Comparison of regression methods for predicting singleleaf pinyon phytomass

Robin J. Tausch
University of Nevada, Reno

Paul T. Tueller
University of Nevada, Reno

Recommended Citation

COMPARISON OF REGRESSION METHODS FOR PREDICTING SINGLELEAF PINYON PHYTOMASS

Robin J. Tausch[1] and Paul T. Tueller[1]

ABSTRACT.—Succession, nutrient cycling, production, and competition studies in plant communities require estimation of plant biomass. This is often accomplished by relating weight to easily measured plant dimensions via allometric equations. Dimensions of basal area and crown volume were used to predict phytomass of singleleaf pinyon (*Pinus monophylla* Torr. & Frem.). Two regression methods for fitting the allometric equation to data for phytomass prediction were tested. These methods were the more commonly used logarithmic transformation of both data variables followed by linear regression analysis and an iterative nonlinear analysis without data transformation. The first was consistently less effective for predicting both individual tree and total plot phytomass for pinyon. Basal area was a better predictor of phytomass on a site than was crown volume. Prediction equations were highly site specific. Age of the trees in the stand used to derive the equations significantly affected the phytomass estimates. Other site and stand factors also appear to affect the phytomass-allometric parameter relationship.

Detailed studies of succession, nutrient cycling, production, and competition in plant communities often require estimates of plant biomass and production. Procedures have been developed for making these estimates by relating biomass or production to easily measured plant dimensions. Examples include the prediction of surface area and biomass relationships of woody plants in deciduous forests (Whittaker and Woodwell 1967, Pastor et al. 1983) and prediction of the aboveground biomass and wood volume of western conifers (Cochran 1982, Hatchell et al. 1985) and eucalyptus species (Stewart et al. 1979). Biomass predictions from plant dimensions have also been used for shrubs, grasses, and forbs (Tausch 1980, Hymphrey 1985, Alaback 1986, Hughes et al. 1987). The most commonly used plant dimensions are stem or trunk diameter, plant height, and crown volume.

Various types of regression analysis techniques are used for these procedures. Usually they involve some type of regression fit to the allometric equation: $Y = aX^b$, where Y = weight and X = the dimension measured (Sprent 1972). The allometric equation describes the constant specific or relative growth rates between measurements of plant size and shape (Batschelet 1973). Commonly the equation is fit to specific data sets through linear regression of logarithmically transformed data known as log-log or power regression (Grove and Malajczuk 1985). Log-log regression has the form: $\ln(Y) = a + b \ln(X)$. Its preferred use is the result of a more precise fit to the transformed data as indicated by higher values for the coefficient of determination (R^2), and lower values for the standard errors of the estimate ($S_{y.x}$) (Stewart et al. 1979, Huges et al. 1987).

Such transformation of data has been demonstrated to sometimes introduce bias into the results (Baskerville 1972, Lee 1982, Sprugel 1983). Theoretically this bias represents the introduction of an underestimate of individual plant biomass values resulting from the log-antilog transformations. In some studies this bias appeared to make only a small difference (Madgwick and Satoo 1975, Turner and Long 1975, Pastor and Bockheim 1981), and a proposed correction factor (Lee 1982, Sprugel 1983) has been shown to result in overestimation of plant biomass in some instances (Westmann and Rogers 1977). Transformation can also result in difficulties in appropriate evaluation by the usual measures of goodness of fit such as the coefficient of determination (R^2) and standard error of the estimate. This occurs because they apply only to the logarithmically transformed data (Payandeh 1981, Chiyenda and Kozak 1982). Previous attempts at using the allometric equation for biomass determination in woodlands dominated by singleleaf pinyon and Utah juniper

(*Juniperus osteosperma* [Torr.] Little) did not consider the bias problem (Tausch 1980).

Avoidance of the potential introduction of bias into the results of regression analyses for biomass estimation by the allometric equation is possible by using untransformed data in an iterative nonlinear regression technique (Payandeh 1981, Chiyenda and Kozak 1982). Determination of the possible presence of bias is best accomplished by the simultaneous use of both log-log and nonlinear regression techniques and comparison of the results (Schlaegel 1981, Brand and Smith 1985).

The first objective of this study was a comparison of the use of log-log and iterative nonlinear regression for the estimation of leaf biomass or phytomass of singleleaf pinyon. The second objective was the comparison of measurements of basal area and crown volume for predicting the phytomass of both individual trees and entire stands. Regression results were also evaluated for how they were affected by different ages of the trees in a stand.

STUDY SITE DESCRIPTION

Tree data used for most of the analyses were associated with four locations in the pinyon-juniper woodlands of the Sweetwater Mountains in western Nevada, first sampled by Meeuwig (1979) and Meeuwig and Budy (1979). These locations ranged in elevation from just over 2,000 m to 2,300 m (Table 1). All four sites were in woodlands that were fully stocked or fully tree-occupied as described by Meeuwig and Budy (1979). Some juniper was present in plot S4, but plots S1, S2, and S3 contained only singleleaf pinyon. Data from four additional sites (Table 1) in central Nevada from Meeuwig (1979) plus data from a site in southwestern Utah from Tausch (1980) were included in one analysis. All five additional sites were closed-stand woodlands with mixed pinyon and juniper.

METHODS

Sampling Procedures

The Sweetwater Mountains plots from Meeuwig (1979) and Meeuwig and Budy (1979) were 30 × 30 m in size. All trees in the four plots were measured for height, average crown diameter, and basal diameter. All trees

TABLE 1. Elevation, slope, and aspect for eight plots from Meeuwig (1979) and one plot from Tausch (1980). Mountain ranges represented are the Sweetwater Mountains of Nevada and California (S), Paradise Range of central Nevada (P), Monitor Range of central Nevada (M), and the Needle Range of southwestern Utah (UT).

Plot number	Aspect (degrees)	Slope (degrees)	Elevation (meters)
S1	80	3	2210
S2	40	11	2100
S3	120	8	2300
S4	345	20	2020
P1	354	3	2060
P2	55	9	2040
P3	85	3	2190
M1	60	11	2220
UT	285	2	2015

were harvested and a subsample was used for the determination of bole, bark, branch, twig, and foliage wet and dry biomass. Multiple regression techniques were used to derive the reported total dry weight of the components for each tree. Paradise and Monitor Range plots from Meeuwig (1979) were sampled in a similar manner. Tree data from the Needle Range, southwestern Utah, were sampled following the Sweetwater Mountains procedures but on a 0.1-ha plot.

A random sample of 12 singleleaf pinyon, stratified by age class, was harvested adjacent to each of the previously sampled plots, S1, S3, and S4, in an area with the same site and stand conditions. Plot S2 (Table 1) did not have a suitable adjacent area. Age classes were determined as described by Blackburn and Tueller (1970). A baseline was located in the adjacent area parallel to, and 10 m from, the side of each plot, and the selected trees were those closest to a random point on that line. Each tree was measured for average crown diameter, tree height, and basal diameter. Basal diameter measurements were just above the root crown and about 15 cm above the ground surface. Where multiple trunks branched below this point, each trunk was individually measured and an average basal diameter determined as described by Meeuwig and Budy (1979).

The harvested trees were sampled by two methods depending on tree size. All green foliage was collected on trees with basal trunk diameters up to about 15 cm. A trunk cross section was collected at the point of basal diameter measurement. The relationship of

sapwood area to phytomass has been shown to remain linear throughout the crown for several species of conifers (Long et al. 1981, Kaufmann and Troendle 1981). This relationship was used to subsample trees with basal trunk diameters greater than 15–18 cm. Two subsamples were collected, one from the main stem and one from a random major branch in the middle one-third of the tree. Main stem and branch subsamples were cut at a diameter of about 12 cm. A cross section was collected from the base of each stem and branch subsample, and a trunk cross section was collected from the base of the tree. All green foliage was collected from each stem and branch subsample during the first week of August.

Analysis Techniques

Crown volume was computed for each harvested tree from the average crown diameter and tree height using the formula for one-half of an ellipsoid (Tausch 1980). Crown volumes in plots S1 through S4 were computed from the average crown diameter and tree height data provided by Meeuwig (1979) and Meeuwig and Budy (1979). Basal diameter measurements were converted to basal area.

Trunk, stem, and branch cross sections from the subsampled trees were measured for heartwood and sapwood area with a dot grid. Sapwood areas and phytomass data from the two subsamples of each tree were combined and an overall ratio of kg of phytomass per cm^2 of sapwood area determined. This ratio was multiplied by the sapwood area of the trunk cross section to estimate the total phytomass of the tree.

Two prediction relationships were used at each site, one using basal area and the other using crown volume, to predict phytomass. Both log-log regression and nonlinear regression were used to derive prediction equations for each relationship for a total of four equations for each site. For nonlinear regression the parameters a and b were determined by an iterative least-squares technique.

Regression results were compared for each prediction relationship on each plot. Log-log regression results were converted back to arithmetic form (anti-log) prior to comparison with results from nonlinear regression analyses. Because R^2 from log-log analysis applies only to transformed data, it cannot be used for

direct comparison with nonlinear regression results (Payandeh 1981). Results from the two types of analyses were compared on the basis of a Fit Index (FI) and a standard error of the estimate ($S_{y.x}$) computed from untransformed data for both analyses as recommended by Payandeh (1981) and Brand and Smith (1985). The term *Fit Index* is from Brand and Smith (1985). The FI is a measure of the variation in Y explained by the variation in X:

$$FI = 1 - (\Sigma(\hat{Y}_i - Y_i)^2 / \Sigma (Y_i - \bar{Y})^2).$$

The FI and R^2 are equivalent for linear regression without data transformation. For the log-log and nonlinear regression analyses the maximum value for FI is also 1.0, but unlike R^2 it can be negative (Payandeh 1981). The standard errors of the estimate for all analyses were computed as the square root of the sums of deviations squared (based on untransformed data) divided by the number of observations minus 2 (Steel and Torrie 1960).

Equations derived from the trees collected next to plots S1, S3, and S4 were used to predict the total pinyon phytomass of the respective adjacent plot. All four possible combinations of regression techniques and prediction relationships were used for each site. Basal areas and crown volumes used on these plots were derived from data provided by Meeuwig (1979) and Meeuwig and Budy (1979). Predicted biomass values were then compared with the published values.

The nonlinear regression results for the basal area and crown volume to phytomass relationships for plots S1, S3, and S4 were used to predict the total plot phytomass for plot S2. This provided a test of how accurately an equation derived for one plot could predict total plot phytomass on a different plot. The results were then evaluated on the basis of differences in singleleaf pinyon age between predicting and predicted sites. Average tree ages for sampled plots from Meeuwig (1979) and Meeuwig and Budy (1979) were computed from individual tree age data they provided.

The effects of average pinyon age in a stand on biomass prediction were further evaluated by comparing the differences over nine sites with trees of different ages. Eight of the sites are from Meeuwig (1979) and one is from Tausch (1980). Nonlinear regression was used

TABLE 2. Basal area (cm²) to tree phytomass (g) log-log and nonlinear regression equation parameters for trees harvested adjacent to three plots studied by Meeuwig (1979) and Meeuwig and Budy (1979) on the Sweetwater Mountains. Plot designations are described in Table 1.

Plot number	Log-log regression					Nonlinear regression			
	a	b	R^2	Fit index	Standard error (g)	a	b	Fit index	Standard error (g)
S1	11.75	1.235	0.98	0.83	9159.1	85.85	0.921	0.88	7806.5
S3	14.82	1.170	0.98	0.90	4299.4	56.63	0.958	0.92	3934.2
S4	14.88	1.139	0.96	0.12	14592.2	397.61	0.620	0.93	4193.9

TABLE 3. Crown volume (m³) to tree phytomass (g) log-log and nonlinear regression equation parameters for trees harvested adjacent to three plots studies by Meeuwig (1979) and Meeuwig and Budy (1979) on the Sweetwater Mountains. Plot designations are described in Table 1.

Plot number	Log-log regression					Nonlinear regression			
	a	b	R^2	Fit index	Standard error (g)	a	b	Fit index	Standard error (g)
S1	852.48	0.988	0.96	0.58	14563.2	3896.0	0.568	0.82	9486.0
S3	912.78	0.935	0.99	0.89	4619.5	1851.5	0.739	0.91	4088.5
S4	462.44	1.068	0.97	0.87	5555.6	698.0	0.950	0.88	5412.1

to derive equations predicting phytomass from basal area for all nine sites. The estimated phytomass of a tree on each site with a basal area equivalent to a basal diameter of 25.4 cm (10 in) was determined. The 25.4-cm (10-in) diameter is the one most commonly used for site comparisons in singleleaf pinyon (Chojnacky 1986). These tree phytomass values and the average pinyon ages in the plots were used in the regression analysis.

RESULTS

Nonlinear regression consistently provided a better fit for the basal area to tree phytomass relationship for harvested trees from each site than log-log regression with transformed data (Table 2). This is evident in both higher Fit Index values (closer to 1.0) and lower standard error of the estimate values. Results were similar for the crown volume to tree phytomass relationships in Table 3. Fit Index values from each log-log regression are always lower than the respective R^2 (coefficient of determination) from the same analysis. This reflects the unreliability of the R^2 values based on transformed data. Differences are particularly evident for plot S4 in Table 2 and plot S1 in Table 3 where high R^2 values from log-log regression are associated with low Fit Index values and poor standard errors computed from untransformed (anti-log) data.

Crown volume was consistently less precise than basal area in predicting tree phytomass. This appears related to variability of density and distribution of phytomass in the tree crowns, thus making these measurements more variable than those from basal area. Values of the regression parameter b are nearly always lower for the nonlinear regression method (Tables 2, 3).

Equations for trees harvested adjacent to plots S1, S3, and S4 were used to estimate the total pinyon phytomass of their respective plots as a direct check on predictability (Table 4). Estimates of total phytomass in the three plots by nonlinear regression had average errors of +0.5 % for prediction by basal area and +5.7% for prediction by crown volume. The equivalent figures for log-log regression were +39.3% and +35.3%. A substantial overestimate related to the bias associated with data transformation for power regression was evident. The error, however, is opposite that usually reported in the literature. Application of a correction factor to the log-log regression results (Lee 1982, Sprugel 1983) would further increase the overestimate, similar to the results reported by Westmann and Rogers (1977).

Prediction of total tree phytomass from basal area on the test plot S2 by nonlinear regression equations from plots S1, S3, and S4 (Table 2) revealed an inverse relationship

Table 4. Prediction of total pinyon phytomass from crown volume and basal area of individual trees on three plots studied by Meeuwig and Budy (1979) and Meeuwig (1979) and comparison with the published values. Plot designations used are those from Meeuwig (1979). Nonlinear and log-log regression equations used are for data from trees randomly collected adjacent to the plots.

| Test plot | Computed by nonlinear regression equations | | | | Computed by log-log regression equations | | | |
	Phytomass from crown volume (1)	Percent error	Phytomass from basal area (2)	Percent error	Phytomass from crown volume (1)	Percent error	Phytomass from basal area (2)	Percent error
S1	1171.6	+11.5	1052.8	+0.2	1608.3	+53.0	1257.3	+19.6
S3	1300.2	+2.5	1284.2	+1.2	1701.6	+34.1	1729.4	+36.3
S4	878.6	+3.2	851.4	+0.05	1101.8	+18.9	1378.6	+62.0
Average error		+5.7		+0.5		±35.3		+39.3

[1]Equations used from Table 3
[1]Equations used from Table 2

Table 5. Estimation of total plot phytomass for test plot S2 from Meeuwig (1979) by equations based on trees harvested adjacent to sample plots S1, S3, and S4 in Table 1. Equations were derived from nonlinear regression analyses for the basal area to tree phytomass relationship. Average pinyon age data are from Meeuwig (1979).

Plot used for estimation equation	Estimated phytomass for plot S2 (kg)	Deviation from actual phytomass (%)	Estimated average tree age (yrs)
S1	1623	+12.6	88.2
S3	1364	−5.4	122.4
S4	1064	−26.2	242.7
Actual S2	1441	0	128.6

between the average tree age on the plot used to develop the equation and estimated total tree phytomass for plot S2 (Table 5). The equation from plot S4, with trees older than the test plot, underestimated actual total tree phytomass of S2. The equation from plot S1, with younger trees, overestimated actual phytomass. Plot S3, with an average tree age the most similar to S2, came the closest to predicting its actual total tree phytomass.

Effects of age on the basal area to phytomass relationship were further investigated using the data from the eight plots from Meeuwig (1979) and the plot from Tausch (1980). A regression analysis between the average age of the trees on each plot and the respective phytomass of a tree with a basal area equivalent to a diameter of 25.4 cm (10 in) was significantly negative (Fig. 1). The variability in Figure 1 indicates that other environmental and stand differences are involved in the estimation errors between sites. These could include differences in site potential, tree size, competing species, and edaphic and topographic conditions between sites.

Discussion

The more commonly used log-log regression was consistently the poorest performer in both precision and prediction accuracy. Nonlinear regression provided better equations for predicting tree biomass for singleleaf pinyon on the sites studied. Differences in the equations between the two regression methods stem from an interaction between logarithmic transformation of the data and the distribution of tree sizes in a stand. There is a skewed distribution with more small trees than large ones. Smaller trees also generally have denser crowns with more foliage biomass per unit of crown volume or unit of basal area than the larger trees. Larger trees, however, usually contribute most of the total tree phytomass in a plot. Logarithmic transformations used for power regression decrease the values for the larger trees more than for the smaller trees. As a result, the smaller trees with their denser crowns have greater weight in the log-log than in the nonlinear regression analyses.

The effects of data transformation on the resulting equations are evident in the higher values of the regression parameter b for log-log than for nonlinear regression (Tables 2, 3). Consistent overestimation of the phytomass values for the largest trees and for the total tree phytomass in the individual plots is the result. Each tree contributes to the regression result more in proportion to its size with nonlinear regression, thus tending to reduce or eliminate the overestimation. For this type of
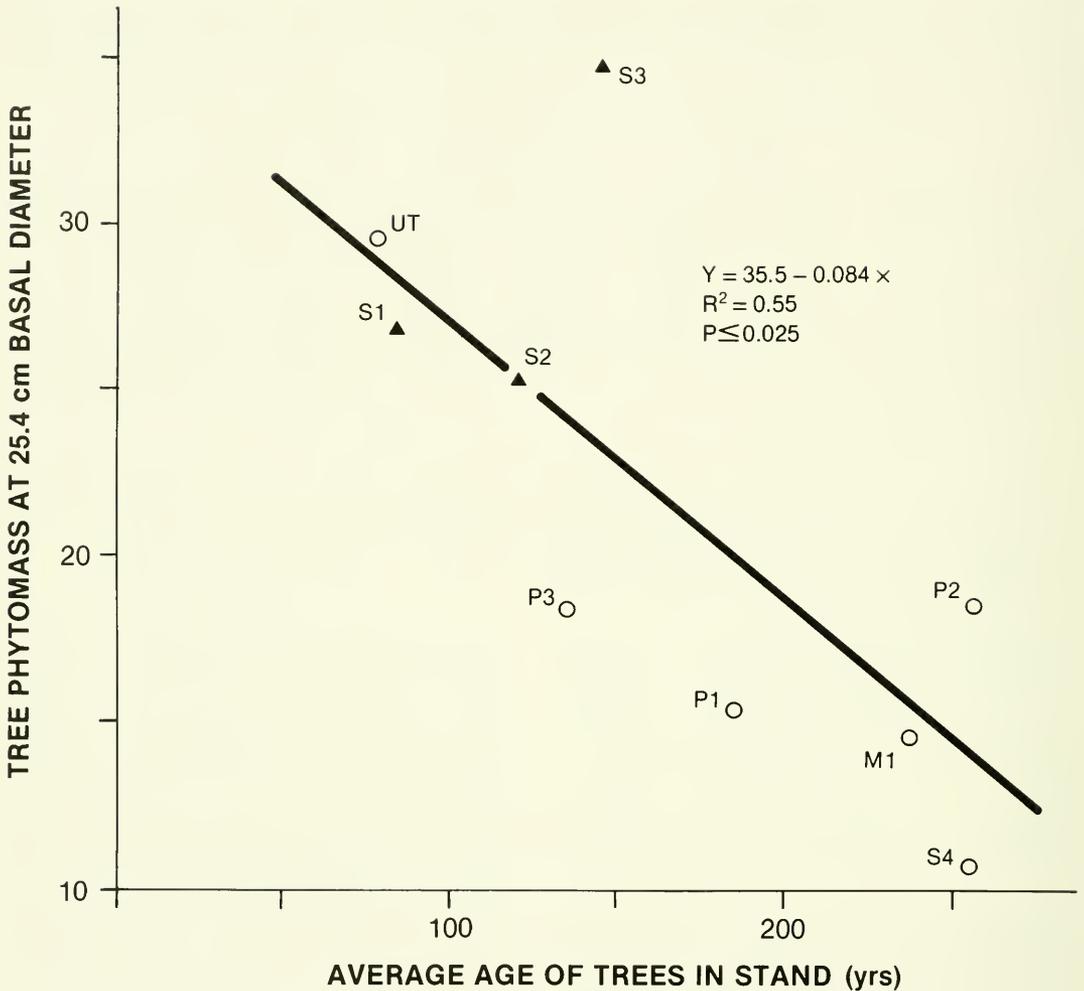
Fig. 1. Regression analysis between the average age of a tree in a stand and the estimated average phytomass of a tree with a basal diameter of 25.4 cm (10 in). Tree phytomass was determined from a nonlinear regression analysis between basal area and tree phytomass for all the trees on each site. Plot definitions are in Table 1.

data, nonlinear regression provides more accurate estimates on an individual tree and total plot basis than the more commonly used log-log regression.

The highly site-specific nature of tree age on analysis results is apparently related to growth characteristics of trees of semiarid climates. Studies of semiarid sites in Australia have shown that the phytomass of mature trees becomes constant after a certain size is reached (Sharpe et al. 1985). Basal area continues to increase after phytomass has become constant, however, resulting in a decline in the ratio of phytomass to basal area with increasing age of the trees. Because younger trees have higher phytomass to basal area ratios, equations developed from plots with younger trees overestimate the phytomass of plots with older trees. Conversely, equations developed from plots with older trees, with their lower ratios, underestimate plots with younger trees. In most instances new equations should be developed for each site where biomass estimates are needed.

and Range Experiment Station agreement number 22-C-4-INT-26.

LITERATURE CITED

ALABACK, P. B. 1986. Biomass regression equations for understory plants in coastal Alaska: effects of species and sampling design on estimates. Northwest Sci. 60: 90–103.

BASKERVILLE, G. L. 1972. Logarithmic regression in the estimation of plant biomass. Canadian J. For. Res. 2: 49–53.

BATSCHLET, E. 1973. Introduction to mathematics for life sciences. Springer-Verlag, New York. 495 pp.

BLACKBURN, W. H., AND P. T. TUELLER. 1970. Pinyon and juniper invasions in black sagebrush communities in east central Nevada. Ecology 51: 841–848.

BRAND, G. J., AND W. B. SMITH. 1985. Evaluating allometric shrub biomass equations fit to generated data. Canadian J. Bot. 63: 64–67.

CHIYENDA, S. S., AND A. KOZAK. 1982. Some comments on choosing regression models for biomass prediction equations. For. Chron. 58: 203–204.

CHOJNACKY, D. C. 1986. Pinyon-juniper site quality and volume growth equations for Nevada. USDA For. Serv. Res. Pap. INT–372.

COCHRAN, P. H. 1982. Estimating wood volumes for douglas-fir and white fir from outside bark measurements. Forest Sci. 28: 172–174.

GROVE, T. S., AND N. MALAJCZUK. 1985. Biomass production by trees and understory shrubs in an age-series of *Eucalyptus diversicolor* F. Muell. Stands. For. Ecol. and Manage. 11: 59–74.

HATCHELL, G. E., C. R. BERRY, AND H. D. MUSE. 1985. Nondestructive indices related to aboveground biomass of young loblolly and sand pines on ectomycorrhizal and fertilizer plots. For. Sci. 31: 419–427.

HUGES, H. G., L. W. VARNER, AND L. H. BLANKENSHIP. 1987. Estimating shrub production from plant dimensions. J. Range Manage. 40: 367–369.

HUMPHREY, L. D. 1985. Use of biomass predicted by regression from cover estimates to compare vegetation similarity of sagebrush-grass sites. Great Basin Nat. 45: 94–98.

KAUFMANN, M. R., AND C. A. TROENDLE. 1981. The relationship of leaf area and foliage biomass to sapwood area in four subalpine forest tree species. For. Sci. 27: 477–482.

LEE, C. Y. 1982. Comparison of two correction methods for the bias due to the logarithmic transformation in the estimation of biomass. Canadian J. For. Res. 12: 326–331.

LONG, J. N., F. W. SMITH, AND D. R. M. SCOTT. 1981. The role of douglas-fir stem sapwood and heartwood in the mechanical and physiological support of crowns and development of stem form. Canadian J. For. Res. 11: 459–464.

MEEUWIG, R. O. 1979. Growth characteristics of pinyon-juniper stands in the western Great Basin. USDA For. Serv. Res. Pap. INT–238. 22 pp.

MEEUWIG, R. O., AND J. D. BUDY. 1979. Pinyon growth characteristics in the Sweetwater Mountains. USDA For. Serv. Res. Pap. INT–227. 26 pp.

PASTOR, J., AND J. G. BOCKHEIM. 1981. Biomass and production of an aspen mixed hardwood-spodosol ecosystem in northern Wisconsin. Canadian J. For. Res. 11: 132–138.

PASTOR, J., J. D. ABER, AND J. M. MELILLO. 1983. Biomass prediction using generalized allometric regressions for some northwest tree species. For. Ecol. and Manage. 7: 265–274.

PAYANDEH, B. 1981. Choosing regression models for biomass prediction equations. For. Chron. 57: 229–232.

SCHLAEGEL, B. D. 1981. Testing, reporting, and using biomass estimation models. Pages 95–112 *in* Proceedings of 1981 Southern Forest Biomass Workshop, June 11, 12, 1981, Clemson University, Clemson, South Carolina. USDA For. Serv. S. For. Exp. Sta., New Orleans.

SHARPE, P. J. H., J. WALKER, L. K. PENDRIDGE, AND H. WU. 1985. A physiologically based continuous-time markov approach to plant growth modelling in semi-arid woodlands. Ecol. Model. 29: 189–213.

SPRENT, P. 1972. The mathematics of size and shape. Biometrics 28: 23–27.

SPRUGEL, D. G. 1983. Correcting for bias in log-transformed allometric equations. Ecology 64: 209–210.

STEEL, R. G. D., AND J. H. TORRIE. 1960. Principles and procedures of statistics. McGraw-Hill Book Co., New York.

STEWART, H. T. L., D. W. FLINN, AND B. C. AEBERLI. 1979. Above-ground biomass of a mixed eucalypt forest in eastern Victoria. Australian J. Bot. 27: 725–740.

TAUSCH, R. J. 1980. Allometric analysis of plant growth in woodland communities. Unpublished dissertation, Utah State University, Logan.

TURNER, J., AND J. N. LONG. 1975. Accumulation of organic matter in a series of douglas-fir stands. Canadian J. For. Res. 5: 181–190.

WESTMANN, W. E., AND R. W. ROGERS. 1977. Biomass and structure of a subtropical eucalypt forest, north Stradbroke Island. Australian J. Bot. 25: 171–191.