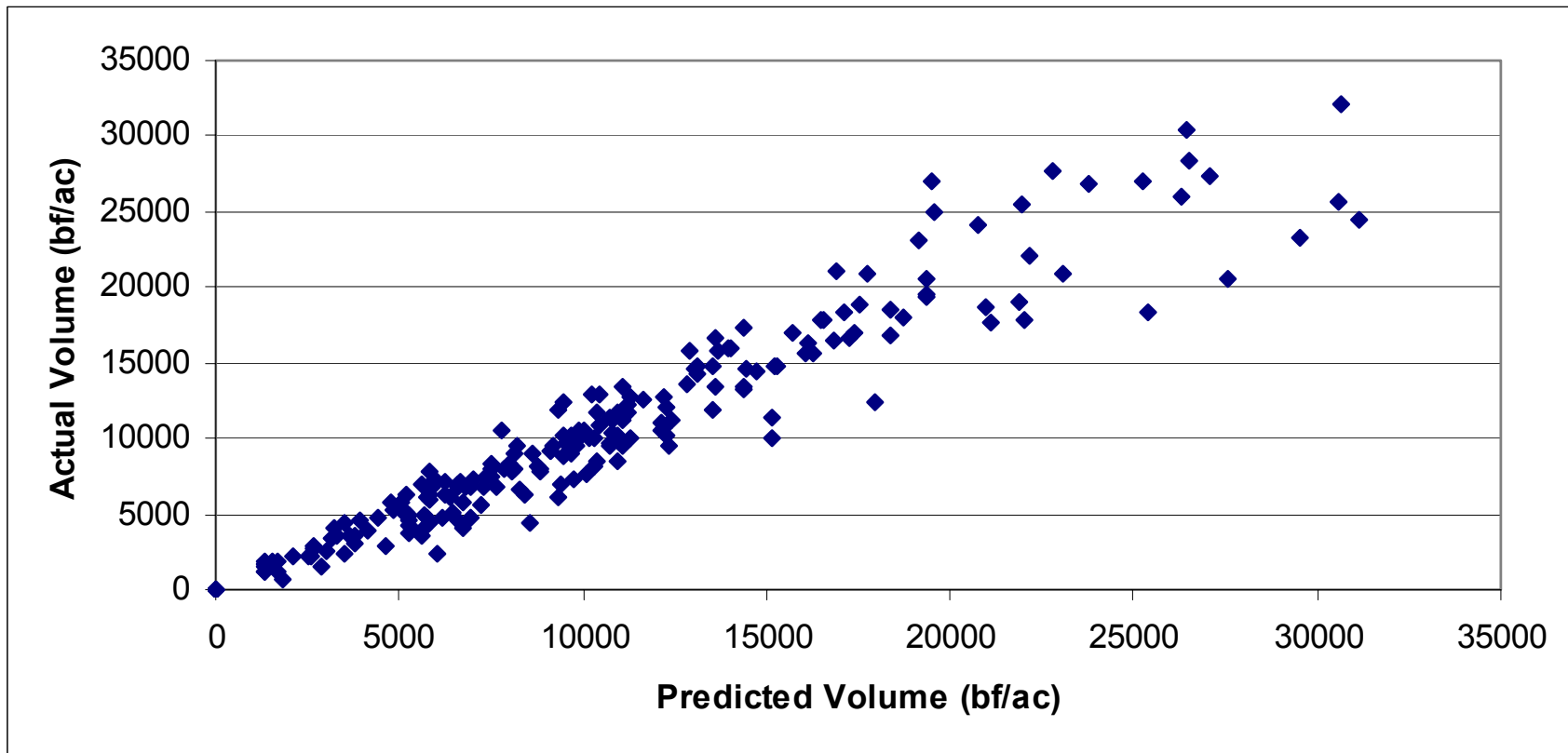


Figure 23. Relationship between actual and predicted volumes for all points on Compartment 14 (Correlation Coefficient = 0.975).



**Figure 24. Relationship between actual and predicted volumes for all points on Savage River (Correlation Coefficient = 0.960).**

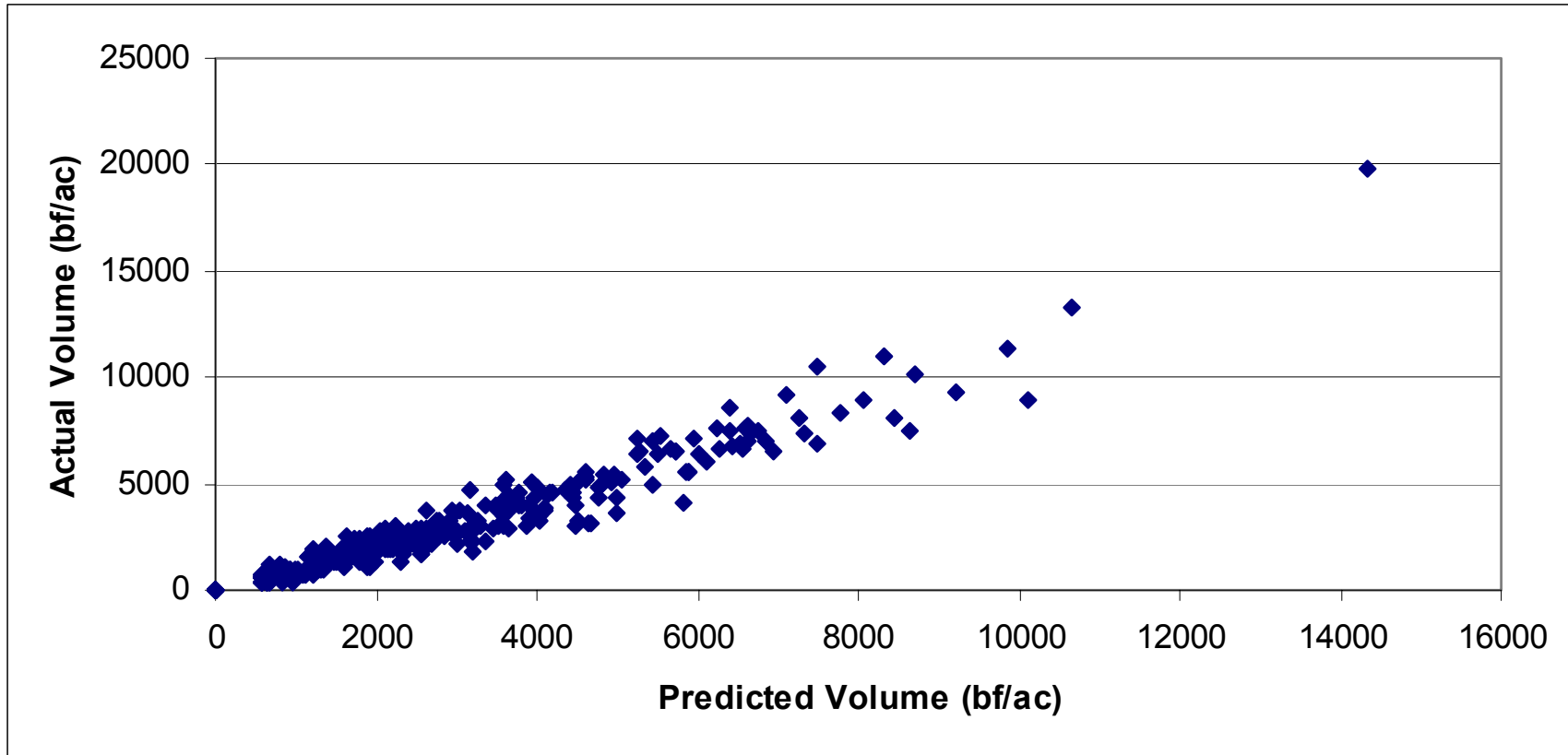
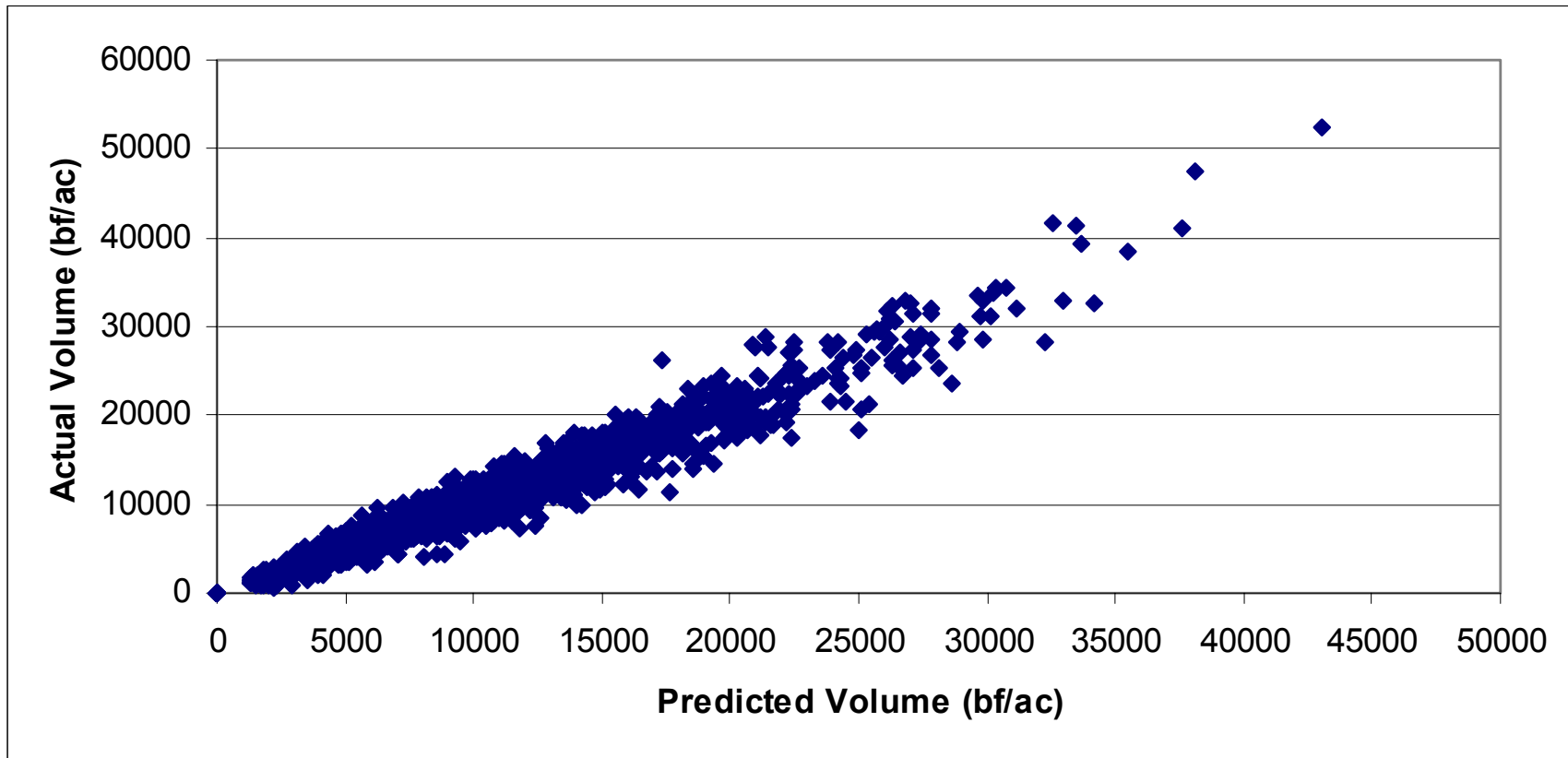


Figure 25. Relationship between actual and predicted volumes for all point on the 1967 single species dataset (Correlation Coefficient = 0.973).



**Figure 26. Relationship between actual and predicted volumes all points on the WVU Research Forest (Correlation Coefficient = 0.977).**



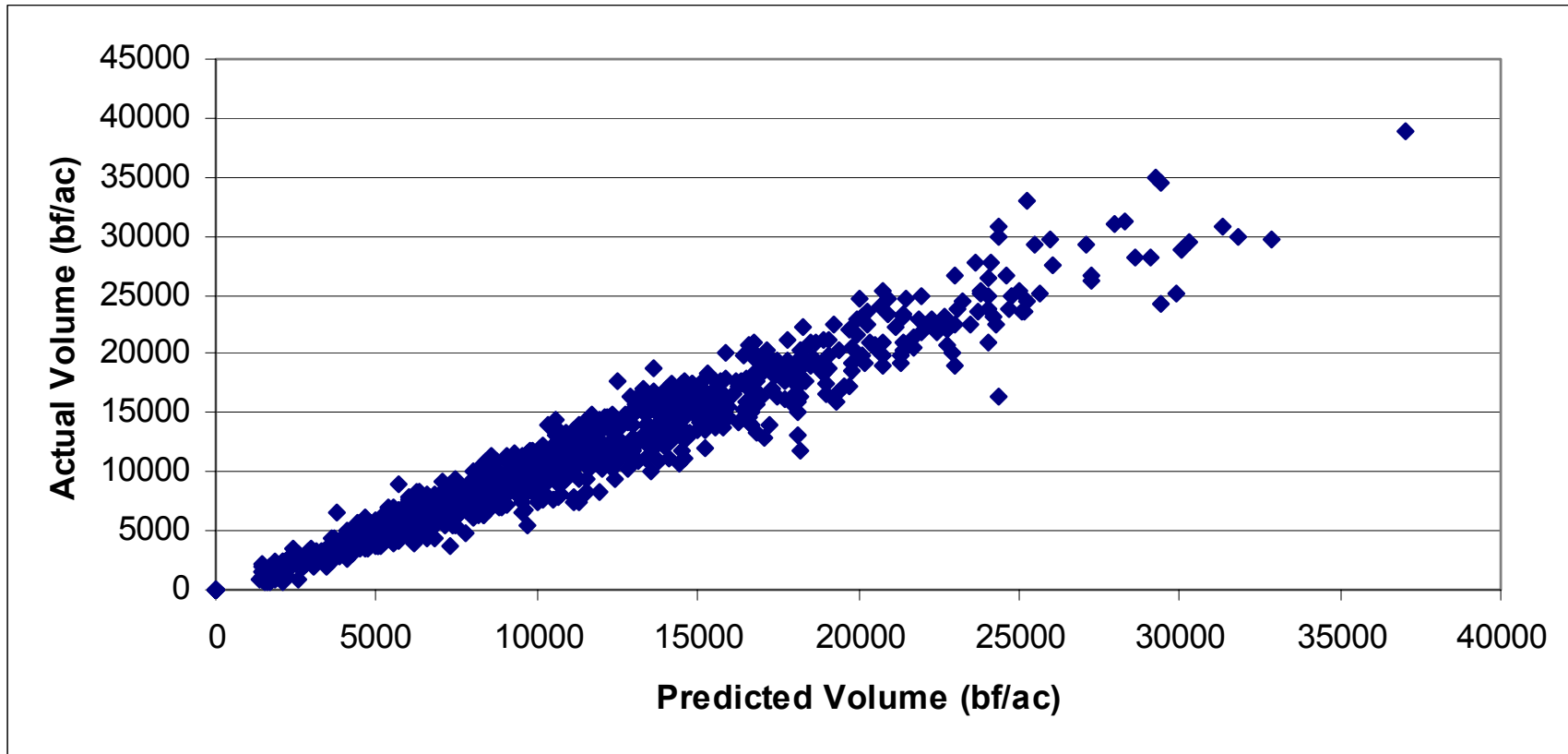


Figure 27. Relationship between actual and predicted volumes for all points on the Coopers Rock State Forest (Correlation Coefficient = 0.973).

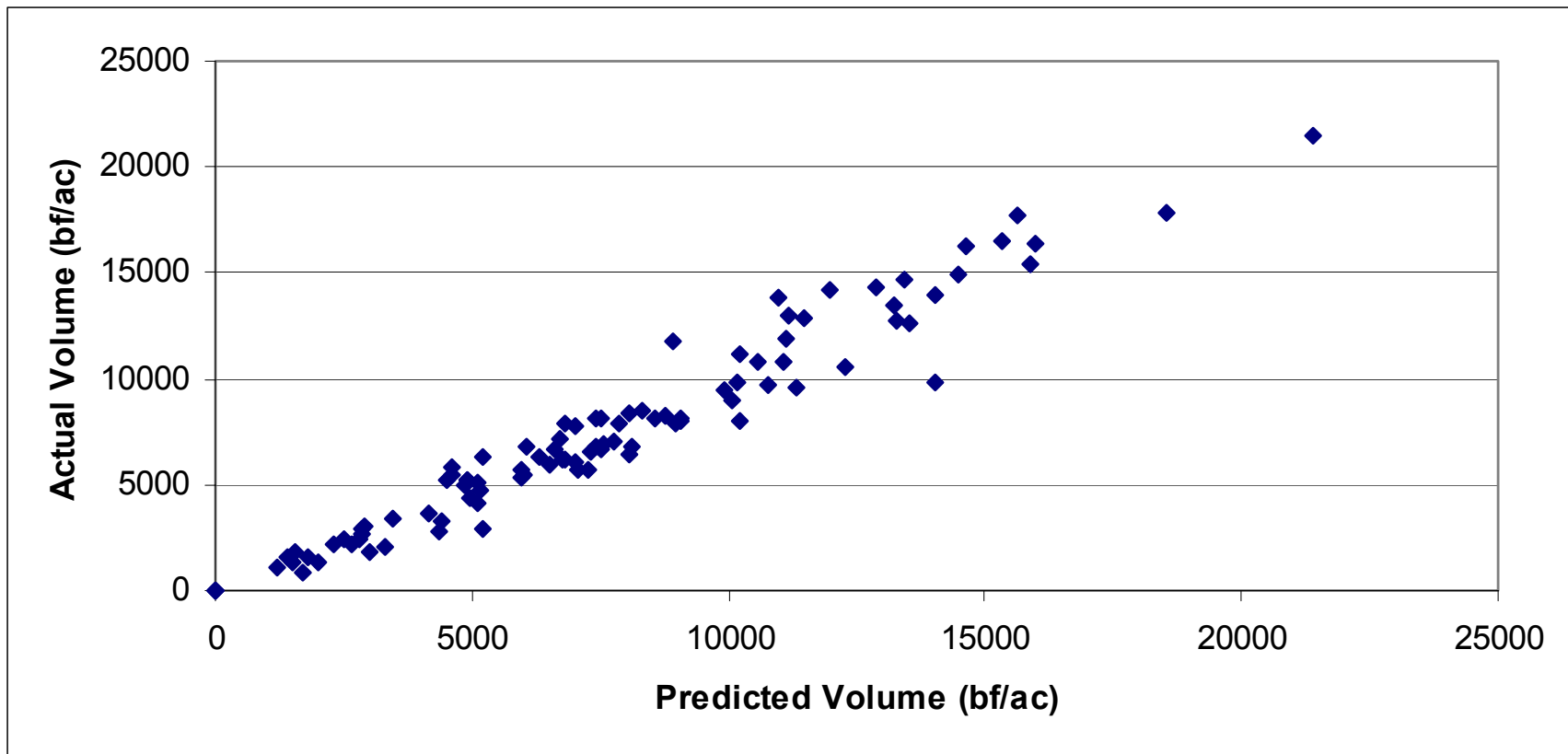


Figure 28. Relationship between actual and predicted volumes for all points on Coopers Rock Annex (Correlation Coefficient = 0.977).

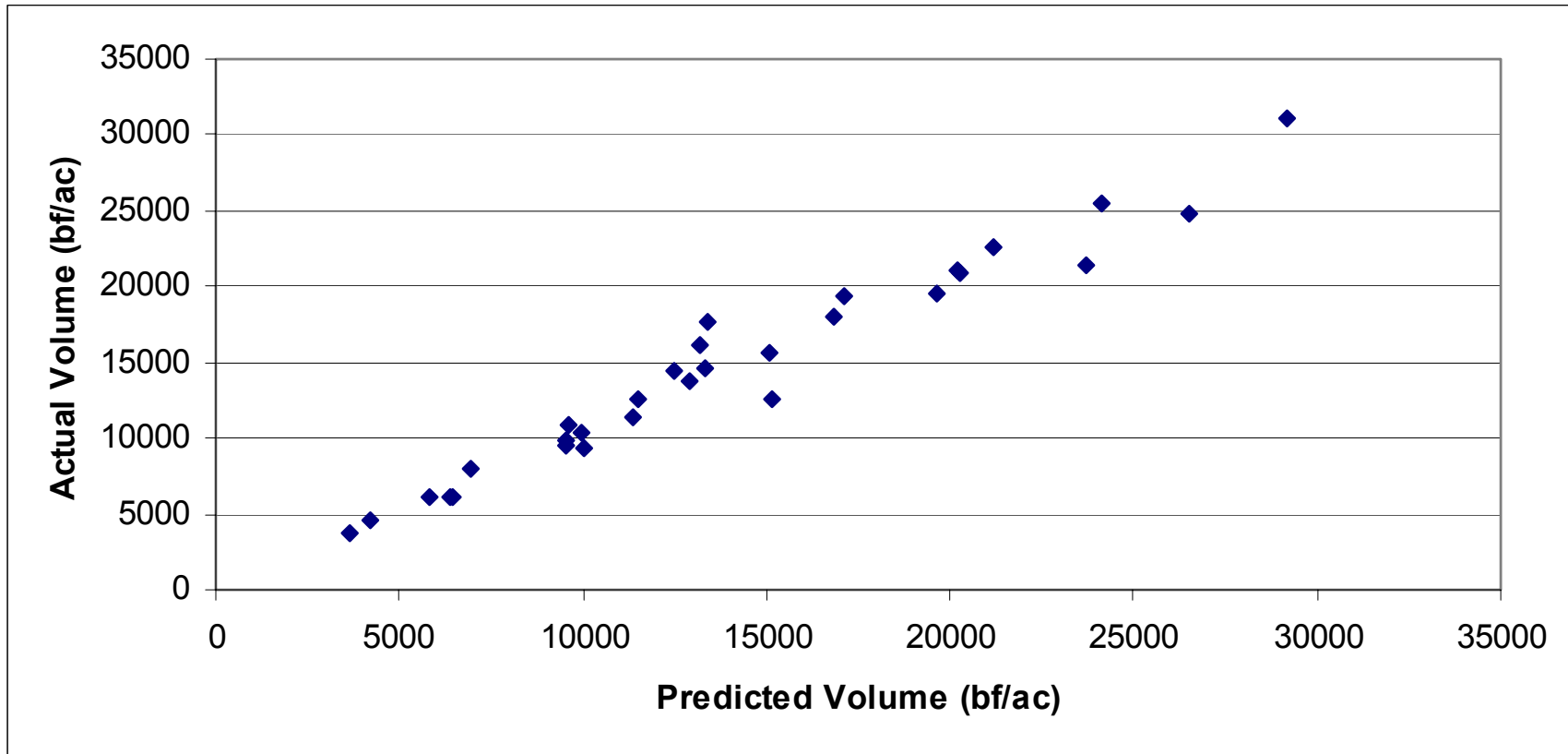


Figure 29. Relationship between actual and predicted volumes for intensive points for 500 simulations of the Trout Pond dataset (Correlation Coefficient = 0.979).

**Table 6. Boardfoot volume per acre by species group for each sampling method (Compartment 14).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	112.7	498.3	112.7	111.1	342.15	0.00	-1.42
Mixed Oak	3,081.8	1,957.2	3,127.6	3,084.9	-36.49	1.49	0.10
Other Hardwoods	124.8	423.5	149.6	147.5	239.34	19.87	18.19
Red Maple	54.3	0.0	42.7	42.1	-100.00	-21.36	-22.47
Red Oak	906.9	1,016.6	961.4	948.0	12.10	6.01	4.53
White Oak	188.1	292.8	225.8	222.7	55.66	20.04	18.39
Yellow-poplar	6,608.1	6,720.7	6,590.0	6,498.3	1.70	-0.27	-1.66

**Table 7. Boardfoot volume per acre by species group for each sampling method (Savage River).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	622.5	295.2	693.4	705.4	-52.58	11.39	13.32
Mixed Oak	1,573.0	1,907.0	1,567.3	1,594.2	21.23	-0.36	1.35
Other Conifers	1,346.9	1,141.3	1,431.0	1,455.5	-15.26	6.24	8.06
Other Hardwoods	1,155.8	912.2	1,212.1	1,232.9	-21.08	4.87	6.67
Red Maple	1,265.6	1,077.9	1,289.2	1,311.3	-14.83	1.86	3.61
Red Oak	3,032.0	1,959.7	3,000.1	3,051.6	-35.37	-1.05	0.65
Sugar Maple	540.9	295.5	542.9	552.2	-45.37	0.37	2.09
White Oak	524.1	484.9	538.4	547.7	-7.48	2.73	4.50
Yellow-poplar	284.3	226.2	258.9	263.4	-20.44	-8.93	-7.35

**Table 8. Boardfoot volume per acre by species group for each sampling method (WVU Research Forest).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	575.9	502.4	565.0	572.1	-12.76	-1.89	-0.66
Mixed Oak	1,629.1	1,564.3	1,608.8	1,629.1	-3.98	-1.25	0.00
Other Conifers	46.4	13.6	37.2	37.7	-70.69	-19.83	-18.75
Other Hardwoods	225.0	203.3	213.0	215.7	-9.64	-5.33	-4.13
Red Maple	506.8	529.6	505.1	511.5	4.50	-0.34	0.93
Red Oak	2,619.3	2,658.3	2,577.0	2,609.5	1.49	-1.61	-0.37
Sugar Maple	11.8	10.6	10.9	11.1	-10.17	-7.63	-5.93
White Oak	360.3	309.5	340.6	344.9	-14.10	-5.47	-4.27
Yellow-poplar	3,701.5	3,939.3	3,662.9	3,709.2	6.42	-1.04	0.21

**Table 9. Boardfoot volume per acre by species group for each sampling method (Coopers Rock).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	819.2	631.8	786.2	797.7	-22.88	-4.03	-2.62
Mixed Oak	1,356.8	1,497.3	1,359.1	1,378.9	10.36	0.17	1.63
Other Conifers	46.6	41.7	56.1	56.9	-10.52	20.39	22.10
Other Hardwoods	360.5	346.8	345.3	350.3	-3.80	-4.22	-2.83
Red Maple	1,510.2	1,460.3	1,508.2	1,530.2	-3.30	-0.13	1.32
Red Oak	1,477.3	1,773.8	1,474.3	1,495.8	20.07	-0.20	1.25
Sugar Maple	69.8	69.5	65.5	66.5	-0.43	-6.16	-4.73
White Oak	374.6	441.5	381.4	387.0	17.86	1.82	3.31
Yellow-poplar	4,628.7	4,164.2	4,612.9	4,680.2	-10.04	-0.34	1.11

**Table 10. Boardfoot volume per acre by species group for each sampling method (Coopers Rock Annex).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	173.2	106.6	167.5	166.0	-38.45	-3.29	-4.16
Mixed Oak	1,700.7	1,695.5	1,797.4	1,781.6	-0.31	5.69	4.76
Other Hardwoods	476.4	413.5	603.5	598.2	-13.20	26.68	25.57
Red Maple	649.1	740.0	647.4	641.7	14.00	-0.26	-1.14
Red Oak	1,200.6	821.3	1,182.8	1,172.4	-31.59	-1.48	-2.35
Sugar Maple	278.2	291.1	323.5	320.7	4.64	16.28	15.28
White Oak	758.0	778.1	768.9	762.1	2.65	1.44	0.54
Yellow-poplar	5,091.7	5,099.1	4,999.6	4,955.6	0.15	-1.81	-2.67



**Table 11. Boardfoot volume per acre by species group for each sampling method (Trout Pond dataset).**

<b>Species</b>	<b>Boardfoot Volume SRS</b>	<b>Boardfoot Volume DS</b>	<b>Boardfoot Volume HRS (srs)</b>	<b>Boardfoot Volume HRS (ratio)</b>	<b>% Difference from SRS Volume DS</b>	<b>% Difference from SRS Volume HRS (srs)</b>	<b>% Difference from SRS Volume HRS (ratio)</b>
Black Cherry	1,740.4	1,352.0	1,279.8	1,263.2	-22.32	-26.47	-27.42
Mixed Oak	660.4	664.9	651.1	642.6	0.68	-1.41	-2.70
Other Hardwoods	347.7	703.0	371.7	366.9	102.19	6.90	5.52
Red Maple	672.9	603.4	665.1	656.5	-10.33	-1.16	-2.44
Red Oak	4,026.2	4,189.8	4,003.8	3,952.0	4.06	-0.56	-1.84
White Oak	1,117.5	1,588.2	1,080.2	1,066.2	42.12	-3.34	-4.59
Yellow-poplar	6,026.8	5,897.0	5,847.2	5,771.4	-2.15	-2.98	-4.24

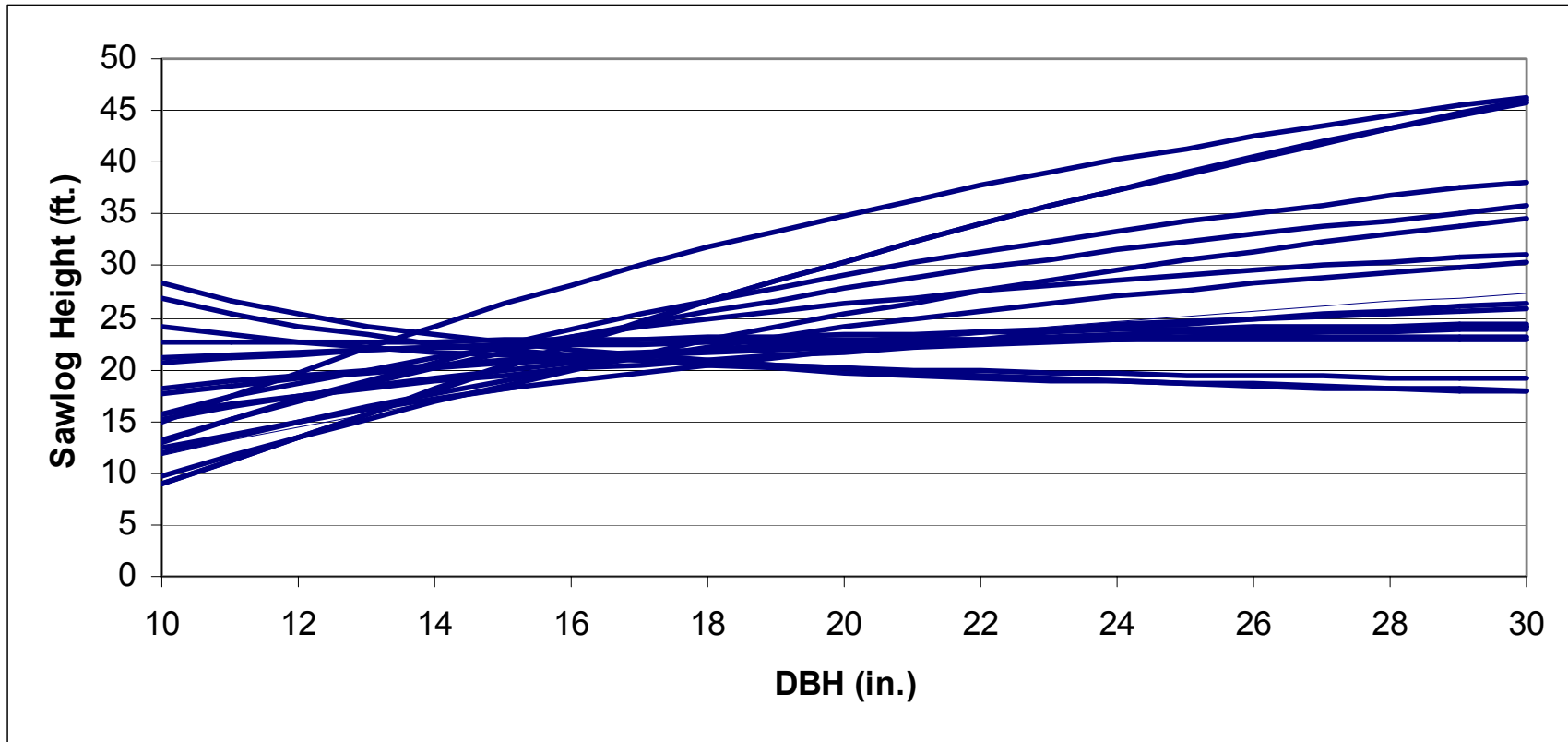


Figure 30. Predicted height curves for chestnut oak in the first 20 of the 500 simulations of Compartment 14. Min=6 ; Max=24.

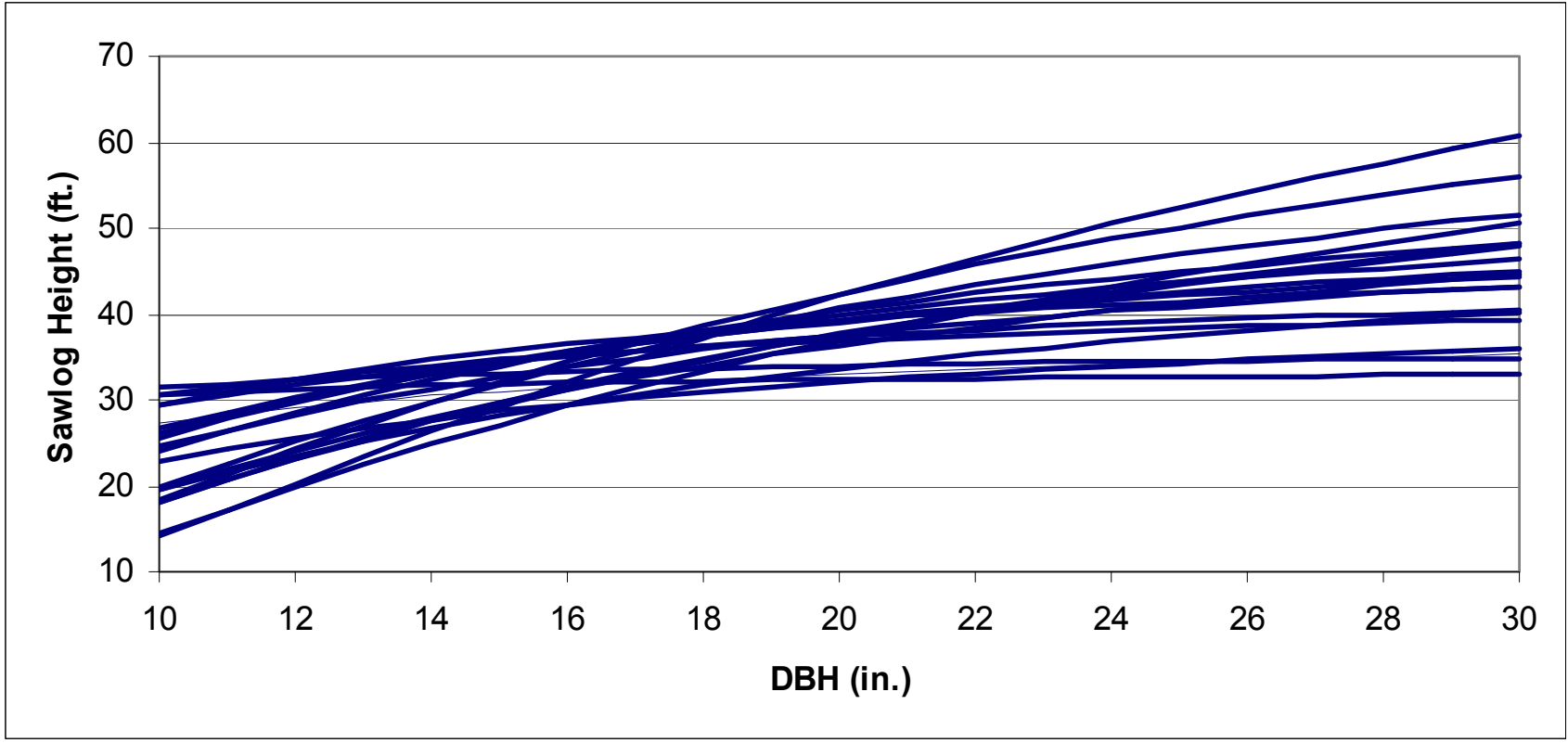
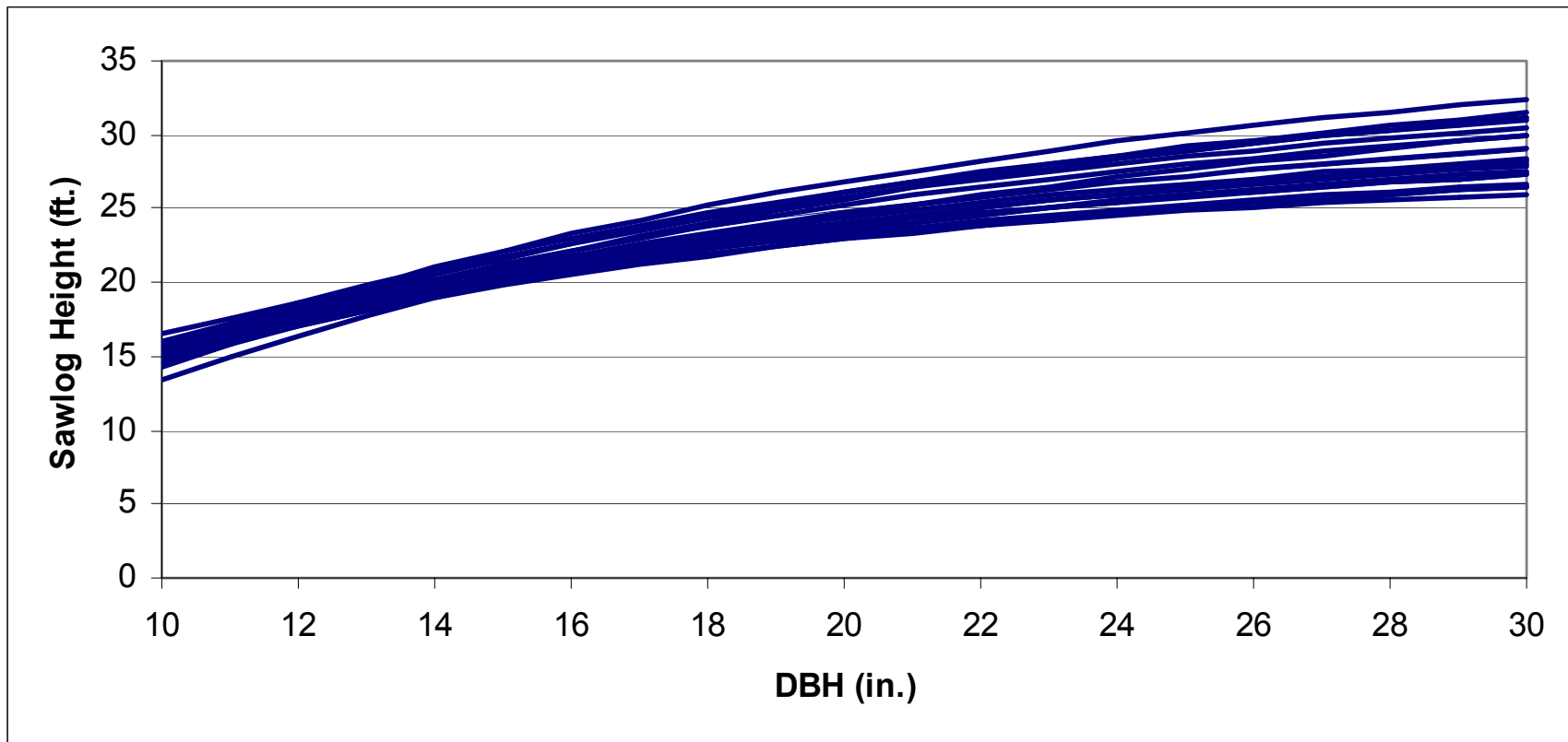
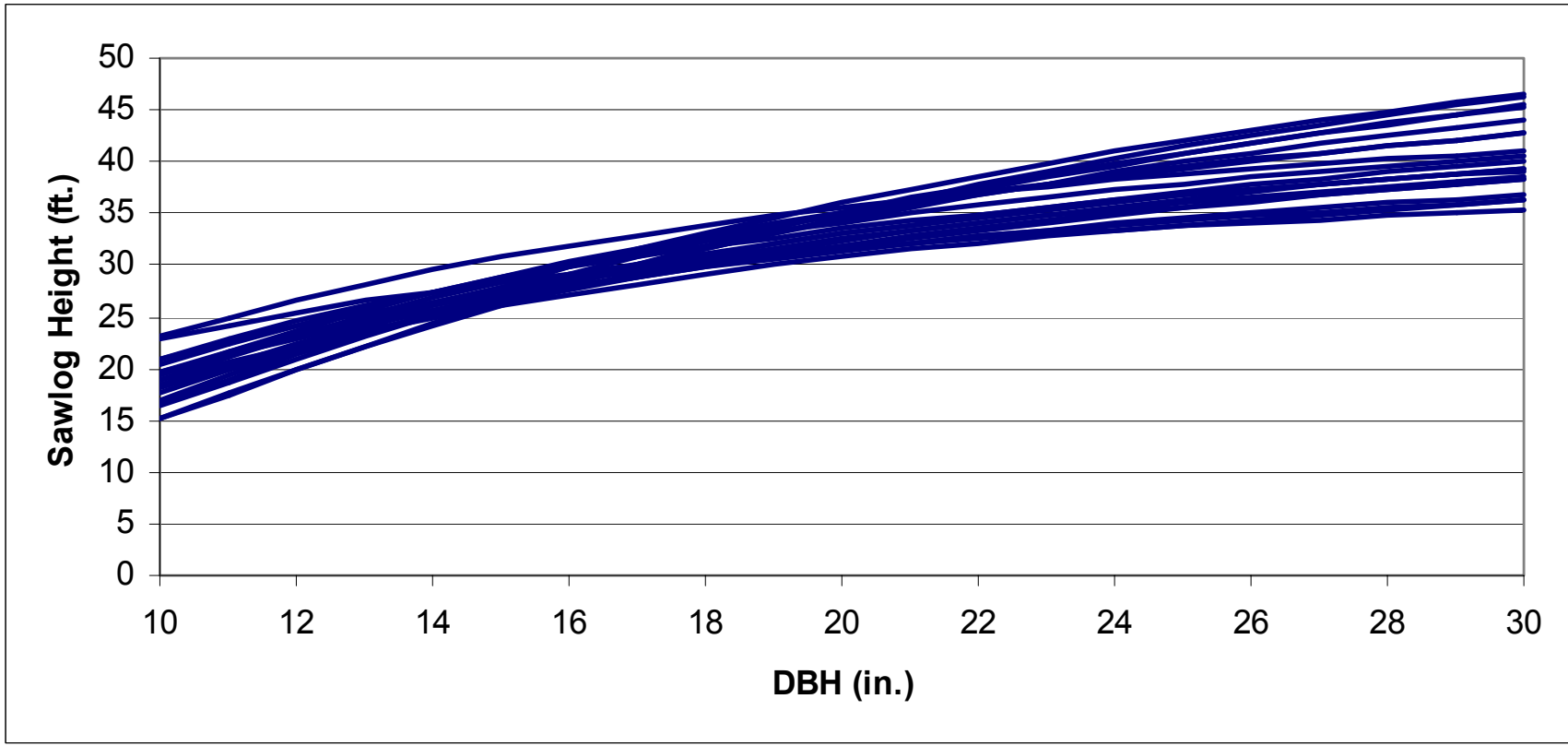


Figure 31. Predicted height curves for white oak in the first 20 of the 500 simulations of Savage River. Min=6; Max=20.



**Figure 32. Predicted height curves for all trees in the first 20 of the 500 simulations of 1967 single species dataset. Min=270; Max=364.**



**Figure 33. Predicted height curves for black cherry in the first 20 of the 500 simulations of WVU Research Forest. Min=99; Max=151.**

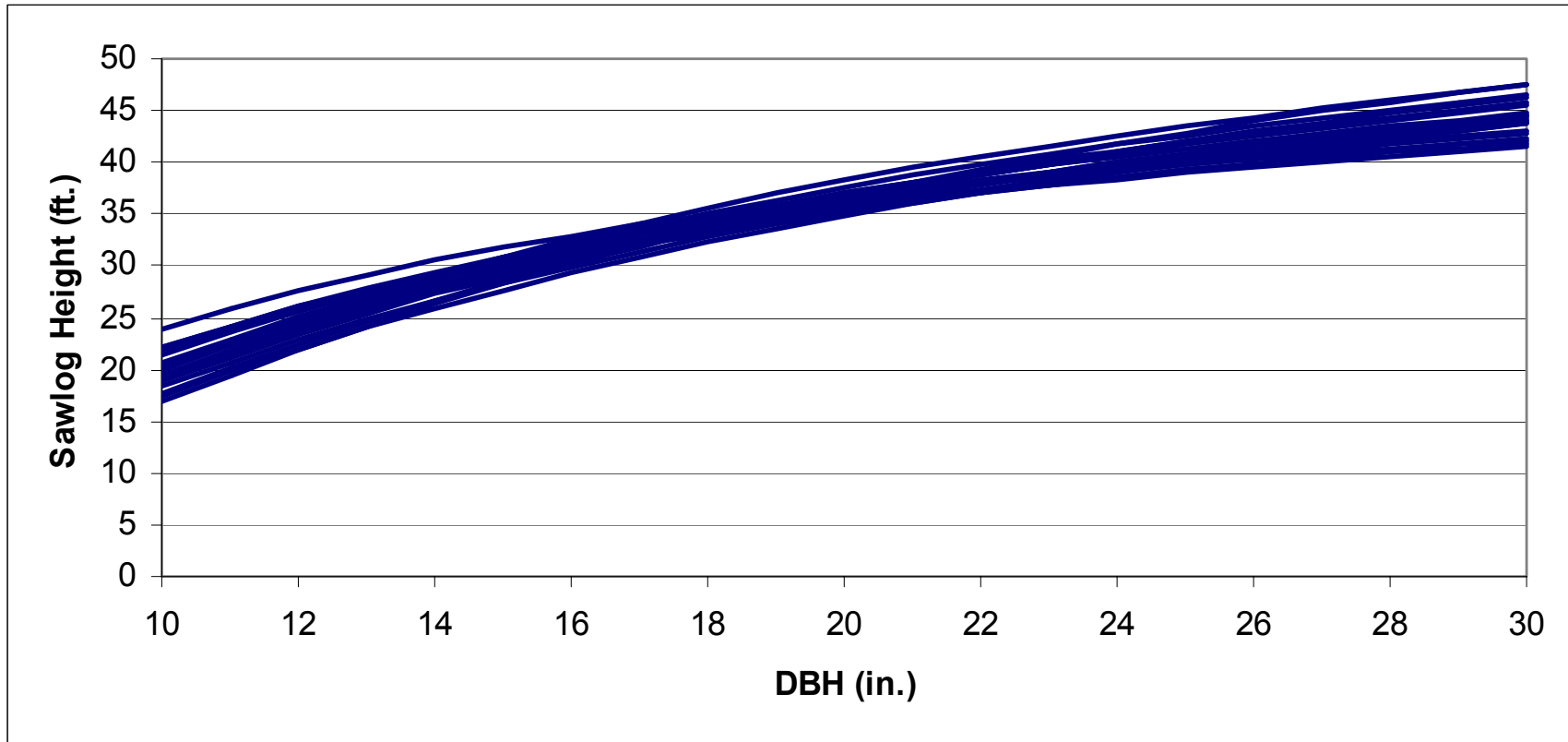
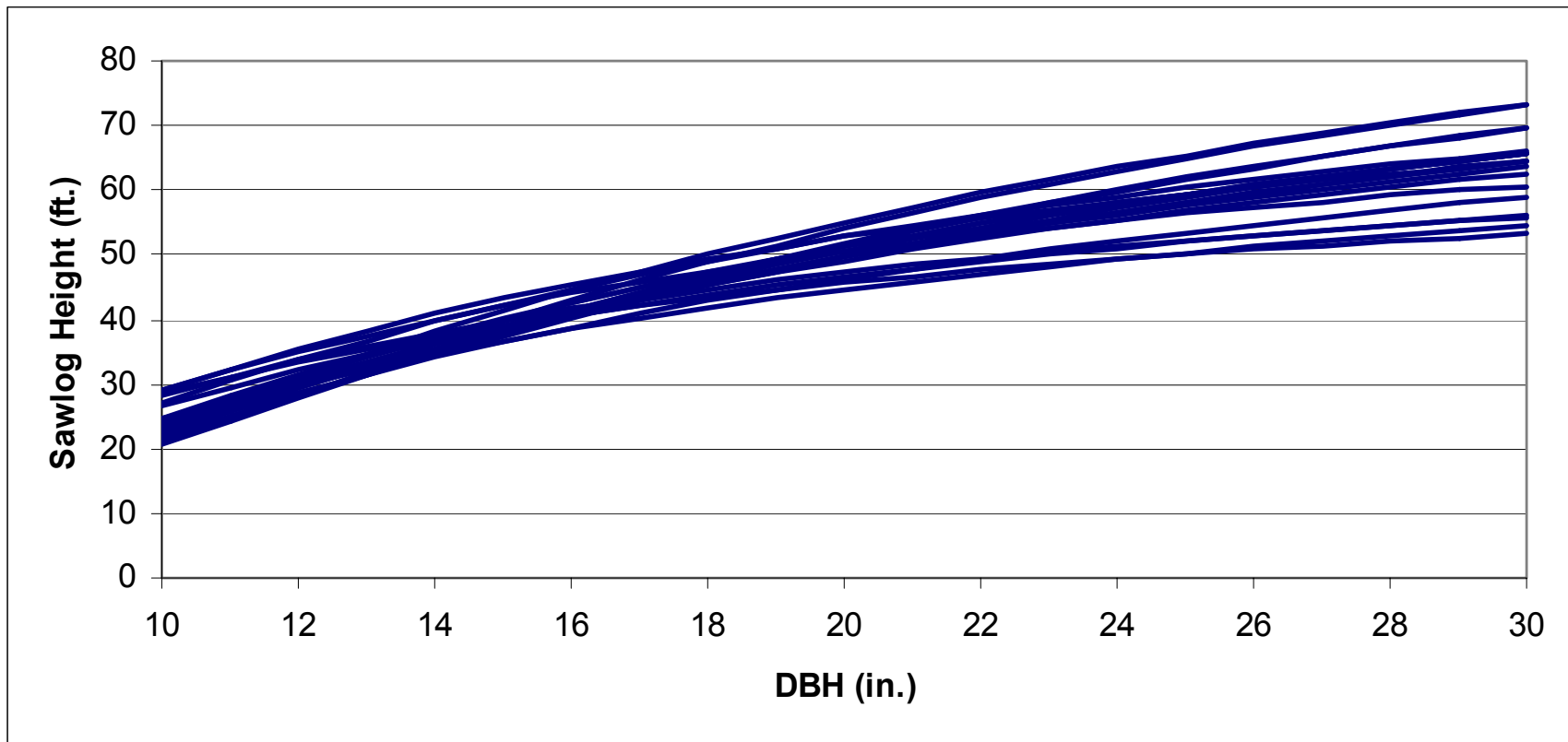
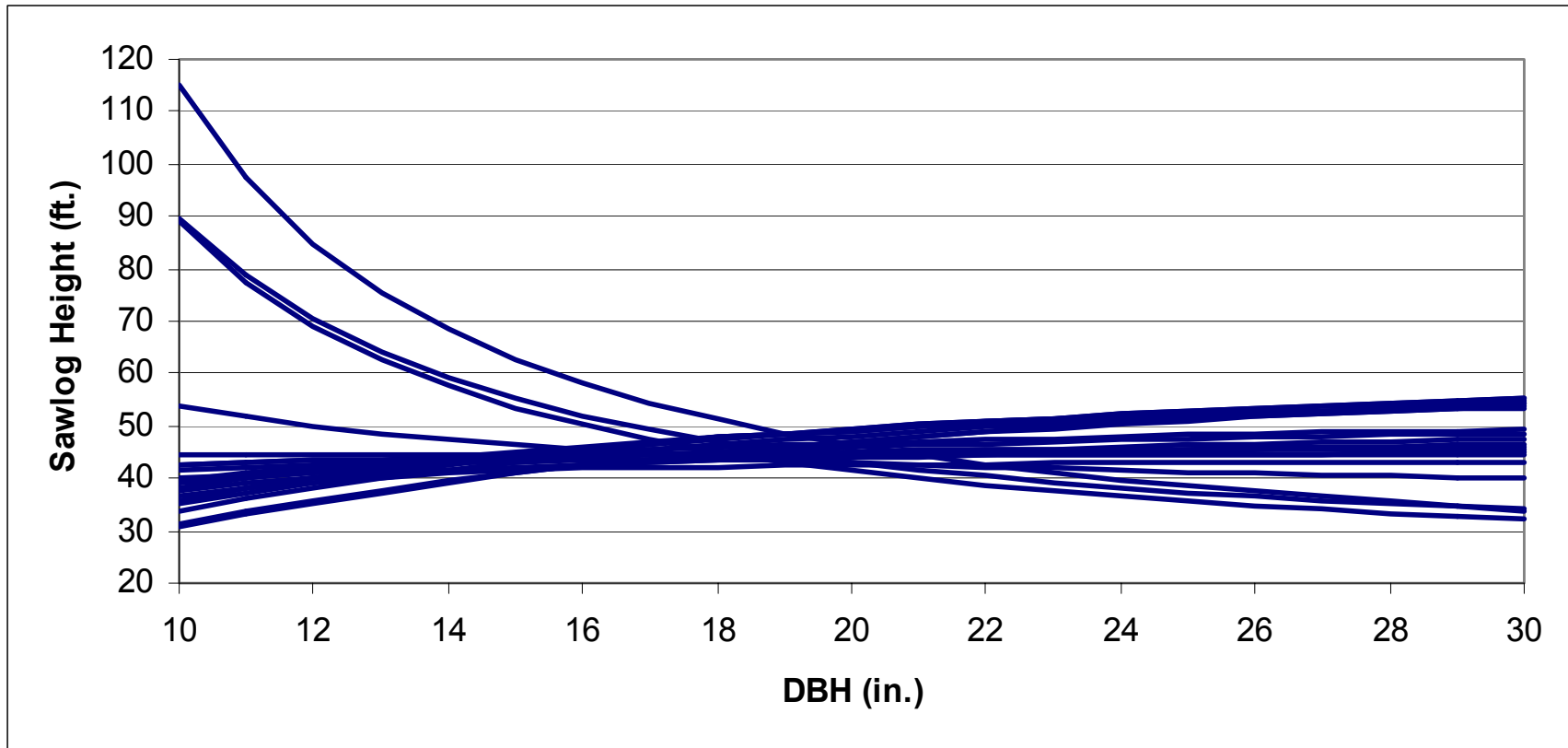


Figure 34. Predicted height curves for red oak in the first 20 of the 500 simulations of Coopers Rock State Forest. Min=149 to 216.



**Figure 35. Predicted height curves for yellow-poplar in the first 20 of the 500 simulations of Coopers Rock Annex. Min= 20, Max=57.**



**Figure 36. Predicted height curves for yellow-poplar in the first 20 of the 500 simulations of Trout Pond dataset. Min=5; Max=34.**



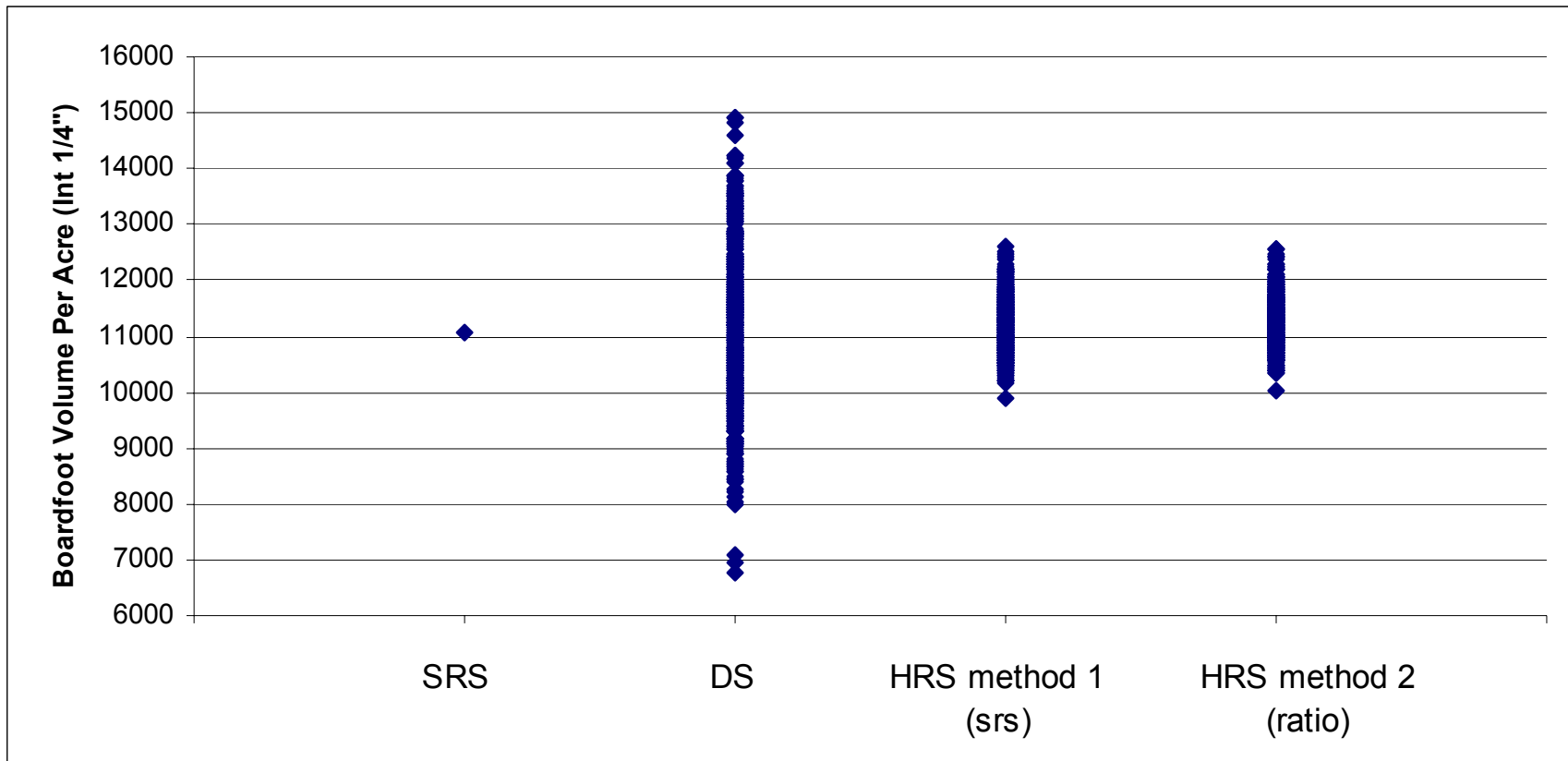


Figure 37. Variation in mean boardfoot volume per acre for Compartment 14 of the WV University Forest after 500 simulations by sampling type.

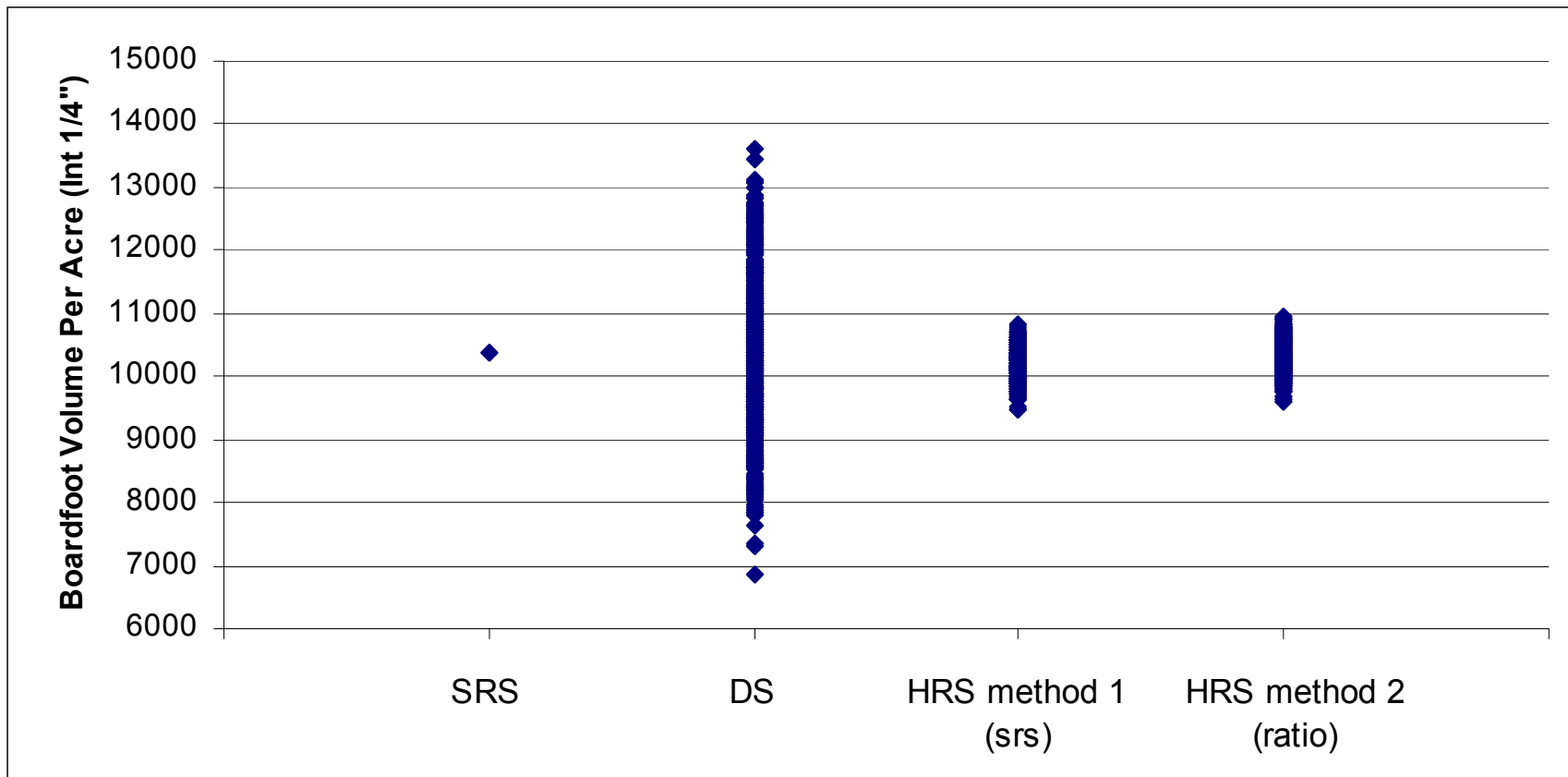


Figure 38. Variation in mean boardfoot volume per acre for Savage River after 500 simulations by sampling type.

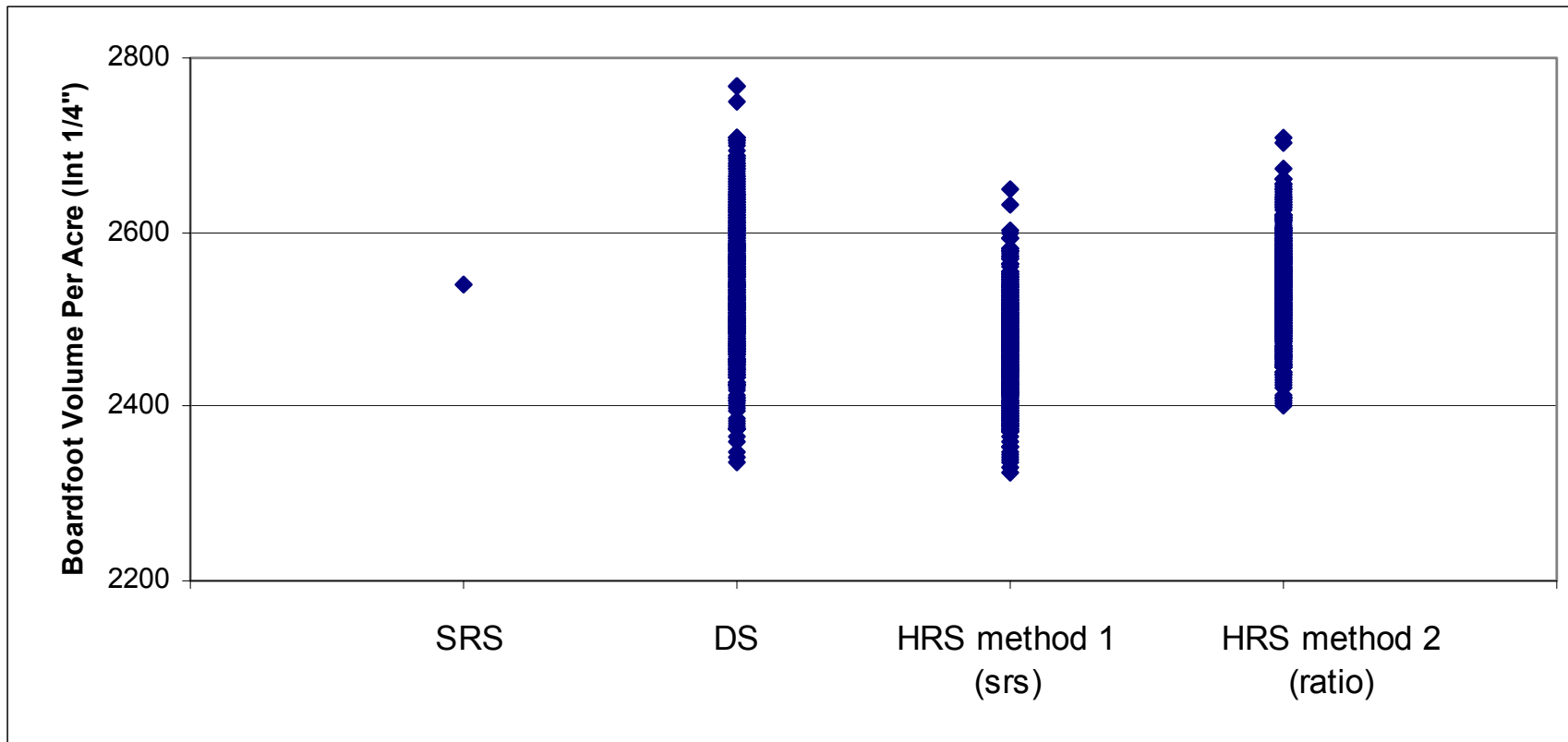


Figure 39. Variation in mean boardfoot volume per acre for 1967 single species data after 500 simulations by sampling type.

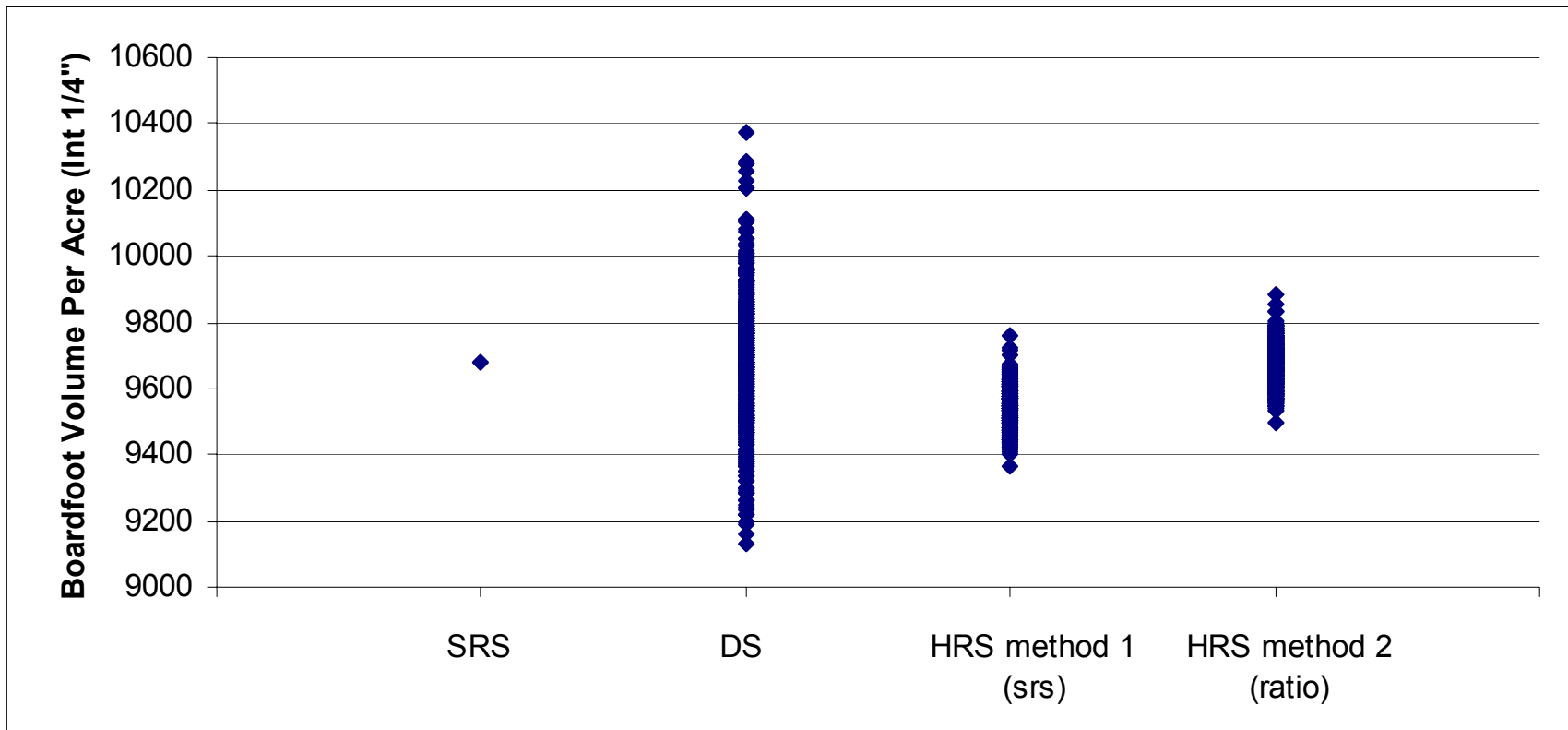


Figure 40. Variation in mean boardfoot volume per acre for WVU Research Forest after 500 simulations by sampling type.

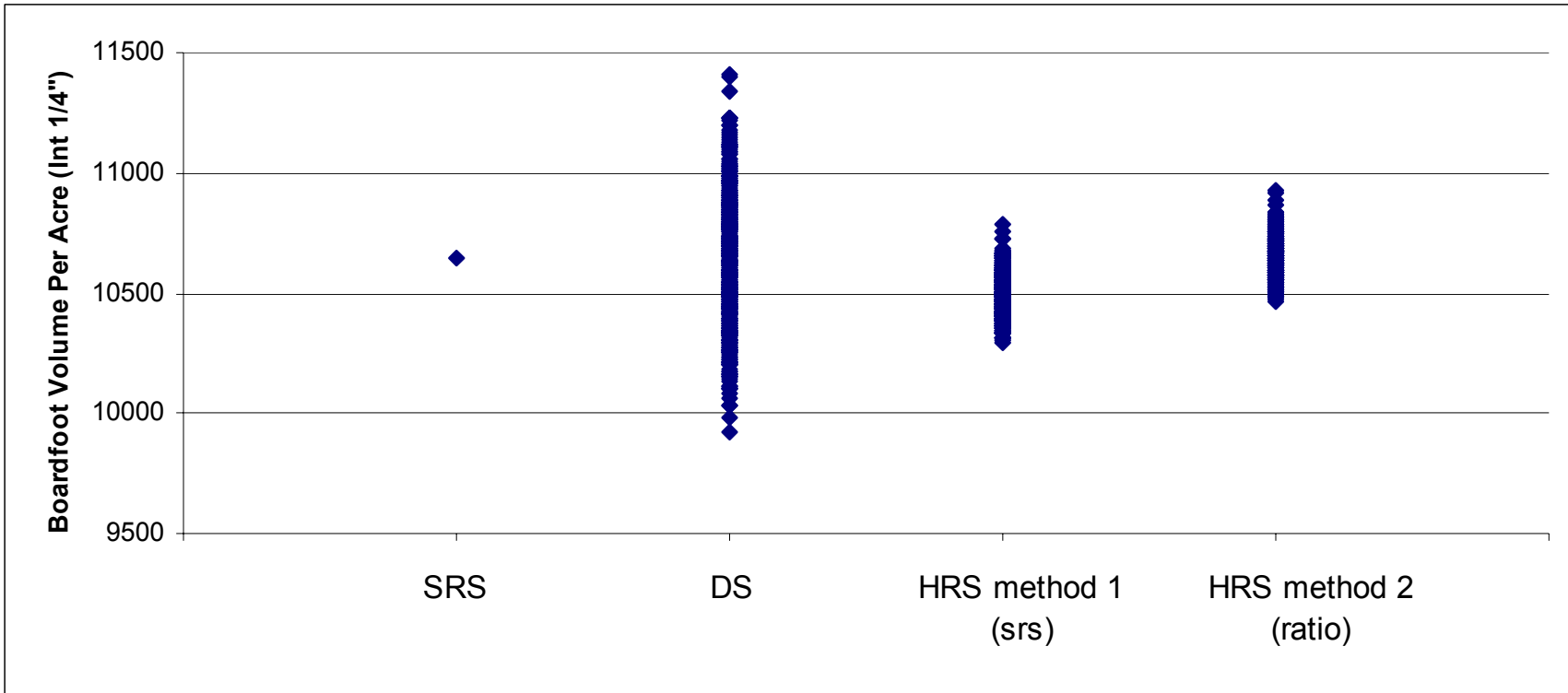


Figure 41. Variation in mean boardfoot volume per acre for Coopers Rock State Forest after 500 simulations by sampling type.

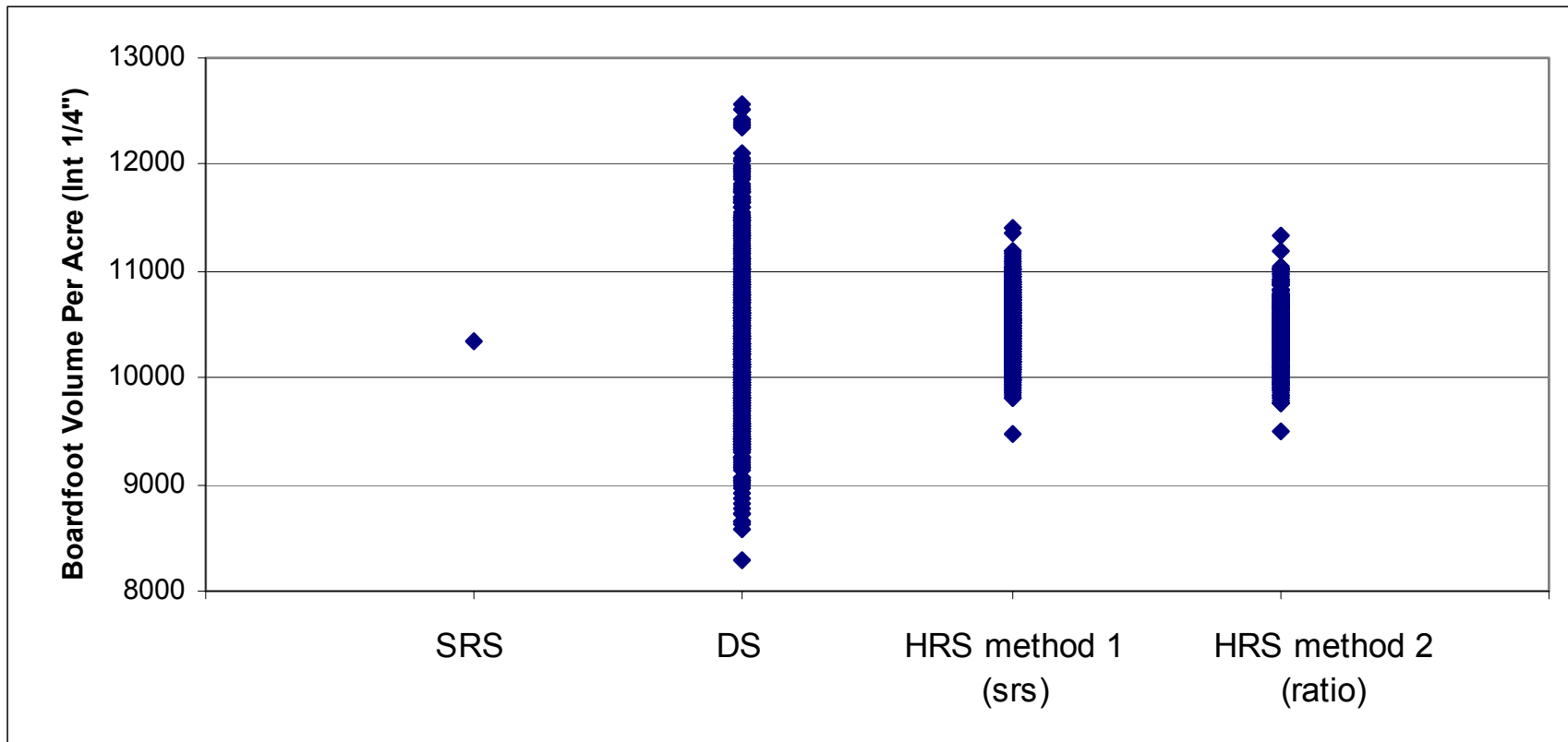


Figure 42. Variation in mean boardfoot volume per acre for Coopers Rock Annex after 500 simulations by sampling type.

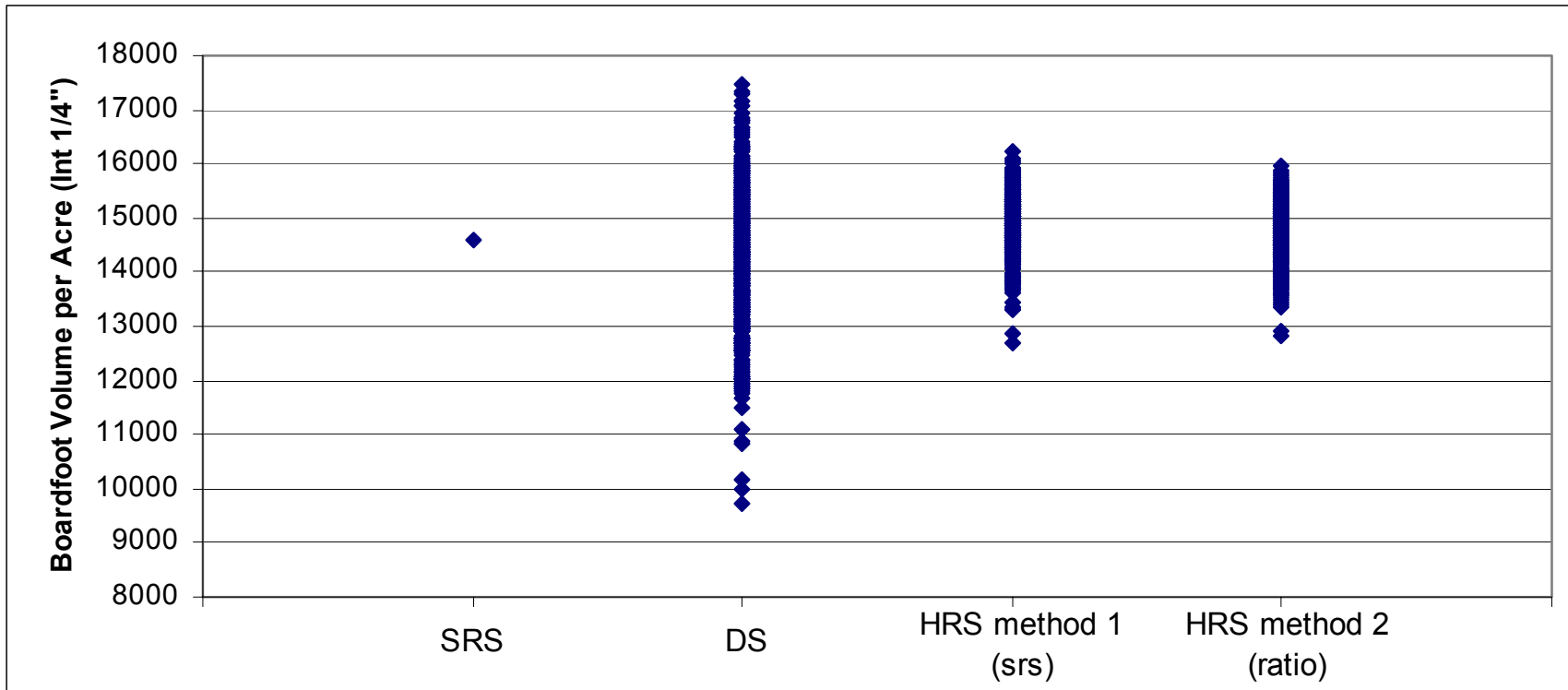


Figure 43. Variation in mean boardfoot volume per acre for Trout Pond dataset after 500 simulations by sampling type.

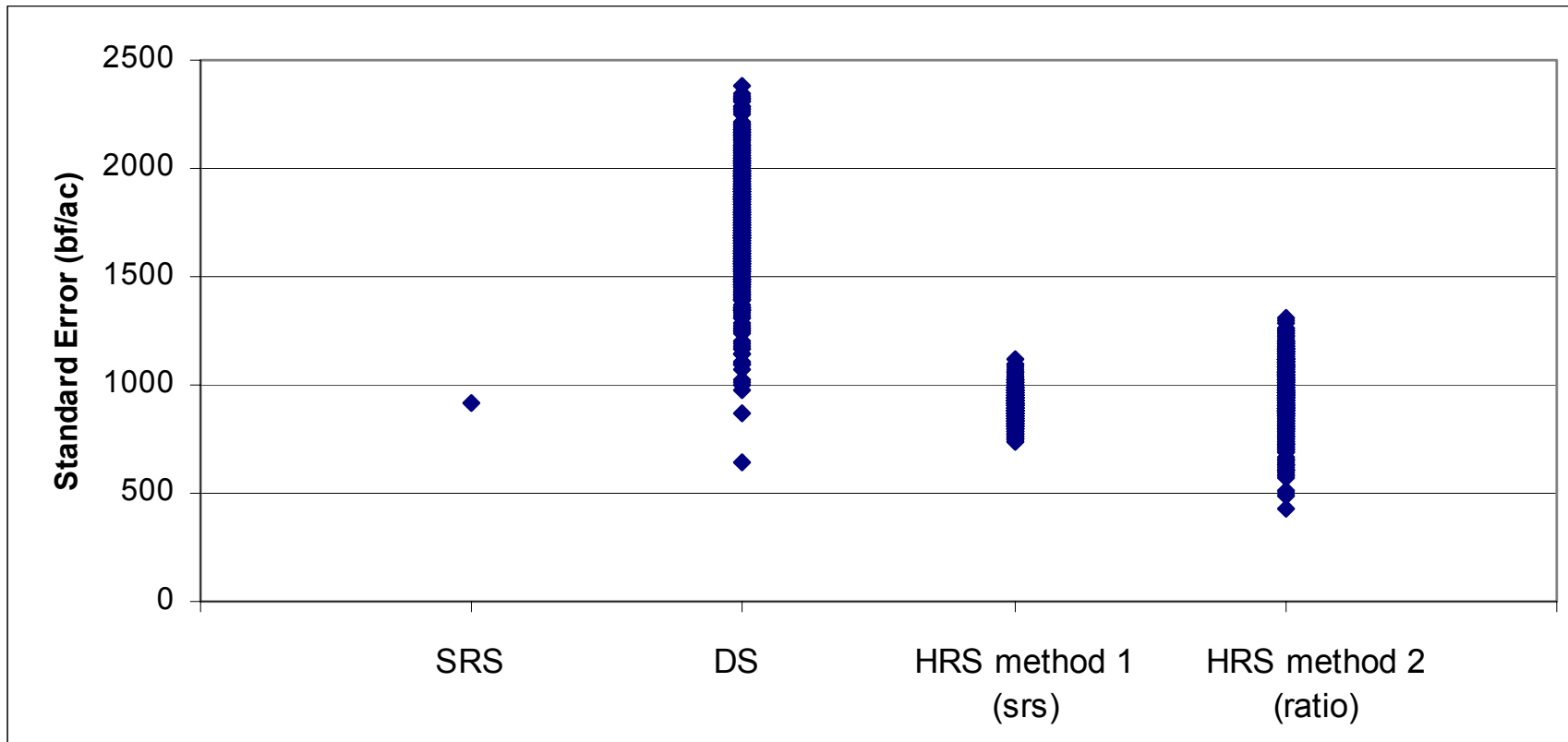


Figure 44. Standard error for Compartment 14 after 500 simulations by sampling type.



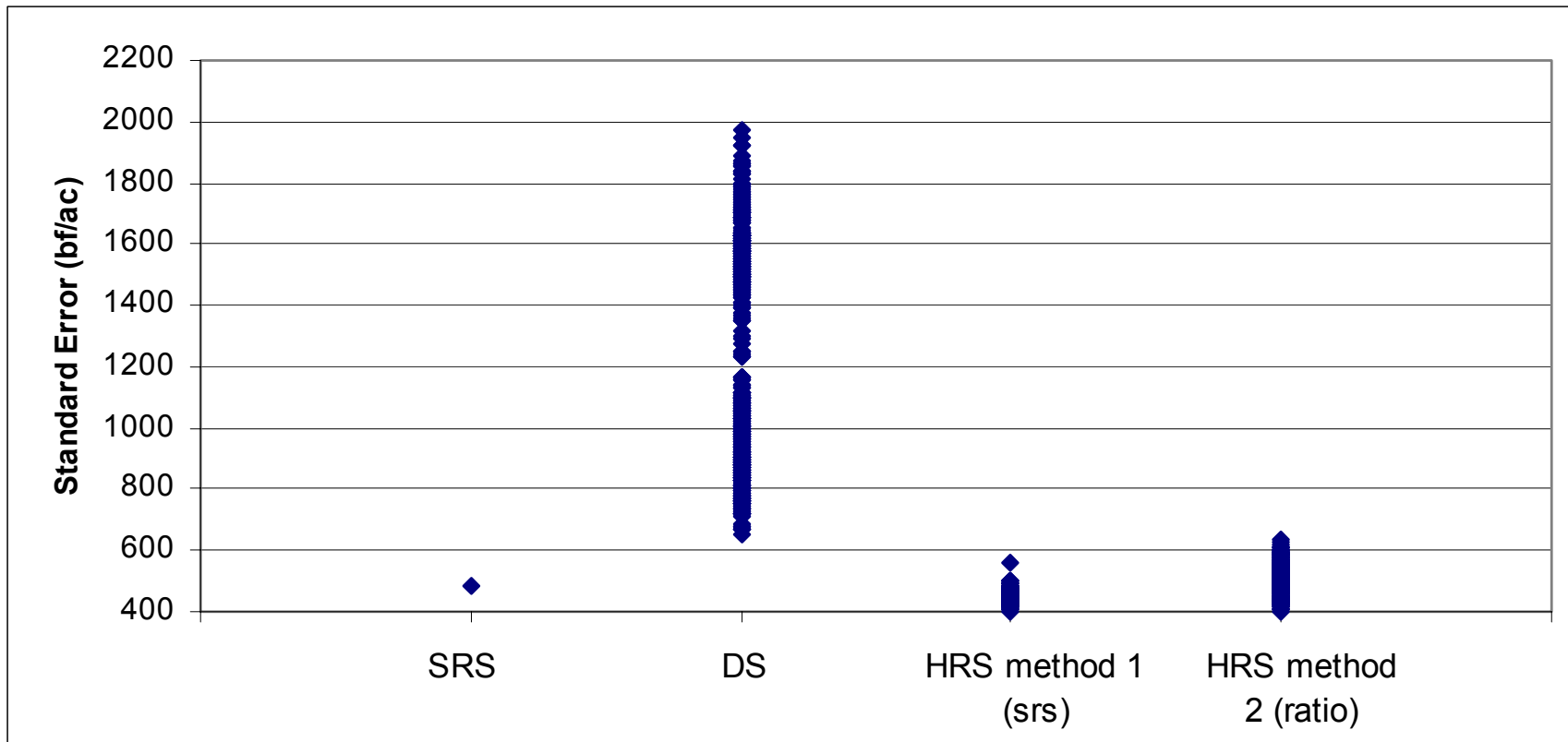


Figure 45. Standard error for Savage River after 500 simulations by sampling type.

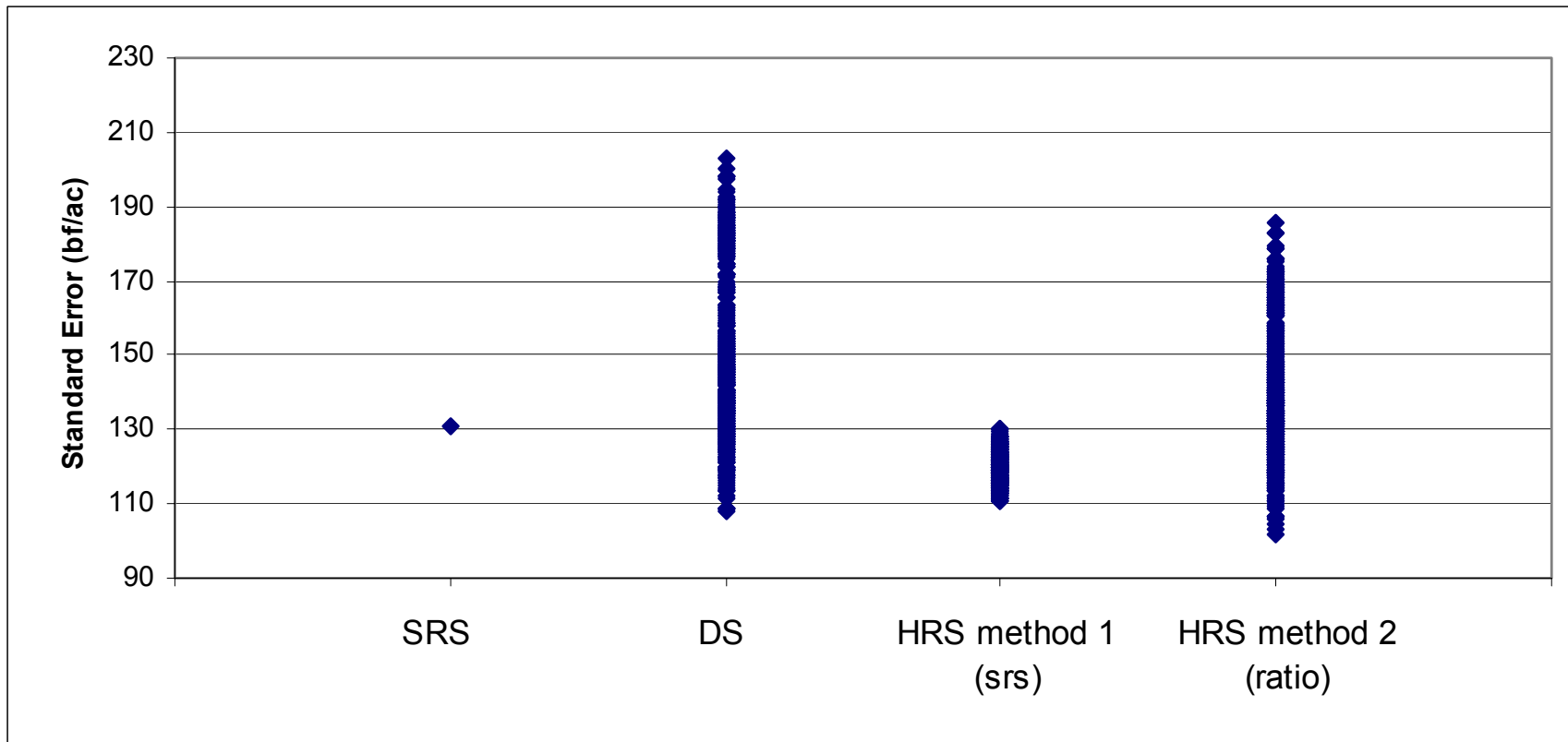


Figure 46. Standard error for 1967 single species data after 500 simulations by sampling type.

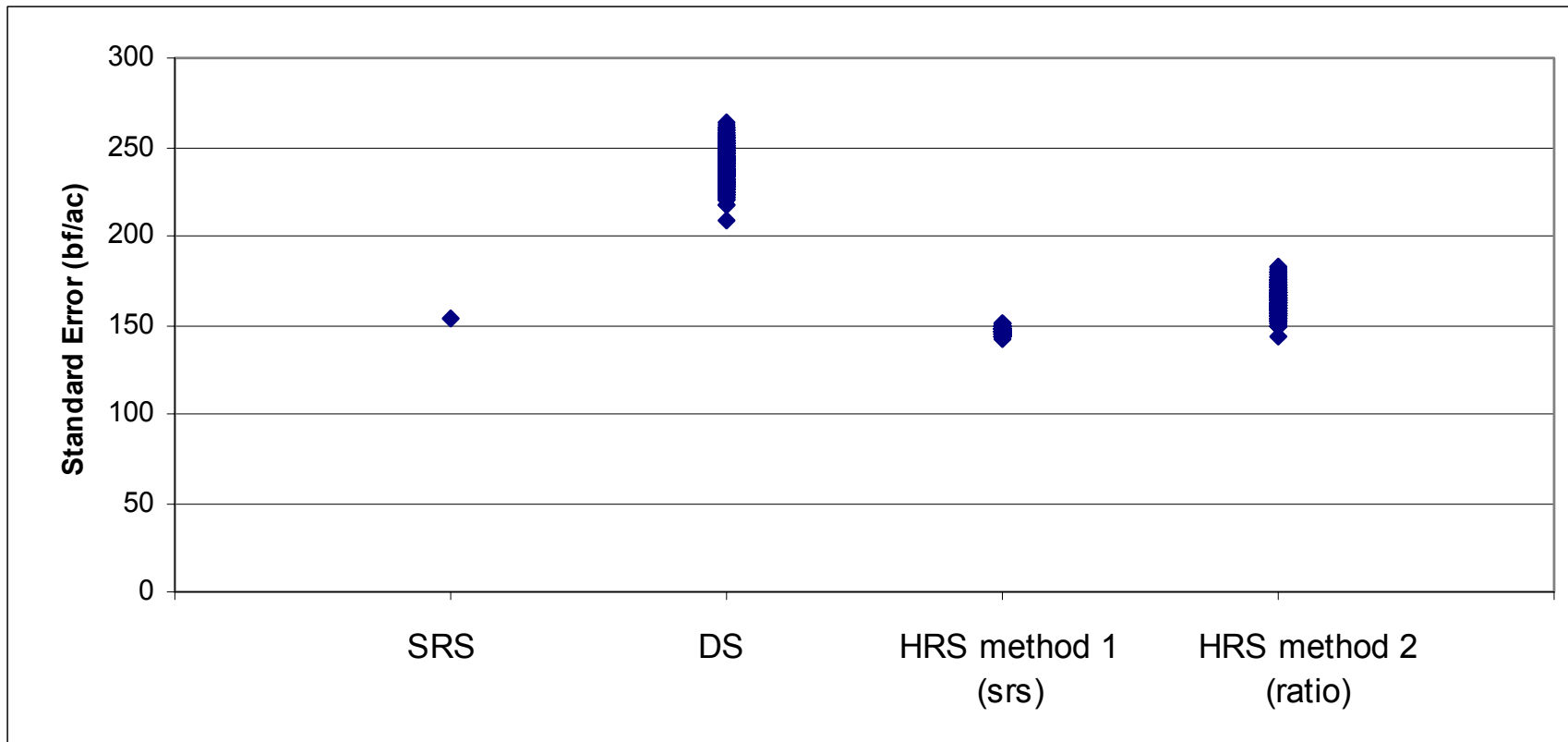


Figure 47. Standard error for WVU Research Forest after 500 simulations by sampling type.

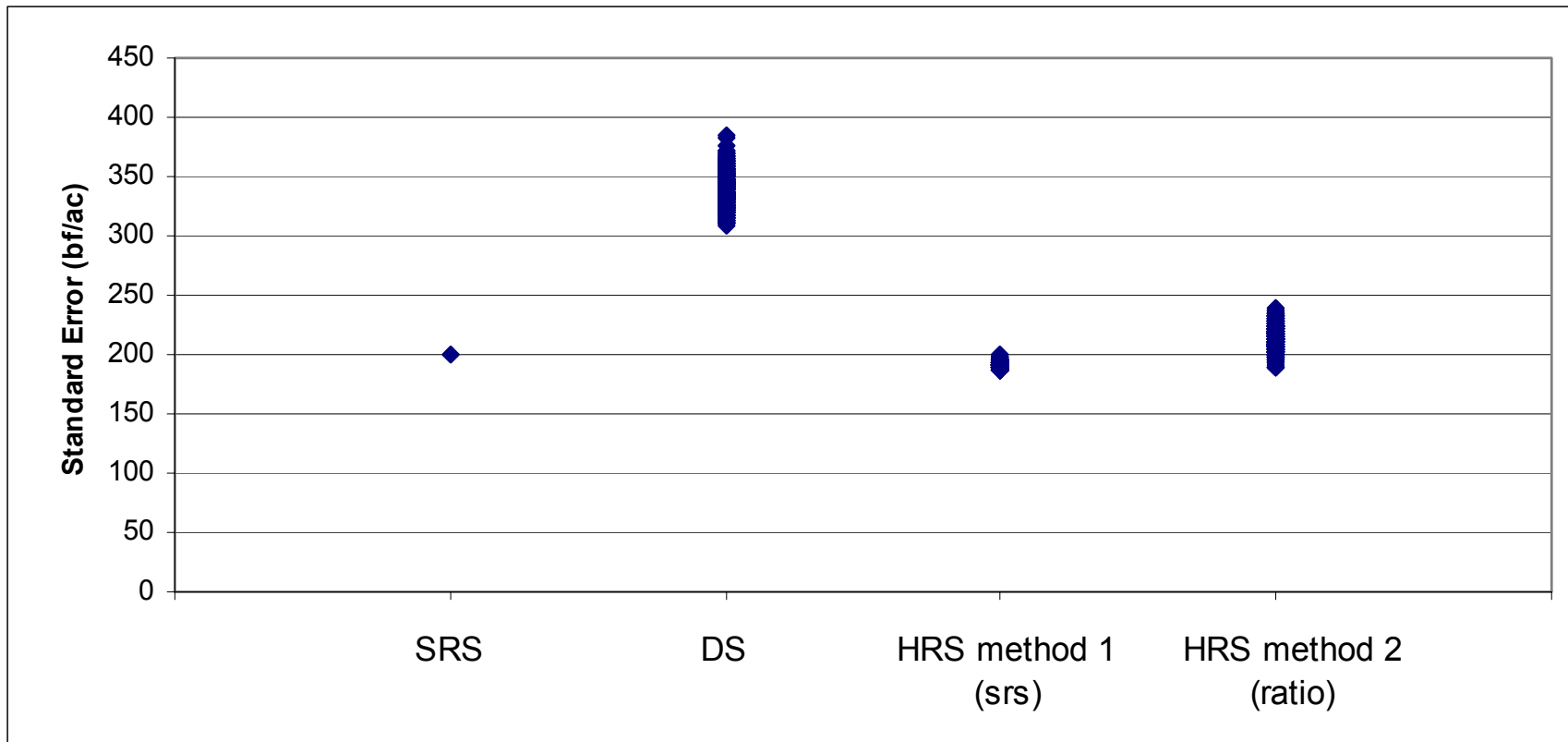


Figure 48. Standard error for Coopers Rock State Forest after 500 simulations by sampling type.

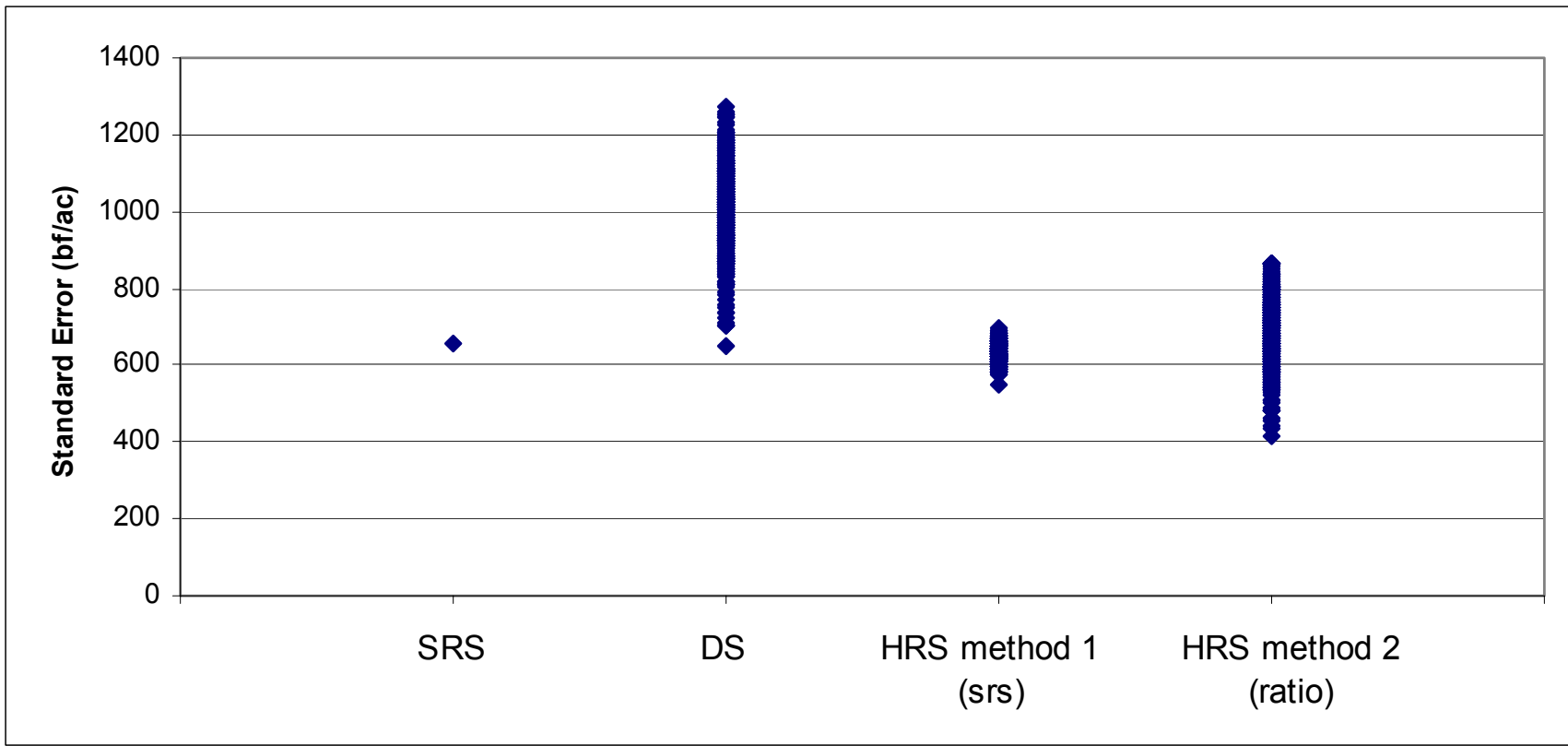


Figure 49. Standard error for Coopers Rock Annex after 500 simulations by sampling type.

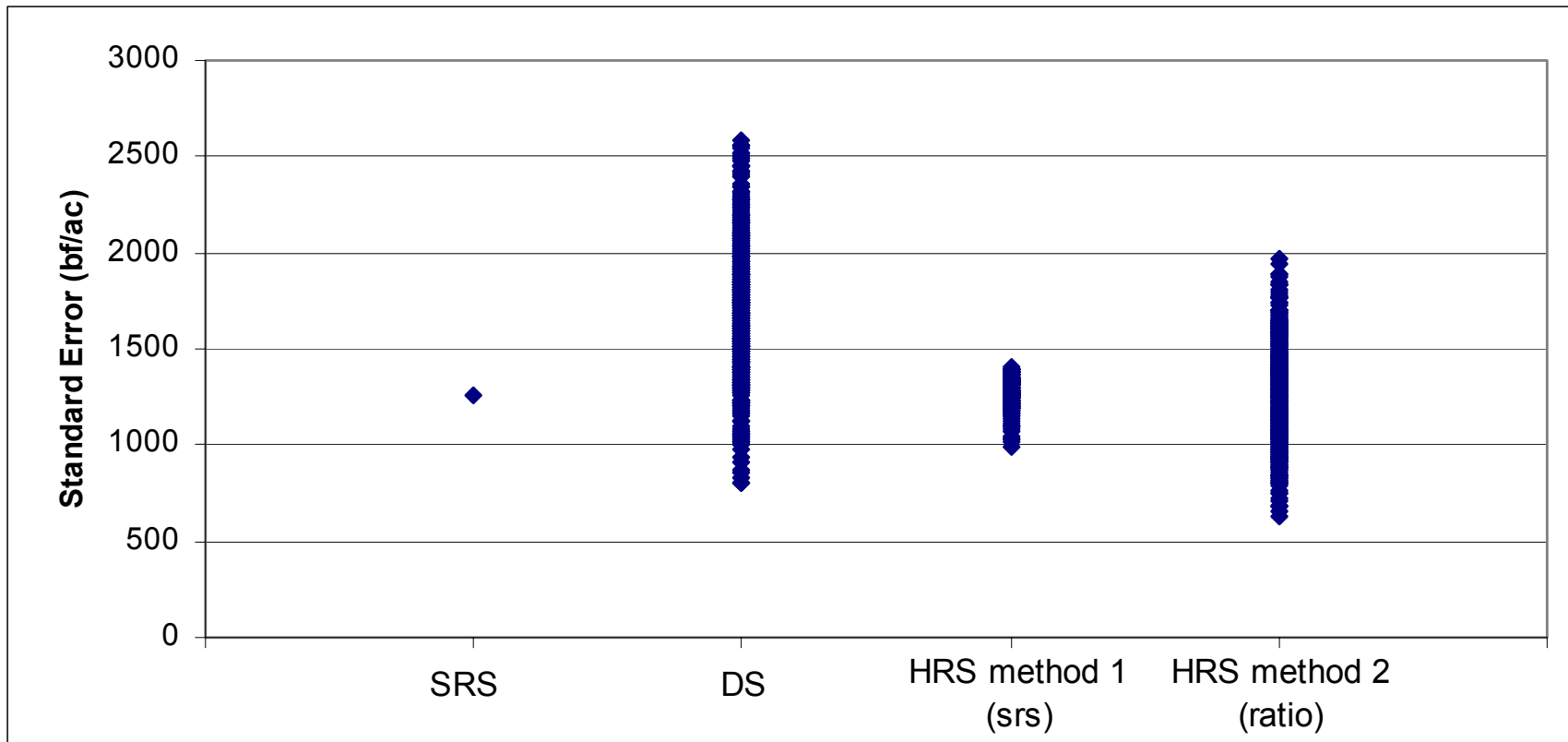


Figure 50. Standard error for Trout Pond dataset after 500 simulations by sampling type.