

# Short-Rotation Woody Crops Program

at  
State University of New York  
College of Environmental Science & Forestry

**Biomass Power for Rural Development**

**Technical Report:**

## **ABOVEGROUND BIOMASS EQUATION DEVELOPMENT FOR FIVE *SALIX* CLONES AND ONE *POPULUS* CLONE Interim Program Report**

**Prepared for the United States Department of Energy  
Under Cooperative Agreement No. DE-FC36-96GO10132**

**Submitted to:  
Edward Neuhauser, Project Manager  
Niagara Mohawk Power Corporation  
Syracuse, New York**

**by:  
B.D. Ballard, S.V. Stehman, R.D. Briggs, T.A.  
Volk, L.P. Abrahamson, and E.H. White**

**March 2000**

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
<b>INTRODUCTION .....</b>	<b>1</b>
BACKGROUND .....	1
THE WILLOW BIOENERGY SYSTEM .....	2
APPLICATION TO COMMERCIALIZATION .....	2
OBJECTIVES .....	2
PRESENT STATUS .....	3
FUTURE PROGRESS .....	3
QUESTIONS TO BE ADDRESSED .....	4
<b>METHODS.....</b>	<b>5</b>
COLLECTION OF BIOMASS EQUATION DATA .....	5
DATA MANAGEMENT .....	6
MODEL DEVELOPMENT AND SELECTION .....	6
VALIDATION.....	7
<b>RESULTS AND DISCUSSION .....</b>	<b>7</b>
SIMPLE LINEAR REGRESSION MODEL USING RAW DATA.....	8
LINEAR REGRESSION MODELS USING $D^2$ AND H AS PREDICTOR VARIABLES .....	8
LOGARITHMIC REGRESSIONS.....	9
NON-LINEAR REGRESSION.....	11
VALIDATION.....	11
<i>Age Generalization</i> .....	12
<i>Comparisons of Bias Correction Methods</i> .....	12
EQUATION COEFFICIENTS FOR THE LOGARITHMIC REGRESSIONS.....	16
<b>CONCLUSION.....</b>	<b>18</b>
<b>LITERATURE CITED.....</b>	<b>19</b>
<b>APPENDIX .....</b>	<b>20</b>

## LIST OF FIGURES

FIGURE 1. LOCATION OF THE FERTILIZATION TRIAL PLANTINGS IN NEW YORK STATE: KING FERRY, SOMERSET, AND TULLY. BASE MAP FROM MICROSOFT <sup>®</sup> AUTOMAP STREETS PLUS. ....	5
FIGURE 2. MEASUREMENT OF STEM DIAMETER WAS MADE AT 30 CM FROM THE BASE OF THE STOOL FOLLOWING THE CURVATURE OF THE STEM (A). SECONDARY BRANCHES FORKING BELOW 30 CM WERE MEASURED FROM THE BASE OF THE FORK (B), WHILE THE MAIN STEM OF A FORK WAS MEASURED FROM THE BASE OF THE STOOL (C). ....	6
FIGURE 3. OVEN-DRY STEM BIOMASS (G) VERSUS STEM DIAMETER (MM) MEASURED AT 30 CM FROM THE BASE OF THE STOOL FOR SAMPLES OF CLONE SV1 COLLECTED IN RESEARCH PLANTINGS IN NEW YORK STATE IN WINTER, 1996 AND 1997. ....	8
FIGURE 4. RESIDUAL PLOT (RESIDUALS VS. PREDICTED VALUES) OF THE NON-TRANSFORMED REGRESSION FOR STEMS OF CLONE SV1. ....	8
FIGURE 5. RESIDUAL PLOT (RESIDUALS VS. PREDICTED) OF THE REGRESSION USING $D^2$ AS THE PREDICTOR VARIABLE FOR STEMS OF ALL CLONES. ....	9
FIGURE 6. RESIDUAL PLOTS FOR THE STEM-BASED LOGARITHMIC REGRESSION FOR CLONE SV1: (A) RESIDUAL VERSUS PREDICTED VALUES AND (B) NORMAL PROBABILITY PLOT. ....	10

FIGURE 7. RESIDUAL PLOT (RESIDUALS VERSUS PREDICTED) OF THE NON-LINEAR REGRESSION FOR STEMS OF CLONE SV1. $R^2$ FOR REGRESSION WAS 0.95. ....	11
FIGURE 8. PREDICTION ERRORS OF THE GENERALIZED LOGARITHMIC EQUATION FOR THE SV1 VALIDATION DATA SET. (NOTE: PREDICTION ERROR = OBSERVED – PREDICTED). ....	14
FIGURE 9. PREDICTION ERRORS OF THE LOGARITHMIC REGRESSION FOR THE SA2 VALIDATION DATA SET: A) CLONE-SPECIFIC (SA2) PREDICTION EQUATION, MSEP = 54.0 AND B) GENERALIZED (WILLOW) PREDICTION EQUATION, MSEP = 33.8. ....	14
FIGURE 10. PLOT OF PREDICTION ERRORS OF THE CLONE-SPECIFIC EQUATIONS FOR THE SA2 VALIDATION DATA SET: A) LOGARITHMIC REGRESSION ESTIMATE; B) LOGARITHMIC REGRESSION ESTIMATE WITH BASKERVILLE’S (1972) BIAS CORRECTION; C) LOGARITHMIC REGRESSION ESTIMATE WITH SNOWDON’S (1991) BIAS CORRECTION. ....	15

## LIST OF TABLES

TABLE 1. WILLOW AND POPLAR CLONES AND THEIR PARENTAGE AND ORIGIN IN VARIOUS PLANTINGS IN NEW YORK STATE. ....	4
TABLE 2. THE MEAN SQUARE ERROR OF THE PREDICTION (MSEP) WAS CALCULATED FOR THE FOLLOWING STEM-BASED MODELS USING VERIFICATION DATA: (1) CLONE-SPECIFIC LOGARITHMIC REGRESSIONS, (2) GENERALIZED (ALL CLONES, INCLUDING POPLAR) AND (3) WILLOW ONLY LOGARITHMIC REGRESSIONS: (4) NON-LINEAR CLONE-SPECIFIC MODELS: $W = AD^B$ AND (5) $W = AD^B + C$ . THE MSEP WAS ALSO CALCULATED USING THE MODEL-SELECTION DATA SET (ON THE MODEL-SELECTION EQUATIONS) AND IS REPORTED FOR REFERENCE. ....	13
TABLE 3. THE MEAN SQUARE ERROR OF THE PREDICTION (MSEP) WAS CALCULATED FOR THE CLONE-SPECIFIC, STEM-BASED LOGARITHMIC REGRESSION MODELS USING VERIFICATION DATA: (A) “UNCORRECTED” LOGARITHMIC REGRESSION ESTIMATES, (B) BIAS CORRECTED REGRESSION ESTIMATES USING BASKERVILLE’S MULTIPLICATIVE CORRECTION ( $\exp(S_e^2/2)$ ) AND (C) SNOWDON’S RATIO ESTIMATOR CORRECTION; THE MSEP WAS ALSO CALCULATED USING THE MODEL-SELECTION DATA SET (ON THE MODEL-SELECTION EQUATIONS) AND IS REPORTED FOR REFERENCE. ....	15
TABLE 4. STATISTICS OF THE GENERALIZED LOGARITHMIC REGRESSIONS USING THE MODEL-SELECTION, THE VALIDATION, AND THE COMPLETE DATA SETS. BCF IS FROM SNOWDON (1991). ....	16
TABLE 5. STATISTICS OF THE CLONE-SPECIFIC LOGARITHMIC REGRESSIONS USING THE MODEL-SELECTION, THE VALIDATION, AND THE COMPLETE DATA SETS FOR EACH CLONE. BCF IS FROM SNOWDON (1991). ....	17

## LIST OF APPENDICES

APPENDIX TABLE 1. BASIC STATISTICS OF STEM DIAMETER (MM), MEASURED AT 30 CM FROM THE BASE OF THE STOOL, FOR THE COMPLETE, MODEL-SELECTION, AND VALIDATION DATA SETS. ....	20
APPENDIX TABLE 2. BASIC STATISTICS OF OVEN DRY WEIGHT (G, 65°C) FOR THE COMPLETE, MODEL-SELECTION, AND VALIDATION DATA SETS. ....	21

## EXECUTIVE SUMMARY

Researchers and managers of commercial bioenergy plantations often require estimation of standing aboveground biomass prior to harvesting. At the time of harvest, biomass can be weighed, but obtaining estimates prior to that time requires the use of an alternate, non-destructive technique. At an operational level, willow plantations will be harvested in winter, chipped and transported to power facilities. Chips will have to be delivered on an as needed basis because of the limited amount of storage space at power plants. In order to ensure a steady supply of willow biomass to the power plants, estimates of biomass before the time of harvest will be necessary. Landowners will likely find estimation of biomass useful for decision-making. For example, rotations are typically three years, but could be extended to four or even five years if the amount of standing biomass does not make harvesting profitable. Because of the strong allometric relationship between stem diameter and dry stem biomass, regression estimation is a reasonable, non-destructive approach to estimating biomass.

The objective of this study was to develop regression equations that accurately predict aboveground biomass in a cost-effective manner. Equations were developed for one- and two-year-old coppice for willow clones S25, S301, S546, SA2 and SV1, and poplar clone NM6. Several regression models were considered for estimation of stem dry weight (biomass) for each clone individually and combined (i.e., one equation for all clones). Several explanatory variables were considered in the various regressions, including height (H), diameter (D),  $D^2$  and  $D^2H$ . Validation was conducted by randomly splitting the data. Transformational bias can be an issue when predictions are made using logarithmic regression. Two bias correction approaches were evaluated.

Regression equations based solely on diameter (D), or the natural log (D), fit the data well (e.g.,  $r^2$  values ranged from 0.98 to 0.99). Application of equations that use diameter as the only predictor variable will allow for cost-effective measurement and estimation of standing biomass. Adding height to the stem-based regression equations increased the cost of data collection and did not significantly improve the regression equations. Clone-specific models appeared to be necessary, although estimates of biomass could be made using equations developed for more than one clone or site, combined (i.e., generalized), depending on the intended use or required degree of accuracy. Neither bias correction approach consistently improved predictions for logarithmic regressions; therefore, neither was recommended for use in this instance. With the exception of the poplar clone, NM6, the non-linear and logarithmic regressions performed equally well, suggesting that the two models are both reasonable; logarithmic estimators had a slight predictive advantage overall. For NM6, the logarithmic regression performed substantially better.

The biomass estimation equations developed to date are being used in various research trials. Additional data collection will take place in the future to expand the range of ages and diameters. Further verification of the equations will be conducted when three-year-old coppice willow is harvested. This will allow the application of these equations to commercial size trials in central and western New York.

## INTRODUCTION

### Background

The use of short-rotation woody crop (SRWC) plantations for bioenergy production has become increasingly attractive as alternative sources of renewable energy are being explored. Environmental concerns over acidic deposition and global warming, in relation to the burning of fossil fuels, primarily coal, in industrial and utility plants, have sparked an interest in clean sources of renewable energy (e.g.,  $\text{CO}_2$ -neutral and low  $\text{SO}_x$  and  $\text{NO}_x$  emissions). Fast-growing plantations of hardwoods, such as willow or poplar, have the potential to meet some of these needs in the northeastern United States. Co-firing willow biomass with coal results in consistently reduced  $\text{SO}_x$  emissions, and reduced  $\text{NO}_x$  emissions under certain conditions. Both of these compounds are precursors of acid rain. The amount of  $\text{CO}_2$  released in the burning of willow biomass is equivalent to the amount recaptured by growing willow trees, and therefore, can be considered a  $\text{CO}_2$ -neutral fuel.

Led by Niagara Mohawk Power Corporation, the Salix Consortium, formed in 1994, is an association of more than 20 corporations, industrial and government agencies, and farming and research organizations. Current research at the State University of New York, College of Environmental Science and Forestry (SUNY-ESF), in a

coordinated effort with other Salix Consortium members, is aimed at establishing operational bioenergy production on a regional level to supply wood for co-firing with coal at power plants in the region.

### **The Willow Bioenergy System**

Plantations of trees (shrubs) are planted in a “double-row system” based on a Swedish design. Trees are generally harvested on a three to four year rotation. Cuttings (25 cm-long sticks) are planted at a density of approximately 15,000 stools per hectare (60 x 70 x 150 cm spacing). Cuttings sprout soon after planting and are coppiced (cut back) in the first winter. The following spring, stumps re-sprout, producing stools with multiple stems (3 to 15+ stems per stool). Further details of the production system can be found in Kopp et al. (1997).

### **Application to Commercialization**

At an operational level, willow plantations will be harvested in winter, chipped and transported to power facilities. Chips will have to be delivered on an as needed basis because of the limited amount of storage space at power plants. In order to ensure a steady supply of willow biomass to the power plants, estimates of biomass yield before the time of harvest will be necessary. Landowners will likely find estimation of biomass useful for decision-making. For example, rotations are typically three years, but could be extended to four or even five years if the amount of standing biomass does not make harvesting profitable. Additionally, current research often requires estimation of aboveground biomass prior to harvesting. At the time of harvest, biomass can simply be weighed, but obtaining estimates prior to that time requires the use of an alternate, non-destructive technique. Because of the strong allometric relationship between stem diameter and dry stem biomass and the relative ease of data collection, regression estimation offers a reasonable, non-destructive approach to estimating stem biomass.

### **Objectives**

- 1) To develop regression equations to predict biomass in a cost-effective manner using stem samples taken from existing plantings in New York State for six different clones (five willow clones and one poplar clone; summarized in

Table 1).

- 2) To establish an efficient approach for collecting measurements necessary to use the biomass estimators.

#### **Present Status**

Objective 1: One- and two-year-old stems were collected and regression equations have been developed.

Objective 2: A reasonable approach has been developed for measuring individual stems. Research plots can readily be measured using a census of trees.

#### **Future Progress**

Objective 1: Three-year-old stems will be collected and used to expand the range of valid ages and diameters for which regression estimators can be applied. Additional validation of the regression estimators will be conducted by collecting plot biomass data at harvest (winter, 1998-99). Future measurement should include a selection procedure for collecting stems to be used in biomass equation development. In addition to the regression approaches considered, a weighted regression should be evaluated.

Objective 2: A sampling scheme needs to be developed for use of regression estimators to assess standing biomass in larger (demonstration and commercial) plantings.

**Table 1.** Willow and poplar clones and their parentage and origin in various plantings in New York State.

Clone	Parentage, Origin
NM6	<i>Populus nigra</i> x <i>P. maximowiczii</i> . Ontario Ministry of Natural Resources.
S25	<i>S. eriocephala</i> 16 x <i>S. eriocephala</i> 276. University of Toronto.
S301	<i>S. exigua</i> 62 x <i>S. eriocephala</i> 276. University of Toronto.
S546	<i>S. eriocephala</i> 16 x <i>S. eriocephala</i> 24. University of Toronto.
SA2	<i>S. alba</i> var. <i>sanguinea</i> . Novi Sad, Yugoslavia.
SV1	<i>Salix dasyclados</i> . Brantford, Ontario, Canada

### Questions to be Addressed

Considerable research in the area of biomass estimation techniques for short-rotation woody crops (SRWC), *Salix* in particular, has been conducted in Scandanavia (Telenius, 1997; Verwijst, 1991; Verwijst and Nordh, 1992; Telenius and Verwijst, 1995; Hytönen et al., 1987; Tahvanainen, 1996; Verwijst and Telenius, 1998). Many of these researchers have suggested that it is both necessary (cost-effective) and adequate to develop and utilize regression estimators that predict stem weight based solely on stem diameter. Using data collected in clone-site trials in New York State, this approach was evaluated by comparing regression estimators that included height as a predictor variable.

Q: Does height add to the predictive capability of the estimators, and if so, is it cost effective?

Biomass estimation is often conducted on a stem basis; however, the possibility of estimating biomass on a stool basis was considered in this study.

Q: Can stool-based regression estimators be developed that are comparable to estimates obtained from stem-based models?

Species, clone, age of the trees, stand (and sites), and degree of competition may all impact the allometric relationship between diameter and biomass (Telenius and Verwijst, 1995). Additionally, cultural treatments likely influence the allometric relations. Using a very large database, Telenius and Verwijst (1995) found that allometric equations were highly age-specific and species-specific. However, when equations were developed for more than one clone or site combined (i.e., generalized), estimates generally deviated by less than 10% from estimates obtained from the stand-specific equations—a potentially reasonable loss of accuracy in commercial application. Site-specific equations for every age and clone are costly. It would be beneficial to develop generalized equations.

Q: Can a single equation be used for more than one clone, or are individual equations necessary for each clone? What impact does age-generalization have on predictive ability?

Log transformed variables are frequently used in tree biomass equations. The logarithmic regression model is:  $\ln W = \alpha + \beta \ln D + \ln(\epsilon)$ ,  $\epsilon \sim N(0, \sigma^2)$ , where  $W$  is dry stem weight, and  $D$  is diameter, for example. Whittaker and Marks (1975) have shown systematic errors in estimates of  $W$  resulting from logarithmic transformations of this sort. Baskerville (1972) and others (e.g., Beauchamp and Olson, 1973 and Snowdon, 1991) have proposed various corrections for the transformational bias (negative) when predicted values on the log scale are transformed back to the original units (e.g., Baskerville's (1972) multiplicative correction:  $\exp(S_e^2/2)$ , where  $S_e^2$  is the error mean square from the logarithmic regression). However, Snowdon (1991) indicated that these correction terms themselves may be biased. Snowdon therefore recommended the use of the ratio estimation technique familiar in

sampling theory (see Cochran, 1977; estimate of true biomass of population = true biomass of sample/predicted biomass of sample x estimated biomass of population).

Q: Do either of the bias corrections help the logarithmic regression estimators in this situation? Can any general statements be made about situations when the correction should be used?

Due to these difficulties with the logarithmic regression approaches, Telenius (1997) and Telenius and Verwijst (1995) proposed using non-linear regression, since computer facilities allow fitting non-linear models as easily as linear models.

Q: How does the predictive ability of the non-linear regressions compare to the logarithmic regressions?

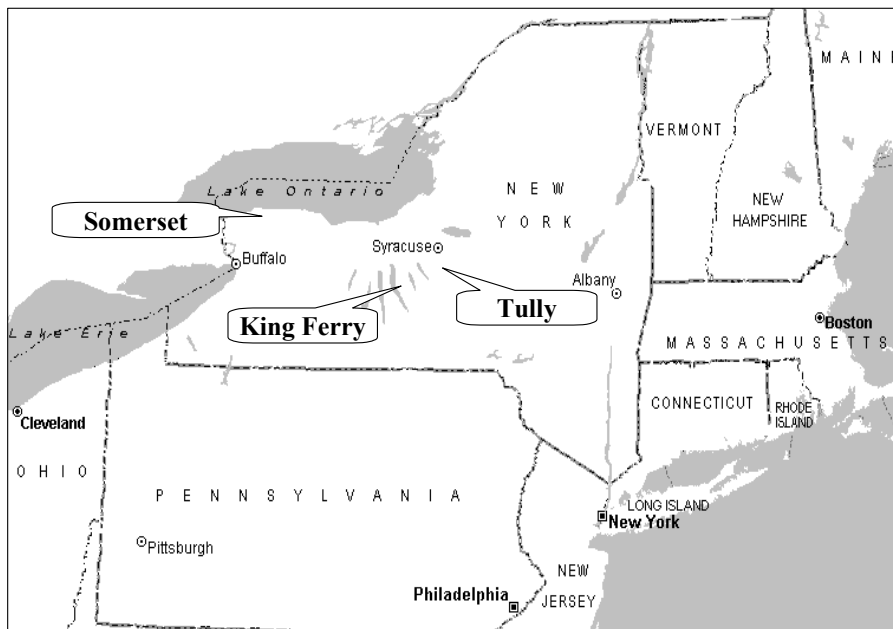
In research plots a census of stems is usually operationally feasible; however, in larger demonstration and commercial plantings of bioenergy crops, the estimation of biomass using regression estimators will require sampling a portion of the trees.

Q: How many stems and/or stools must be measured to adequately assess standing biomass using the regression estimators in a given plantation? How should these be sampled?

## METHODS

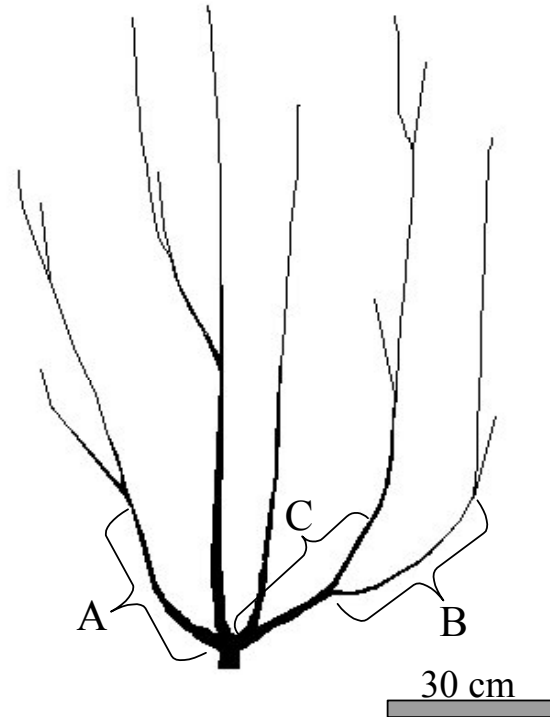
### Collection of Biomass Equation Data

Stem samples for biomass equation development were collected in late winter 1996 using stratified random sampling. Three stools per clone were selected at each of two or three sites from unfertilized areas (Figure 1). All stools in the buffer area were assigned a number on a map frame. One large, medium, and small stool each was selected randomly from the available numbers in the map frame. The first stool selected was assigned to a size class based on relative size for that clone and site. The remaining two size classes were filled by randomly selecting stools until stools meeting the size criterion were selected. Additional stools were selected in 1997 at the Tully site (three stools per clone, via a stratified random sample) and the Somerset site (one medium stool) for expanding the range of diameters used in developing the biomass equations.



**Figure 1.** Location of the fertilization trial plantings in New York State: King Ferry, Somerset, and Tully. Base map from Microsoft<sup>®</sup> Automap Streets Plus.

Six to nine stools per clone or 39 total stools were collected. The number of stems collected for a given clone ranged from 57 to 131 (Table 5; note: several outliers were removed—e.g., rabbit damage, or stems that had been classified and noted as “unusual” in the field). Stem diameter at 30 cm from the base of the stool (Figure 2), average stem height by stool (first year only), and number of stems per stool were measured in the field. Stems were then severed 5 cm above the ground. Corresponding stem oven-dry weights (dried at 65°C to a constant weight) were measured in the laboratory for use in establishing the allometric relationships between biomass and diameter, average stool height, and/or number of stems per stool.



**Figure 2.** Measurement of stem diameter was made at 30 cm from the base of the stool following the curvature of the stem (A). Secondary branches forking below 30 cm were measured from the base of the fork (B), while the main stem of a fork was measured from the base of the stool (C).

### Data Management

The complete data set was used for final model equations, whereas a random subset of the data was used to decide which model was best. Data splitting was used for reserving an independent sample from the complete stem data set to be used for model validation. The complete data set was split by randomly selecting half of the data by clone for equation selection and validation. One-half of the complete data set was referred to as the *validation* set and the other half was referred to as the *model-selection* set (see Appendix Table 1 and 2 for split data set statistics).

### Model Development and Selection

Model development and selection were conducted using the *model-selection* set. Initially, several regression estimators were considered. Regressions were considered for two generalized equations (all clones, including NM6, and all willow clones) and for clone-specific equations. Additionally, stem-based and stool-based (SV1 was used for illustration) regressions were considered. Linear regressions using untransformed (raw) data were evaluated for stem and stool-based data, primarily for illustration and completeness, since the relationship between stem diameter and biomass is not linear. Logarithmic regressions, regressions with  $D^2$  as an explanatory variable, and non-linear regressions (stem-based only) were run for individual clones and for the two generalized cases for stems and stools.

Model development and selection employed typical regression evaluation criteria, including  $R^2$ , adjusted  $R^2$ , Mean Square Error (MSE), and Prediction Sum of Squares (PRESS). Multiple regression was performed using stepwise and  $R^2$  selection procedures for predictor variables in Statistica (StatSoft, 1997). The Quasi-Newton non-linear estimation procedure in Statistica (StatSoft, 1997) was used to fit the non-linear models considered. Assumptions were tested using residual analysis techniques, including normal probability plots and residual plots. Tests for outliers were conducted using standardized residuals and deleted residuals.

### Validation

The *validation data* set was used to evaluate the predictive capacity of the various regression estimators. The Mean Square Error of the Prediction ( $MSEP = \Sigma(y_i - \hat{y}_i)^2 / m$ , where  $m$  is the number of observations used in the validation set, and the  $y_i$ 's are observed and predicted biomass for observation  $i$ ) was calculated based on the prediction equations applied to the independent *validation data* set. The MSEP measures the predictive ability of the models and serves as a basis for comparing model predictions. Neter et al. (1996) suggest that the MSEP is a good evaluation criterion for the predictive ability of regression equations. The MSEP was calculated for the data used to develop the prediction equations for reference.

Clone-specific and generalized stem-based equations were compared, as were the “best” non-linear and logarithmic regressions. Comparisons of “biased” and “unbiased” logarithmic regression predictions for the clone-specific, stem-based estimators were made using the validation data set(s). Two methods were examined for bias correction:

- a) Baskerville's (1972) multiplicative correction:  $\exp(S_e^2/2)$ ,  
where  $S_e^2$  is the error mean square from the logarithmic regression;
- b) Snowdon's (1991) ratio estimator bias correction factor:

$$BCF = \frac{\text{mean of sample biomass}}{\text{mean of predicted sample biomass}} = \frac{\bar{w}}{\hat{w}}$$

The biomass estimates were corrected by a) multiplying the estimates by the correction factor  $\exp(S_e^2/2)$ , and b) multiplying estimates by the BCF.

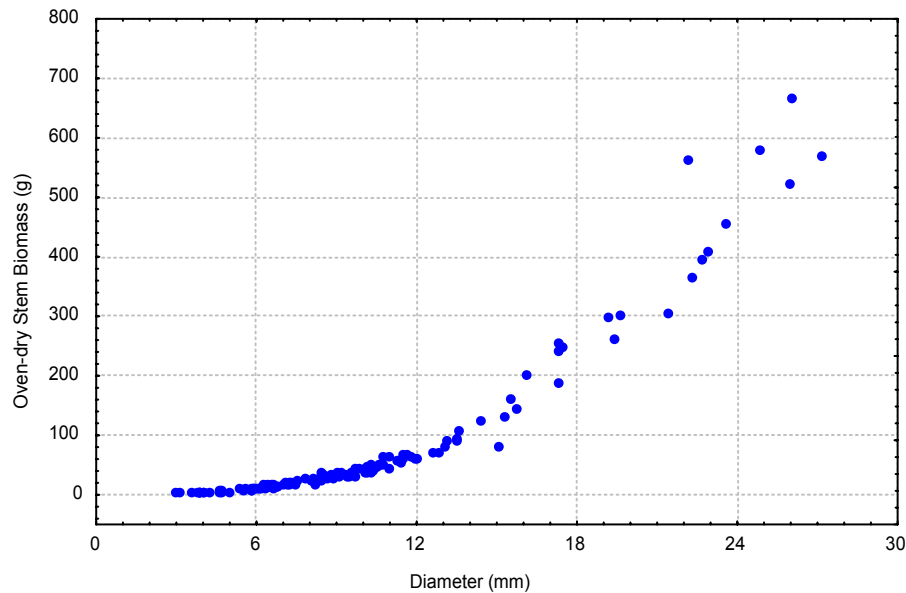
After the initial validation of the “best” stem-based regression estimators, the complete (original) data set was used to re-estimate the regression coefficients. These final versions of the regression equations were compared to the model-selection versions, by clone, for the purpose of “final” validation and creating the useable biomass equation.

NOTE: Stool based models were not validated because there were not enough observations to create a validation data set.

## RESULTS AND DISCUSSION

There was a strong allometric relationship between tree diameter and dry aboveground, leafless biomass (Figure 3). This relationship allows equations to be developed for predicting biomass of willow bioenergy crops.

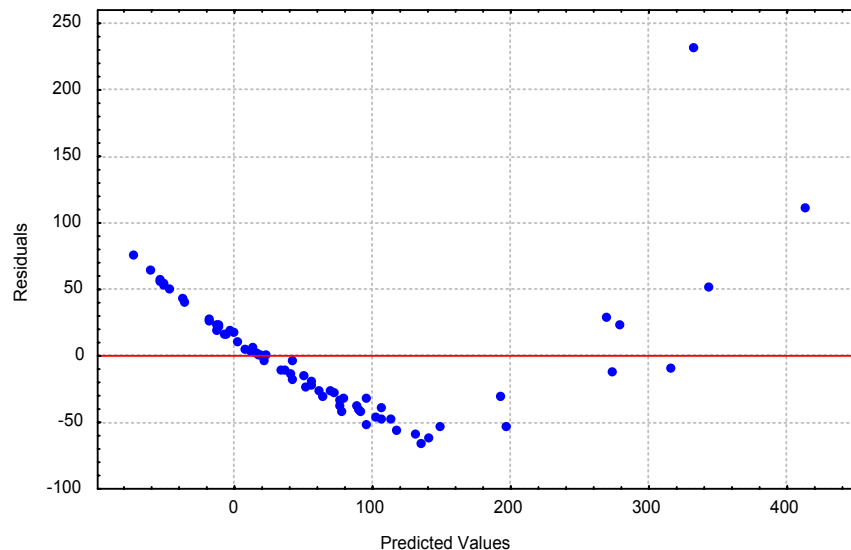
Many researchers have studied these relationships for the purpose of estimating tree biomass (e.g., Telenius and Verwijst, 1995, Telenius, 1997, Tahvanainen, 1996, Baskerville, 1972, Hytönen et al., 1987) and have proposed several reasonable models. Most of the “best” models are exponential (non-linear form:  $W = aD^b\epsilon$ ), linearized versions of the exponential model (log-transformed variables:  $\ln W = \ln a + b \cdot \ln D + \ln(\epsilon)$ ), and other modifications thereof (Tahvanainen, 1996).



**Figure 3.** Oven-dry stem biomass (g) versus stem diameter (mm) measured at 30 cm from the base of the stool for samples of clone SV1 collected in research plantings in New York State in winter, 1996 and 1997.

#### Simple Linear Regression Model Using Raw Data

The untransformed variables were first considered in a simple linear regression model (e.g.,  $W = a + bD$ ). It is rather obvious, and not surprising, that this model is not appropriate, as the relationship depicted in Figure 3 is not linear. Though the regressions had relatively high  $R^2$  values (e.g.,  $R^2 = 0.85$  to  $0.96$ , for stem and stool-based models, respectively), the residual plot in Figure 4 clearly shows the inappropriateness of the linear model with raw data. The residual plots clearly indicated that the model is not correct and that transformation of the data, or a model with a higher-order predictor variable, is necessary.

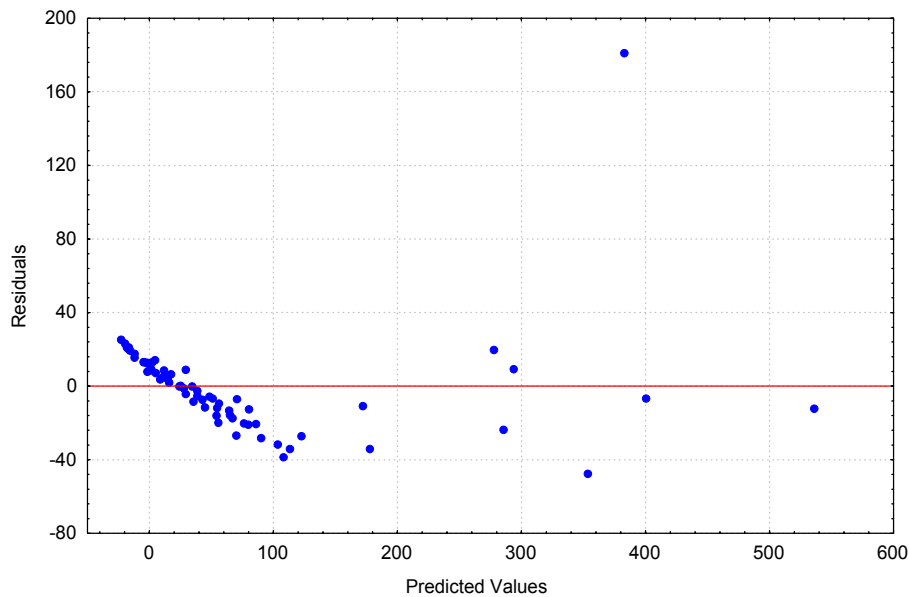


**Figure 4.** Residual plot (residuals vs. predicted values) of the non-transformed regression for stems of clone SV1.

#### Linear Regression Models Using $D^2$ and H as Predictor Variables

The second type of linear regressions considered were those that used diameter-squared ( $D^2$ ) as a predictor variable. Predictor variables considered in the stem-based regressions were diameter (D), height (H),  $D^2$ , and  $D^2 \cdot$

height ( $D^2H$ ). Using a stepwise selection procedure in Statistica (StatSoft, 1997),  $D$ ,  $D^2$ , and  $D^2H$  were selected and significant in the model. Again, the regressions have relatively high  $R^2$  (e.g., 0.96 for stem-based models). However, this data failed to meet the homogenous error variance assumption (i.e., error variance increased with increasing values of  $D$ , see Figure 5). These observations were valid for the generalized and clone-specific stem-based regressions, as well as for the stool-based regressions.



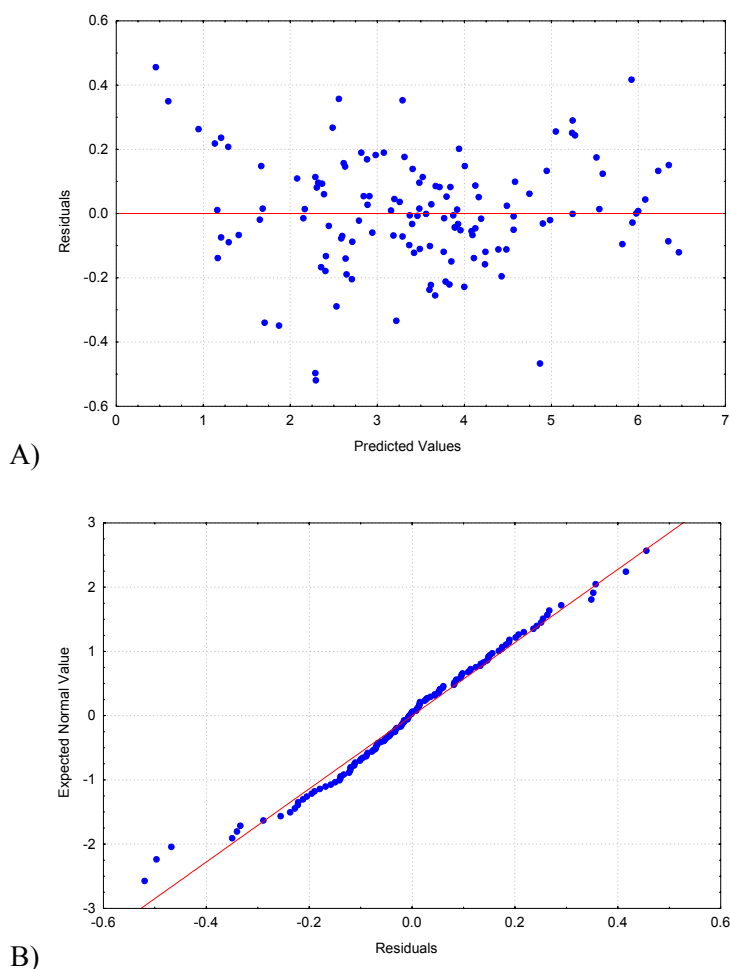
**Figure 5.** Residual plot (residuals vs. predicted) of the regression using  $D^2$  as the predictor variable for stems of all clones.

### Logarithmic Regressions

The logarithmic regression ( $\ln W = \ln a + b \cdot \ln D + \ln(\epsilon)$ ) appears to remedy the issues associated with the previous linear regressions (i.e., models with  $D^2$  predictors and the untransformed models with regard to the homoscedacity assumption and the use of an “incorrect model”). The stem-based logarithmic regression models had a high  $R^2$  and fit the data well in the generalized case as indicated by residual analysis (Figure 6). The results of the stem-based model selection indicated that diameter alone resulted in a good fit. For clones NM6, S301, S546, and SV1 and the generalized (all clones) regression estimator, height was not significant in the model already containing diameter. For the other two clones, S25 and SA2, height contributed little to the explanatory power of the regression (i.e., <1%, or 0.01 addition to the  $R^2$  value). The residual analysis of the logarithmic regression models evaluated suggested that they were more appropriate than the previous models (e.g., error variance was homogeneous, and residuals were normally distributed, Figure 6 A & B).

The fact that height was not significant in the models agrees with findings by Tahvanainen (1996), though it should be noted that the height measurements used in the analysis were the average height of stems on each stool, not measurements of individual stems. Taking individual stem heights may improve the regression. The cost of obtaining height measurements for individual stems may be prohibitive, though cost was not evaluated in these analyses.

The stool-based logarithmic regression models had a high  $R^2$  (e.g., 0.98) and fit the data well in the generalized case as indicated by residual analysis. The residual analysis was favorable, and resembled that of the stem-based models. The number of stems per stool ( $\ln(\text{number of stems})$ ) was not significant in the model ( $p = 0.15$ ), but  $\ln(\text{stool sum of diameters})$  and  $\ln(\text{average stool height})$  were both significant in the model.



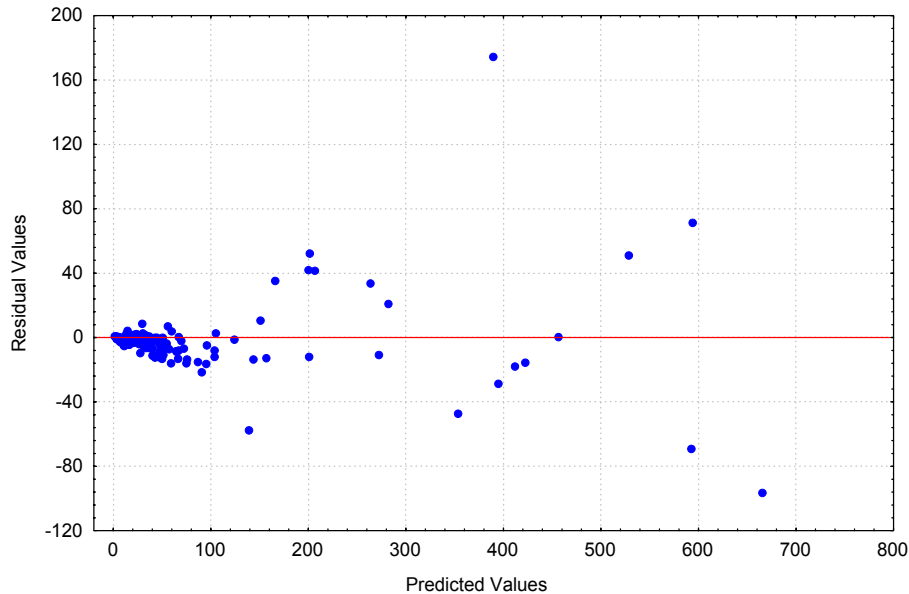
**Figure 6.** Residual plots for the stem-based logarithmic regression for clone SV1: (A) residual versus predicted values and (B) normal probability plot.

The disadvantage of the stool-based approach was that in order to predict biomass, stool height and diameter of all stems on a stool had to be measured (to obtain the sum of stool diameters). One difficulty in this analysis was that there was not a straightforward method for validating the stool-based model due to the small number of stool observations (e.g., cross-validation or data splitting was not a viable option). This was particularly true of the clone-specific, stool-based models. Clone SV1, with the largest number of stools (nine), had a high  $R^2$  value (0.98) and met the regression assumptions. Obviously, with so few observations, the SV1 stool data set could not be split for validation. The stool-based approach should not be completely abandoned. However, better prediction can be made using the stem-based approach. Since all stems on a stool would have to be measured in order to use the stool-based equation, it is more efficient to use the stem-based equations (e.g., development and modification of stem sampling approaches for biomass estimation increases the attractiveness of the stem-based approach).

As suggested in the discussion above, height was not a necessary variable for predicting biomass in stem-based equations. Height did not add to the predictive ability of the equations—at least not the stem-based versions, since it was not even significant in the model—but it could add to the predictive ability of the stool-based estimators.

## Non-linear Regression

The non-linear model that was considered ( $W = aD^b + c + \epsilon$ , recommended by Telenius, 1997) fit the data well. Compared with other possible functions this has been found to be both robust and simple (Verwijst, 1991). The residual analysis did suggest that error variance was non-homogeneous (i.e., increased as  $D$  increased; Figure 7). The pattern seen for clone SV1 in Figure 7 was representative of the other clones as well. The estimator was not discarded since it performed reasonably well in predictions (see the MSE in Table 2) and a weighted version of the nonlinear fit could be considered in the future.



**Figure 7.** Residual plot (residuals versus predicted) of the non-linear regression for stems of clone SV1.  $R^2$  for regression was 0.95.

Two versions of this non-linear model were considered:  $W = aD^b + c + \epsilon$  (with an intercept) and  $W = aD^b + \epsilon$  ( $c = 0$ ; regression passes through the origin). Generally, it is not desirable, nor recommended, to force the regression through the origin (Neter et al., 1996). However, the attractive aspects of such an equation is that there are no negative predictions of biomass for small diameter stems, as the intercept of willow biomass equations is often negative (Telenius, 1997). Only two clones in this analysis did not have negative intercepts: S546 and S25. The benefit is minor, however, because negative biomass predictions for small diameters can easily be set equal to zero, as they would likely contribute little to biomass anyway.

## Validation

The MSE in Table 2 was used as an indicator of predictive ability and served as a basis for the comparisons made between clone-specific versus generalized stem-based models, non-linear versus logarithmic regression estimators (Table 2), and bias correction methods (Table 3).

### Clone-specific versus Generalized Logarithmic Regression Estimators

As indicated in Table 2, clone-specific models were far better than the generalized equations for NM6 and SV1 based on the MSE (calculated from the validation data). The generalized equation underestimated biomass for clone SV1 and overestimated biomass for NM6, particularly as diameter increased (positive and negative residuals, respectively (Figure 8). The generalized logarithmic regression estimator (that included NM6) performed better than the clone-specific equations for clones S25 and S301, based on MSE. The generalized willow regression estimator performed well for SA2 and S546. For example, the generalized willow equation increased the accuracy of the predictions for SA2; residuals of the predictions were more evenly distributed around zero, though accuracy still decreased as diameter increased (Figure 9). In all cases, as would be expected, the accuracy of the predictions decreased as diameter increased (e.g., Figure 8 and Figure 9).

The clone-specific equations had the lowest MSEP when calculated using the model-selection data (with the exception of clone NM6 and S546). The fact that overall performance of the clone-specific versus generalized estimators was inconsistent may have been the result of the level of generalization in the model-selection data set. Specifically, the model-selection data sets comprised data from two or three sites and include one and two year old stems (with a greater number of observations for one year old stems). Telenius and Verwijst (1995) found that species, clone, site, and age all influence the relationship between stem weight and stem diameters. They suggest that clone-specific, site specific regression equations are necessary.

### **Age Generalization**

There was insufficient data to accurately assess the impact that age-generalization had on the predictive ability of the equations. Additional work should be done in this area.

### **Comparisons of Bias Correction Methods**

Neither of the two methods of bias correction, Baskerville's (1972)  $S_e^2/2$  correction nor Snowdon's (1991) ratio estimator correction, significantly improved the predictive power of the logarithmic regressions on a consistent basis (see Table 2). Plots of prediction errors when the model-selection equations were applied to the validation data set (Figure 0) clearly indicated that there was little, if any, improvement in accuracy of the estimates by using either of the correction approaches.

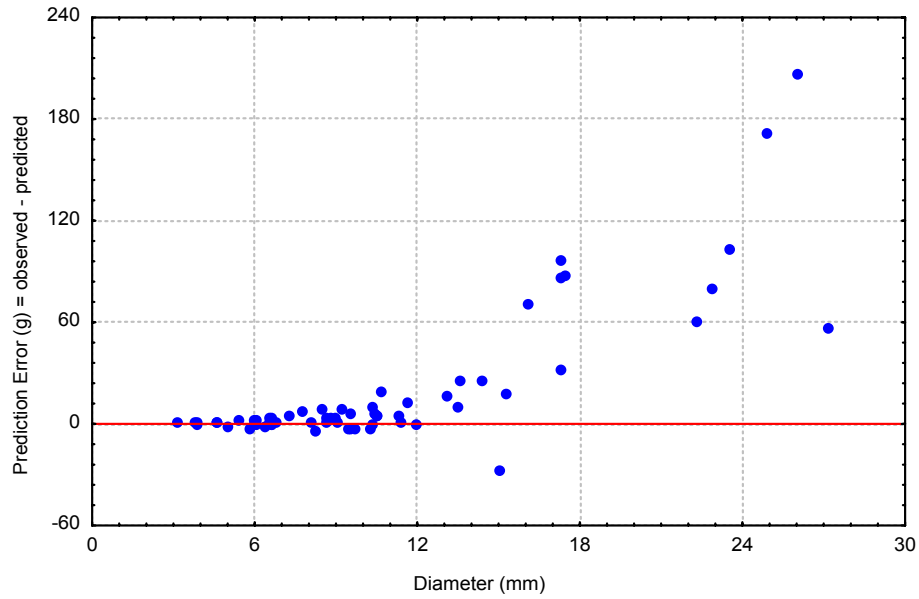
The intrinsic bias in the back-transformation of predictions from logarithmic regression should be corrected, in theory, but the impact on predictions made with these logarithmic regression estimators of willow biomass is relatively small in comparison to other errors of prediction (e.g., MSEP is much greater than the MSE of the regressions). For example, the accuracy of the predictions for clone SA2 was improved more by using the generalized equation than it was by either bias correction method. Prediction errors remained essentially unchanged by using the bias correction methods (Figure), but they were more evenly distributed around zero when using the generalized equation (Figure 9).

### **Non-linear versus Logarithmic Regressions**

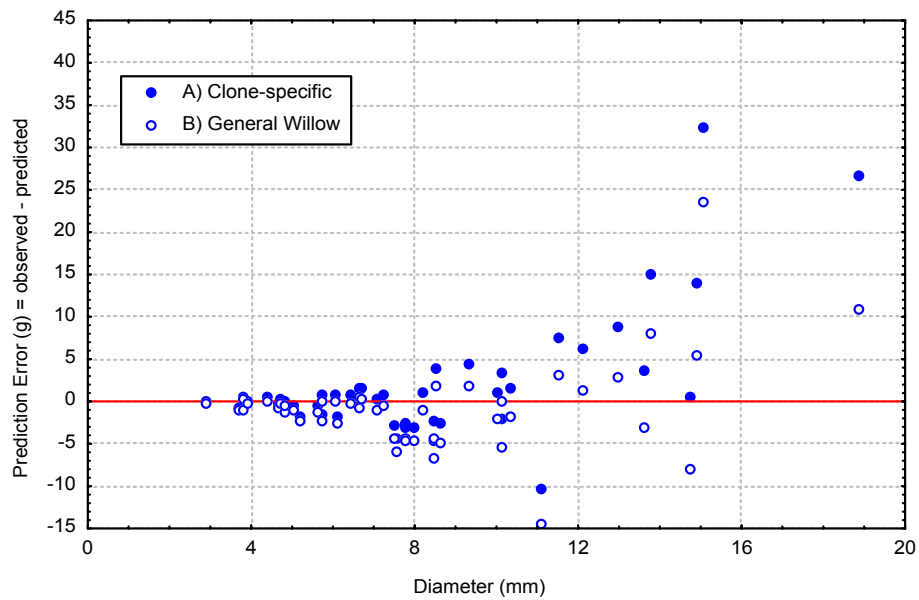
With the exception of poplar clone NM6, the non-linear and logarithmic regressions performed equally well (see Table 2), suggesting that both models are reasonable. Overall, the MSEP was slightly lower for the logarithmic regressions. The non-linear model had increasing error variance as D increased, as indicated in the previous discussion (see Figure 7), but could be corrected using a weighting procedure. For NM6, the logarithmic regression performed better much than the non-linear model (MSEP of 607 versus 1343, respectively).

**Table 2.** The mean square error of the prediction (MSEP) was calculated for the following stem-based models using verification data: (1) clone-specific logarithmic regressions, (2) generalized (all clones, including poplar) and (3) willow only logarithmic regressions: (4) non-linear clone-specific models:  $W = aD^b$  and (5)  $W = aD^b + c$ . The MSEP was also calculated using the model-selection data set (on the model-selection equations) and is reported for reference.

<b>Mean Square Error of the Prediction (MSEP) calculated using:</b>	<b>Validation Data Set on Model Selection Equations</b>	<b>Model Selection Data Set on Model Selection Equations</b>
<b>CLONE / REGRESSION</b>	(number of obs. in parentheses)	(number of obs. in parentheses)
(1) NM6/NM6 Logarithmic Regression	607	998
(2) NM6/General Logarithmic Regression	1744	842
(3) NM6/Willow Logarithmic Regression	3253	1527
(4) NM6 / Non-linear: $W=aD^b$	1359	573
(5) NM6 / Non-linear: $W=aD^b+c$	1342	573
	(n = 33)	(n = 24)
(1) S25 / S25 Logarithmic Regression	485	146
(2) S25 / General Logarithmic Regression	458	172
(3) S25 / Willow Logarithmic Regression	569	191
(4) S25 / Non-linear: $W=aD^b$	602	138
(5) S25 / Non-linear: $W=aD^b+c$	668	132
	(n = 39)	(n = 34)
(1) S301 / S301 Logarithmic Regression	146	117
(2) S301 /General Logarithmic Regression	135	126
(3) S301 /Willow Logarithmic Regression	157	131
(4) S301 / Non-linear: $W=aD^b$	169	112
(5) S301 / Non-linear: $W=aD^b+c$	171	111
	(n = 33)	(n = 27)
(1) S546 / S546 Logarithmic Regression	74	72
(2) S546 /General Logarithmic Regression	68	72
(3) S546 /Willow Logarithmic Regression	63	73
(4) S546 / Non-linear: $W=aD^b$	81	66
(5) S546 / Non-linear: $W=aD^b+c$	89	63
	(n = 51)	(n = 50)
(1) SA2 / SA2 Logarithmic Regression	54	12
(2) SA2 / General Logarithmic Regression	34	19
(3) SA2 / Willow Logarithmic Regression	28	25
(4) SA2 / Non-linear: $W=aD^b$	61	12
(5) SA2 / Non-linear: $W=aD^b+c$	63	11
	(n = 49)	(n = 45)
(1) SV1 / SV1 Logarithmic Regression	570	670
(2) SV1 / General Logarithmic Regression	2065	1496
(3) SV1 / Willow Logarithmic Regression	1721	1312
(4) SV1 / Non-linear: $W=aD^b$	493	619
(5) SV1 / Non-linear: $W=aD^b+c$	478	609
	(n = 63)	(n = 68)



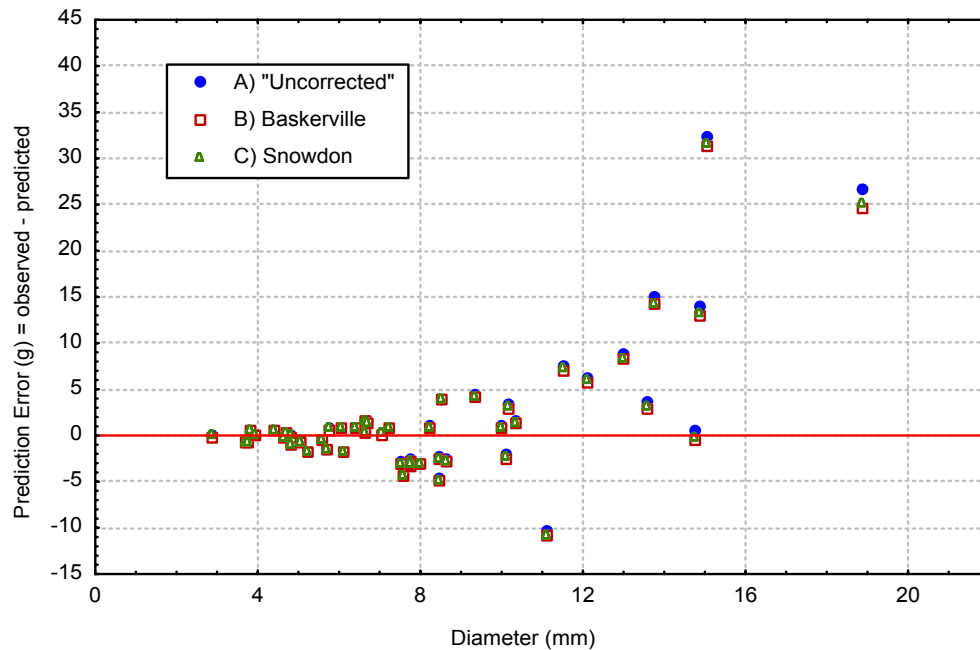
**Figure 8.** Prediction errors of the generalized logarithmic equation for the SV1 validation data set. (Note: prediction error = observed – predicted).



**Figure 9.** Prediction errors of the logarithmic regression for the SA2 validation data set: A) clone-specific (SA2) prediction equation, MSE<sub>EP</sub> = 54.0 and B) generalized (willow) prediction equation, MSE<sub>EP</sub> = 33.8.

**Table 3.** The mean square error of the prediction (MSEP) was calculated for the clone-specific, stem-based logarithmic regression models using verification data: (a) “uncorrected” logarithmic regression estimates, (b) bias corrected regression estimates using Baskerville’s multiplicative correction ( $\exp(S_e^2/2)$ ) and (c) Snowdon’s ratio estimator correction; The MSEP was also calculated using the model-selection data set (on the model-selection equations) and is reported for reference.

CLONE	Validation Data Set on Model Selection Equations (number of obs. in parentheses)			Model Selection Data Set on Model Selection Equations (number of obs. in parentheses)		
	Regression Estimate (a)	Estimate with Bias Correction Applied		Regression Estimate (a)	Estimate with Bias Correction Applied	
		Snowdon (b)	Baskerville (c)		Snowdon (b)	Baskerville (c)
NM6	607 (n = 33)	592	535	998 (n = 24)	608	814
S25	485 (n = 39)	501	513	146 (n = 39)	144	143
S301	146 (n = 33)	156	159	117 (n = 27)	118	119
S546	74 (n = 51)	70	69	72 (n = 51)	71	71
SA2	54 (n = 49)	50	49	12 (n = 45)	13	13
SV1	570 (n = 63)	499	537	670 (n = 68)	624	649



**Figure 10.** Plot of prediction errors of the clone-specific equations for the SA2 validation data set: A) Logarithmic regression estimate; B) Logarithmic regression estimate with Baskerville’s (1972) bias correction; C) Logarithmic regression estimate with Snowdon’s (1991) bias correction.

### Additional Validation

In addition to the predictions made with the validation data (i.e., MSEP evaluation), the statistics for the logarithmic regressions using the model-selection and the complete data sets were compared to validate the

regression estimators. The comparison indicated that by adding the verification set back into the model-selection set(s), the coefficients of the regression equations did not change greatly; the greatest change was for the coefficients of clone SA2 (Table 4). The changes in regression coefficients for SA2 resulted from a few stems with large dry weights randomly reserved in the validation data set (see Table 5). Another method for future validation of the biomass estimators would be to compare harvested biomass of research plots with predictions made from the regression estimators. This approach will be applied following harvests in the winter of 1998/99.

#### Equation Coefficients for the Logarithmic Regressions

The final biomass equation coefficients for the logarithmic regression equations are summarized in Table 5. The general form of the final logarithmic regression equation is:

$$\text{Dry stem biomass (W)} = [e^{a + b \ln(\text{Stem Diameter})}]$$

$$\text{Example: General (all clones) } W = [e^{-2.536 + 2.666 \ln(\text{Stem Diameter})}]$$

**Table 4.** Statistics of the generalized logarithmic regressions using the model-selection, the validation, and the complete data sets. BCF is from Snowdon (1991).

CLONE	ALL (GENERAL)			CLONE	ALL WILLOW		
	Statistic	Model-Selection	Validation		Complete	Statistic	Model-Selection
n	248	268	516	n	224	235	459
a	-2.455	-2.610	<b>-2.536</b>	a	-2.485	-2.653	-2.571
s{a}	0.0537	0.0506	0.0369	s{a}	0.0568	0.0544	0.0393
b	2.634	2.696	<b>2.666</b>	b	2.654	2.722	2.689
s{b}	0.0239	0.0222	0.0163	s{b}	0.0257	0.0241	0.0176
MSE ( $S_{\square}^2$ )	0.036	0.035	0.035	MSE ( $S_{\square}^2$ )	0.034	0.034	0.034
R-square	0.9823	0.9802	0.9812	R-square	0.9821	0.9796	0.9808
BCF (ratio)	1.030	1.007	1.019	BCF (ratio)	1.040	1.031	1.036

**Table 5.** Statistics of the clone-specific logarithmic regressions using the model-selection, the validation, and the complete data sets for each clone. BCF is from Snowdon (1991).

CLONE	NM6			CLONE	S546		
Statistic	Model-Selection	Validation	Complete	Statistic	Model-Selection	Validation	Complete
n	24	33	57	n	50	51	101
a	-2.605	-2.547	-2.570	a	-2.517	-2.522	-2.523
s{a}	0.1665	0.1127	0.0929	s{a}	0.0806	0.0842	0.0579
b	2.642	2.620	2.629	b	2.652	2.669	2.663
s{b}	0.0646	0.0461	0.0371	s{b}	0.0375	0.0380	0.0265
MSE ( $S_{\square}^2$ )	0.036	0.025	0.029	MSE ( $S_{\square}^2$ )	0.019	0.017	0.018
R-square	0.9905	0.9870	0.9891	R-square	0.9902	0.9905	0.9903
BCF (ratio)	1.049	1.041	1.043	BCF (ratio)	1.008	1.011	1.009

CLONE	S25			CLONE	SA2		
Statistic	Model-Selection	Validation	Complete	Statistic	Model-Selection	Validation	Complete
n	34	39	73	n	45	49	94
a	-2.853	-2.844	-2.850	a	-2.590	-2.924	-2.807
s{a}	0.1766	0.1432	0.1102	s{a}	0.1343	0.1035	0.0825
b	2.756	2.771	2.765	b	2.663	2.813	2.761
s{b}	0.0733	0.0593	0.0457	s{b}	0.0643	0.0504	0.0399
MSE ( $S_{\square}^2$ )	0.027	0.030	0.029	MSE ( $S_{\square}^2$ )	0.021	0.023	0.022
R-square	0.9833	0.9779	0.9810	R-square	0.9851	0.9755	0.9812
BCF (ratio)	1.008	0.971	0.988	BCF (ratio)	1.008	1.025	1.014

CLONE	S301			CLONE	SV1		
Statistic	Model-Selection	Validation	Complete	Statistic	Model-Selection	Validation	Complete
n	27	33	60	n	68	63	131
a	-2.269	-2.325	-2.288	a	-2.442	-2.570	-2.499
s{a}	0.1238	0.1162	0.0860	s{a}	0.0918	0.1062	0.0694
b	2.574	2.561	2.561	b	2.695	2.740	2.716
s{b}	0.0550	0.0499	0.0375	s{b}	0.0415	0.0461	0.0308
MSE ( $S_{\square}^2$ )	0.030	0.023	0.027	MSE ( $S_{\square}^2$ )	0.027	0.034	0.030
R-square	0.9884	0.9887	0.9877	R-square	0.9830	0.9846	0.9837
BCF (ratio)	1.013	1.036	1.029	BCF (ratio)	1.040	1.039	1.041

## CONCLUSION

After evaluating the various regression estimators for the prediction of biomass for one poplar and five willow bioenergy crops it was determined that regression equations based solely on diameter could adequately estimate tree biomass. Height was not a significant predictor variable in the stem-based regression models. Stool-based estimators were an exception; height was a significant predictor variable in the regression. The disadvantage of including height in any regression estimation procedure is the added cost of measuring the variable, particularly on a commercial basis.

The disadvantages of using stool-based regressions include the cost of measuring height and the requirement that all stems on a stool be measured (to obtain the sum of stool diameters) in order to predict biomass. In general, stem-based regression estimators are easier to use and provide reasonable biomass predictions. Since all stems on a stool have to be measured in order to use the stool-based equation, it is more efficient to use the stem-based equations (without the need for height measurement). There is the possibility of developing or modifying stem sampling approaches to improve the efficiency of collecting data for biomass estimation.

The generalized (all-clone) logarithmic regression estimator performed well for several clones. However, clone-specific models were far better than the generalized equations for NM6 and SV1. Overall, the improvements in the predictive ability (MSEP) for a given clone using the generalized equation were relatively small. Clone-specific equations should be used whenever possible; though, generalized equations may yield reasonable approximations. For example, when regression equations were generalized for clone or site, estimates generally deviated by less than 10% from the stand-specific equations (Telenius and Verwijst, 1995).

Of the two correction methods evaluated for the intrinsic negative bias associated with log-transformed variables in tree biomass equations, neither the multiplicative error term ( $\exp(S_e/2)$ ) proposed by Baskerville (1972), nor the multiplicative ratio estimator advocated by Snowdon (1991), consistently decreased MSEP. Neither method consistently improved the accuracy of the predictions. The impact the bias corrections had on predictions made with the logarithmic regression equations was relatively small in comparison to other errors of prediction (e.g.,  $MSEP \gg MSE$ ). The use of bias corrections should be considered in more detail if the logarithmic regression is used in practice.

Overall, the results suggested that logarithmic and non-linear regressions were both reasonable. The performances of the two approaches were comparable (MSEP was only slightly better for the logarithmic regressions), with the exception of poplar clone NM6. For NM6, predictive ability of the clone-specific logarithmic regression was superior to that of the non-linear approach. Error variance increased as diameter increased when using the non-linear regression model suggesting that a weighting procedure should be considered in the future. The logarithmic regression was selected as the final model.

There are many questions that have not yet been addressed, including, how many stems or stools must be measured to apply the regression equations to obtain a reasonable estimate of yield in a given plantation or plot and how should these be sampled. In addition, future work should include more intensive sampling of stems of various ages to develop age-specific estimators and quantify the loss of accuracy by using age-generalized estimators.

## LITERATURE CITED

- Baskerville, G. L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Can. J. For. Res.* (1): 49-53.
- Beauchamp, J. J. and Olson, J. S. 1973. Corrections for bias in regression estimates after logarithmic transformation. *Ecology* 54(6): 1403-1407.
- Cochran, W. G. 1977. *Sampling techniques*. 3rd ed. John Wiley & Sons, New York.
- Hytönen, J., Lumme, I., and Törmälä, T. 1987. Comparison of methods for estimating willow biomass. *Biomass*, 14:39-49.
- Kopp, R.F., Abrahamson, L.P., White, E.H., Volk, T.A., and Peterson, J.P. 1997. *Willow Bioenergy Producer's Handbook*. New York State Energy Research and Development Authority (NYSERDA). Albany, NY.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. 1996. *Applied Linear Regression Models*. Third Ed. Richard D. Irwin, Inc., Chicago.
- Snowdon, P. 1991. A ratio estimator for bias correction in logarithmic regressions. *Can. J. For. Res.* 21: 720–724.
- StatSoft, Inc. (1997). *STATISTICA for Windows* [Computer program manual]. Tulsa, OK: StatSoft, Inc., 2300 East 14th Street, Tulsa, OK 74104, phone: (918) 749-1119, fax: (918)749-2217, email: info@statsoft.com, WEB: http://www.statsoft.com
- Tahvanainen, L. 1996. Allometric relationships to estimate above-ground dry-mass and height in *Salix*. *Scand. J. For. Res.* 11:233-241.
- Telenius, B. 1997. A software tool for standardised non-destructive biomass estimation in short rotation forestry. *Bioresource Technology* v.60(3):267-268.
- Telenius, B. and Verwijst, T. 1995. The influence of allometric variation, vertical biomass distribution and sampling procedure on biomass estimates in commercial short rotation forests. *Bioresource Technology* v.51(2/3):247-253.
- Verwijst, T. 1991. Logarithmic transformations in biomass estimation procedures: Violation of the linearity assumption in regression analysis. *Biomass and Bioenergy* Vol. 1, No. 3:175-180.
- Verwijst, T. and Nordh, N.-E. 1992. Non-destructive estimation of biomass of *Salix dasyclados*. *Bioresource Technology* v.41: 59-63.
- Verwijst, T. and Telenius, B. 1999. Biomass estimation procedures in short rotation forestry. *Forest Ecology and Management* 121: 137-146.
- Whittaker, R.H and Marks, P. L. 1975. Methods of assessing terrestrial productivity. In: H. Lieth and R.H. Whittaker (Eds.). *Primary Productivity of the Biosphere*. Pp. 55-118. Springer-Verlag, New York.

## APPENDIX

**Appendix Table 1.** Basic statistics of stem diameter (mm), measured at 30 cm from the base of the stool, for the complete, model-selection, and validation data sets.

CLONE/Data Set	VARIABLE	Valid N	Stem Diameter (mm)			Std. Error
			Mean	Minimum	Maximum	
GENERAL (All)	Diameter	516	10.35	2.91	43.41	0.24
GENERAL (Model-selection)	Diameter	248	10.14	2.93	36.59	0.34
GENERAL (Validation)	Diameter	268	10.53	2.91	43.41	0.35
NM6 (All)	Diameter	57	13.60	3.81	43.41	1.16
NM6 (Model-selection)	Diameter	24	14.49	3.81	36.59	1.69
NM6 (Validation)	Diameter	33	12.95	3.93	43.41	1.58
S25 (All)	Diameter	73	11.72	3.05	25.12	0.57
S25 (Model-selection)	Diameter	34	11.59	4.85	24.61	0.77
S25 (Validation)	Diameter	39	11.83	3.05	25.12	0.83
S301 (All)	Diameter	60	10.68	3.02	23.30	0.69
S301 (Model-selection)	Diameter	27	10.35	3.02	22.67	1.10
S301 (Validation)	Diameter	33	10.94	3.17	23.30	0.90
S546 (All)	Diameter	101	9.44	2.93	19.99	0.45
S546 (Model-selection)	Diameter	50	9.16	2.93	19.89	0.66
S546 (Validation)	Diameter	51	9.71	2.95	19.99	0.62
SA2 (All)	Diameter	94	8.23	2.91	18.88	0.34
SA2 (Model-selection)	Diameter	45	8.32	4.12	17.81	0.45
SA2 (Validation)	Diameter	49	8.15	2.91	18.88	0.51
SV1 (All)	Diameter	131	10.23	2.97	27.14	0.47
SV1 (Model-selection)	Diameter	68	9.72	2.97	25.99	0.60
SV1 (Validation)	Diameter	63	10.77	3.13	27.14	0.73
WILLOW (All)	Diameter	459	9.94	2.91	27.14	0.23
WILLOW (Model-selection)	Diameter	224	9.68	2.93	25.99	0.31
WILLOW (Validation)	Diameter	235	10.19	2.91	27.14	0.33

**Appendix Table 2.** Basic statistics of oven dry weight (g, 65°C) for the complete, model-selection, and validation data sets.

CLONE/Data Set	VARIABLE	Valid N	Dry Weight (g)			Std. Error
			Mean	Minimum	Maximum	
GENERAL (All)	Dry Weight	516	71.00	1.18	1572.7	5.86
GENERAL (Model-selection)	Dry Weight	248	65.70	1.45	1084.8	7.28
GENERAL (Validation)	Dry Weight	268	75.90	1.18	1572.7	9.05
NM6 (All)	Dry Weight	57	152.76	2.42	1572.7	38.92
NM6 (Model-selection)	Dry Weight	24	157.82	2.42	1084.8	51.06
NM6 (Validation)	Dry Weight	33	149.07	3.04	1572.7	56.70
S25 (All)	Dry Weight	73	74.18	1.29	412	9.69
S25 (Model-selection)	Dry Weight	34	68.85	3.65	412	14.32
S25 (Validation)	Dry Weight	39	78.82	1.29	340.72	13.29
S301 (All)	Dry Weight	60	68.07	1.45	316.04	10.48
S301 (Model-selection)	Dry Weight	27	69.39	1.45	316.04	16.10
S301 (Validation)	Dry Weight	33	66.99	1.71	310.01	13.99
S546 (All)	Dry Weight	101	49.10	1.35	246.24	5.76
S546 (Model-selection)	Dry Weight	50	46.19	1.54	246.24	8.32
S546 (Validation)	Dry Weight	51	51.95	1.35	202.29	8.04
SA2 (All)	Dry Weight	94	29.37	1.18	213.91	3.75
SA2 (Model-selection)	Dry Weight	45	27.87	3.6	153.77	4.52
SA2 (Validation)	Dry Weight	49	30.75	1.18	213.91	5.92
SV1 (All)	Dry Weight	131	81.76	2.49	665.4	11.69
SV1 (Model-selection)	Dry Weight	68	69.55	2.49	564.13	13.99
SV1 (Validation)	Dry Weight	63	94.95	2.58	665.4	19.04
WILLOW (All)	Dry Weight	459	60.85	1.18	665.4	4.28
WILLOW (Model-selection)	Dry Weight	224	55.83	1.45	564.13	5.62
WILLOW (Validation)	Dry Weight	235	65.63	1.18	665.4	6.41