OTAGO AND SOUTHLAND TREE VOLUME EQUATIONS

A.G.D. WHYTE

REPORT NO. 36 MAY 1994

Note : This is an unpublished report and must not be cited as a literature reference.

EXECUTIVE SUMMARY

This report reviews the variability in existing radiata pine tree volume and taper equations for the Otago/Southland region and suggest some ways of using current information to possibly better effect. The major aim of the study was to investigate the need for individual equations and to determine whether or not all the relevant data could be better amalgamated in an overall equation with or without recognition of individual sub-populations. There were two secondary objectives, the first being to review the general approach to construction and use of tree volume and taper equations in New Zealand, and the second to provide users in the region with indication of which equation to employ in their own locality.

OTAGO AND SOUTHLAND TREE VOLUME EQUATIONS

A.G.D. WHYTE

INTRODUCTION

This report reviews the variability in existing radiata pine tree volume and taper equations for the Otago/Southland region and suggests some ways of using current information to possibly better effect. The major aim of the study was to investigate the need for individual equations and to determine whether or not all the relevant data could be better amalgamated in an overall equation with or without recognition of individual sub-populations. There were two secondary objectives, the first being to review the general approach to construction and use of tree volume and taper equations in New Zealand, and the second to provide users in the region with an indication of which equation to employ in their own locality.

There are five existing radiata pine equations for Otago and Southland trees, all of the form:

$$v = d^{b_1}(h^2 / (h - 1.40))^{b_2}.\exp(b_3)$$

where

v = total stem volume inside bark in m^3
d = diameter at breast height over bark in cm
h = total tree height in metres
b_i are least-squares non-linear regression coefficients

T126 was prepared for Otago Coast forests in 1974, T227 for Pomahaka, Pukerau Block, compartment 421, 500 stems/ha age 30 in 1986, T232 for Longwood, Woodlaw Block, ages 14-24 in 1986 and T235 for Longwood age 30, 370 stems/ha in 1986. T13 was based on all Southland radiata pine and prepared in 1952.

The coefficients and data statistics are set out below in Table 1

where

b_i are least-squares non-linear regression statistics
n is number of trees for which a regression was derived
$\overline{v}$ is the average volume sampled
dmin, hmin are minimum dbhob's and heights sampled
dmax, hmax are maximum dbhob's and heights sampled

Table 1 Existing Volume Equations

Equation	b1	b2	b3	n	$\overline{V}$ (m^3)	dmin (cm)	dmax (cm)	hmin (m)	hmax (m)
T126	1.801022	1.148053	-10.30559	265	1.000	12	47	8	30
T227	1.889000	1.122600	-10.54122	87	2.290	37	65	28	38
T232	1.765996	1.104103	-10.09289	105	0.180	9	32	7	19
T235	1.869605	1.202020	-10.72714	96	2.260	30	67	26	39
T13	1.865000	1.071000	-10.28214	365	4.070	10	94	9	43

Indications are that there is not much difference in the coefficients, which suggests that there is an opportunity to combine the data from the various sub-populations.

There are only three taper equations F227, F232 and F235 corresponding to T227, 232 and 235 respectively. All are of the form:

$$dib^2 = \frac{v}{kh}\left(b_1(\ell/h) + b_2(\ell/h)^2 + b_3(\ell/h)^3 + b_4(\ell/h)^4 + b_5(\ell/h)^5 + b_6(\ell/h)^{b_7} + b_8(\ell/h)^{b_9}\right)$$

where

dib is stem diameter inside bark in cm at a distance ℓ from the tip of the tree
v is volume in m^3 inside bark predicted by the corresponding volume equation
h is total height of tree in m
ℓ is distance along stem from tip also in m
$k = \pi / 40\,000$

Table 2 Existing Taper Equations

Equation	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	n	S_{dib}%
F227	0.2915	6.414	-7.984	3.440	-	0.55596	40	0.97279	90	87	1.50
F232	0.98748	-	8.004	-14.25	8.076	0.84266	90	-	-	105	0.83
F235	0.5958	3.6414	-2.1572	-	-	1.6814	60	-	-	96	1.55

A simulation was carried out to compare estimates for inputs of the same diameter at breast height over bark (d) and total height (h) over a range of both.

Table 3 Estimates for a range of tree dimensions of volume

d (cm)	h (m)	Estimated Volume (v) in m3 by T13	T126	T227	T232	T235
12	8	0.0401	0.0399	0.0411	0.0409	0.0351
37	28	1.0789	1.0857	1.0812	1.0199	1.0951
50	33	2.2366	2.2344	2.2760	2.0629	2.3205
65	38	4.2171	4.1863	4.3488	3.8067	4.4588

This comparison showed that the two largest sample sizes (365 trees for T13 from Southland and 265 trees for T126 from Otago Coast) gave almost identical results: the maximum difference was 0.76 per cent, well within the standard error of over 6.6 per cent. T227 and T235 were also close to both T13 and T126 estimates within their sampled range of data and certainly within their standard errors. Even T232, constructed from a maximum d = 32 cm and h = 19 m, was reasonably consistent for d as high as 65 cm and h as high as 38 m. This evidence would suggest that there is a very real opportunity to consolidate a representative tree volume equation for Otago and Southland jointly. Eggleston (1992) pointed out that T126, T232 and T235 had the narrowest range of dbhob among all those in the national list of equations.

The two largest sub-populations, T13 and T126, have no corresponding taper equations. Eggleston (1992) recommends that F227 (taper equation based on only 87 trees, all of age 30 and in stands of 500/ha) be used in tandem with T13 for the whole of Southland and that F230 for Golden Downs be paired with T126, the Otago Coast volume equation. These pairings could be re-evaluated, though that is not strictly within the terms of reference for this study.

ANALYTICAL PROCEDURES

Data were made available for: T126A (metric, 3 m sections, early format); T126B (imperial, 10 feet sections for tree heights > 56 feet and 5 feet sections for shorter trees); T232, T235 (metric, 3 m sections, current format); and T227 (metric, irregular lengths of section). There is, therefore, a mix of measuring methodology which confounds locality and, hence, the analysis. Other confounding factors are the different ages and stocking regimes for the four forest populations.

The data were supplied on diskette, but, because of the lack of consistency from one sub-population to another, it was decided to key in the data again to conform with an easier programming format. Consequently, there may be some data transcription errors that have occurred. On the other hand, several anomalies in the data, in addition to the ones noted on the FRI data files, were revealed.

The apparent reliability of each stem profile was visually assessed on the computer screen. Unusual readings (excessive departures in the smoothness of over and under bark profiles, unrepresentative breast-height diameters and the like) were queried. Generally, if the error was not a genuine mis-read or an obvious mis-record, the apparent anomaly was allowed to stand. All the readings at 0.15 m were retained in deriving volume, but eliminated from the taper analysis. A few more (but less than 1 per cent) other individual diameter and height readings were removed later during the analysis of stem taper, while a dozen or so trees were excluded from the stem volume equations because of excessive deviation (four standard deviations or more from the average) and anomalous measurements that could not be corrected.

Polynomial equations were used for both the volume and taper models. All volume models were weighted, the apparently appropriate weight in each case being $(1/d^2h)$. When taper equations were analysed, there was no apparent need for polynomials of order higher than 5, probably because the diameters at 0.15 m had been removed from the analysis. The individual coefficients finally chosen were derived both with and without the restriction that they sum to 1, in order to assess the bias incurred through forcing volume and taper equation compatibility.

The approach adopted here was volume based compatibility (i.e. the taper equations were made compatible with pre-determined volume equations) because the study was aimed primarily at volume equations. Nevertheless, a taper-based system approach is intuitively more appealing and should not be discounted in future work in this field.

An example is shown from two extremes of the data to show how consistent the volume and taper trends are. The trees making up T227 are from one compartment at a stocking of 500 stems/ha in the Pukerau Block of Pomahaka Forest, while those in T232 are from a 14 to 24 year old range of crop ages and several different stockings in the Woodlaw Block of Longwood Forest. There is absolutely no overlap in tree size in these two sub-populations.

Statistics for the volume equations pertaining to the two forests and the overall combined one are shown in Table 4 below for $v = b_0 + b_1 (d^2h)$, the solutions weighted by $(1/d^2h)$. Other polynomial terms (e.g. h, d^2, dh^2) were tried but found to be unsuccessful additions to d^2h, contributing no statistically significant reduction in the sums of squares when used in any combination with d^2h.

Table 4 Preliminary Analysis of Population Extremes, Woodlaw & Pukerau Data

	b_0	b_1	r^2	Root MSE
Woodlaw	0.008 976 ± 0.002 015	0.000 027 195 ± 0.000 000 41	0.9802	0.000 19
Pukerau	0.073 709 ± 0.055 641	0.000 026 530 ± 0.000 000 69	0.9426	0.000 51
Combined	0.008 811 ± 0.002 677	0.000 027 303 ± 0.000 000 15	0.9947	0.000 38

There is no apparent bias in the combined overall equations, as is shown in the accompanying plots of standardised residuals and in the studentised residual statistics shown on the following pages.

The corresponding taper equation with this combination of data also shows no difference between the two populations, though this cannot be seen as graphically as can be shown in the volume data. The form of the equation is:

$$d_k^2 = 2b_1z + 3b_2z^2 + 5b_4z^4 + 6b_5z^5$$

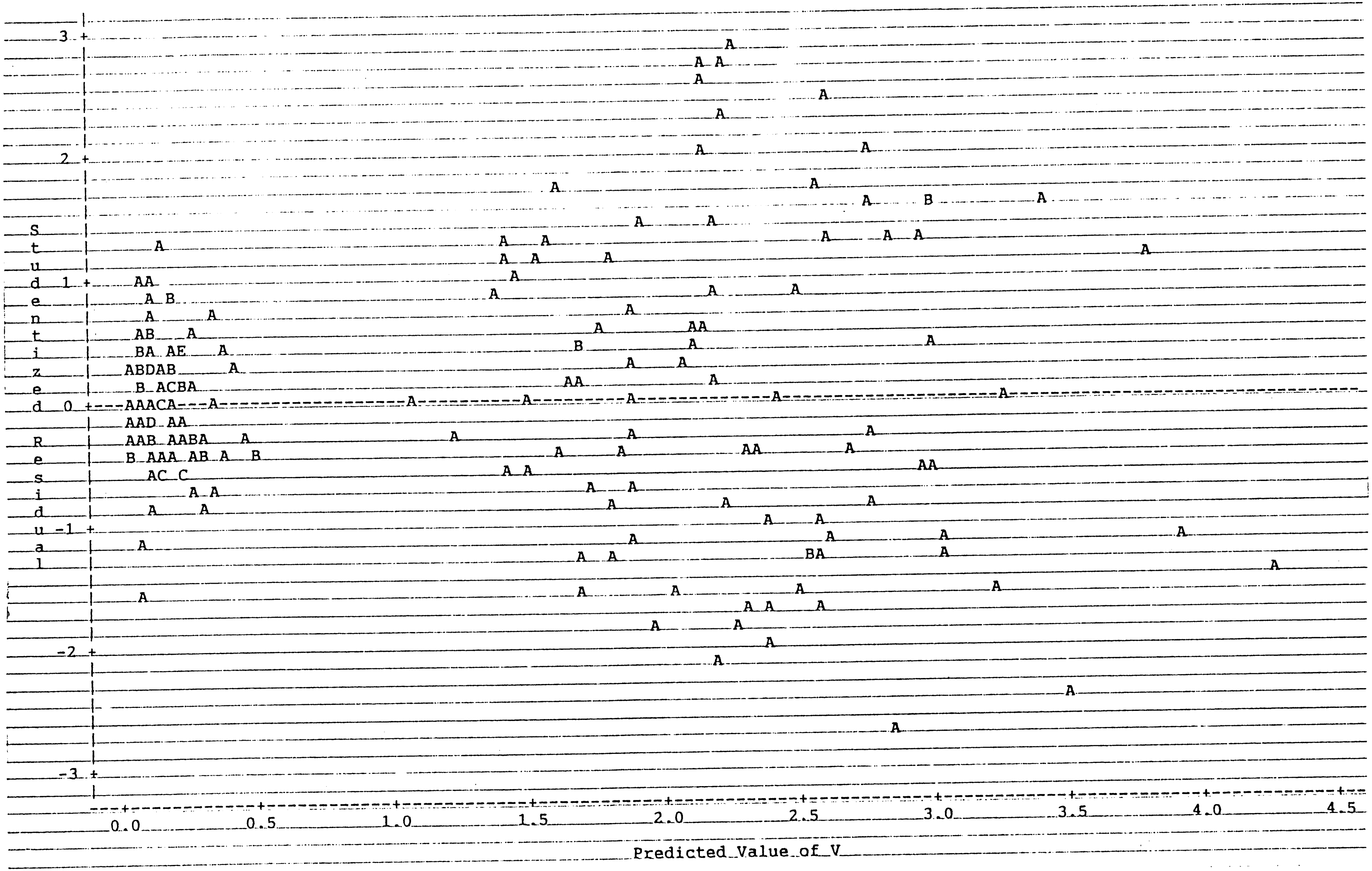
where

$$z = ((h - h_k)/h)$$

Note that the $4b_3z^3$ term was discarded because of insignificance. The results are summarised in the pages following the volume equation results. Note that the

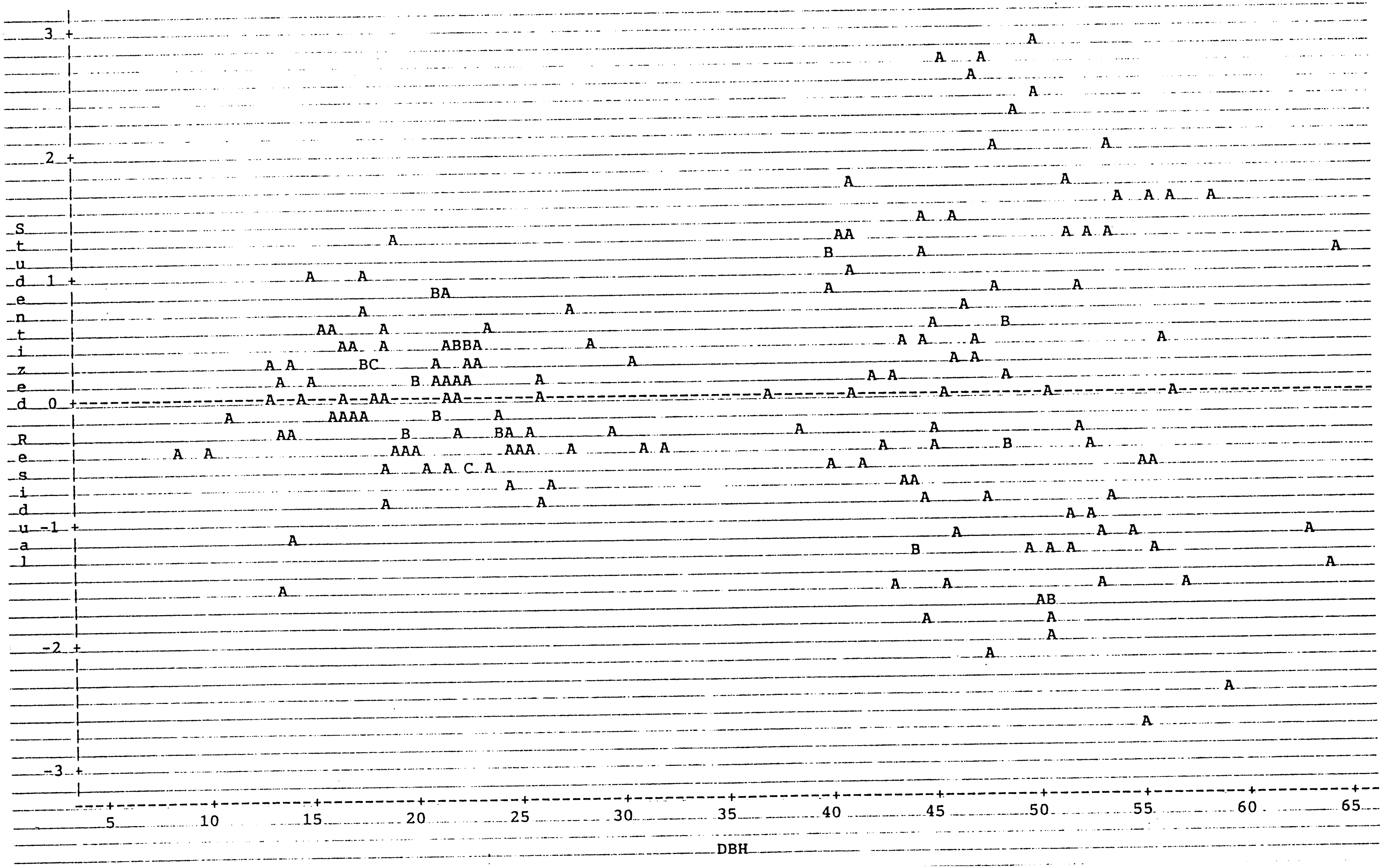
SAS

Plot of ST*PRED. Legend: A = 1 obs, B = 2 obs, etc.



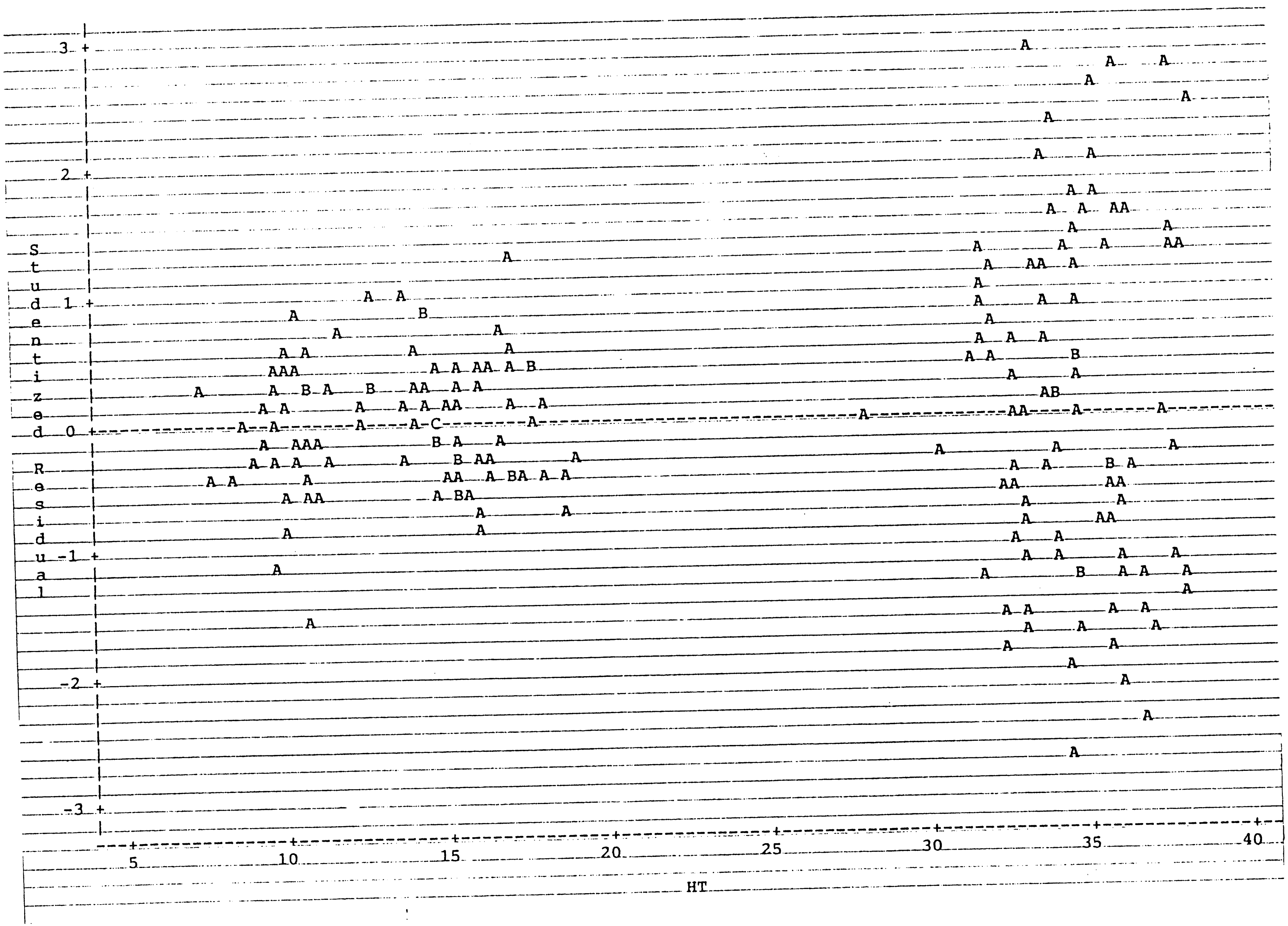
SAS

Plot of ST*DBH. Legend: A = 1 obs, B = 2 obs, etc.



SAS

Plot of ST*HT. Legend: A = 1 obs, B = 2 obs, etc.



SAS

UNIVARIATE PROCEDURE

```
Variable=ST          Studentized Residual

              Moments                       Quantiles(Def=5)                              Extremes

                                                                                 Lowest    Obs    Highest    Obs
N               182  Sum Wgts        182    100% Max   2.877374     99%  2.759252   -2.74103(   47)  2.483109(   41)
Mean       0.010642  Sum        1.936825     75% Q3    0.488275     95%  1.686117   -2.47187(   16)  2.596563(   42)
Std Dev    1.002998  Variance   1.006004     50% Med   0.000502     90%   1.25661   -2.11909(    2)   2.68742(   10)
Skewness   0.251559  Kurtosis   0.585342     25% Q1    -0.52565     10%  -1.28866     -1.971(    6)  2.759252(   27)
USS        182.1074  CSS        182.0868      0% Min   -2.74103      5%  -1.63636   -1.82973(   21)  2.877374(   17)
CV         9424.992  Std Mean   0.074347                             1%  -2.47187
T:Mean=0   0.143138  Prob>|T|     0.8863     Range     5.618406
Sgn Rank     -128.5  Prob>|S|     0.8573     Q3-Q1     1.013929
Num ^= 0        182                          Mode      -2.74103
W:Normal   0.976615  Prob<W       0.1751

                                                #   Boxplot                 Normal Probability Plot
   Stem Leaf                                                        2.75+                                            ** **
      2 56789                                   5      0                 |                                        ****+++
      2 003                                     3      0                 |                                     ****++
      1 555677                                  6      |                 |                                 +****
      1 00112222233344                         14      |                 |                             ++*****
      0 5555566667788889999                    19   +-----+              |                        ********
      0 00001111111111222222333333333334444444444444   44   *--+--*      |                   *******
     -0 444444444444333333333333222211111110000    39   |     |          |               *****
     -0 9988888776666655555555555              24   +-----+              |          +****
     -1 44333322211110                         14      |                 |     ********
     -1 8887766655                             10      |                 |+****
     -2 10                                      2      0                 |
     -2 75                                      2      0            -2.75+*
        ----+----+----+----+----+----+----+----                           +----+----+----+----+----+----+----+----+----+----+
                                                                               -2        -1         0        +1        +2
```

RESTRICTED MODEL
NB:COEFFS ARE *2 ,*3 ,*5,*6

Model: MODEL1
NOTE: Restrictions have been applied to parameter estimates.
NOTE: No intercept in model. R-square is redefined.
Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	4079.81287	1359.93762	82730.929	0.0000
Error	2213	36.37747	0.01644		
U Total	2216	4116.19034			

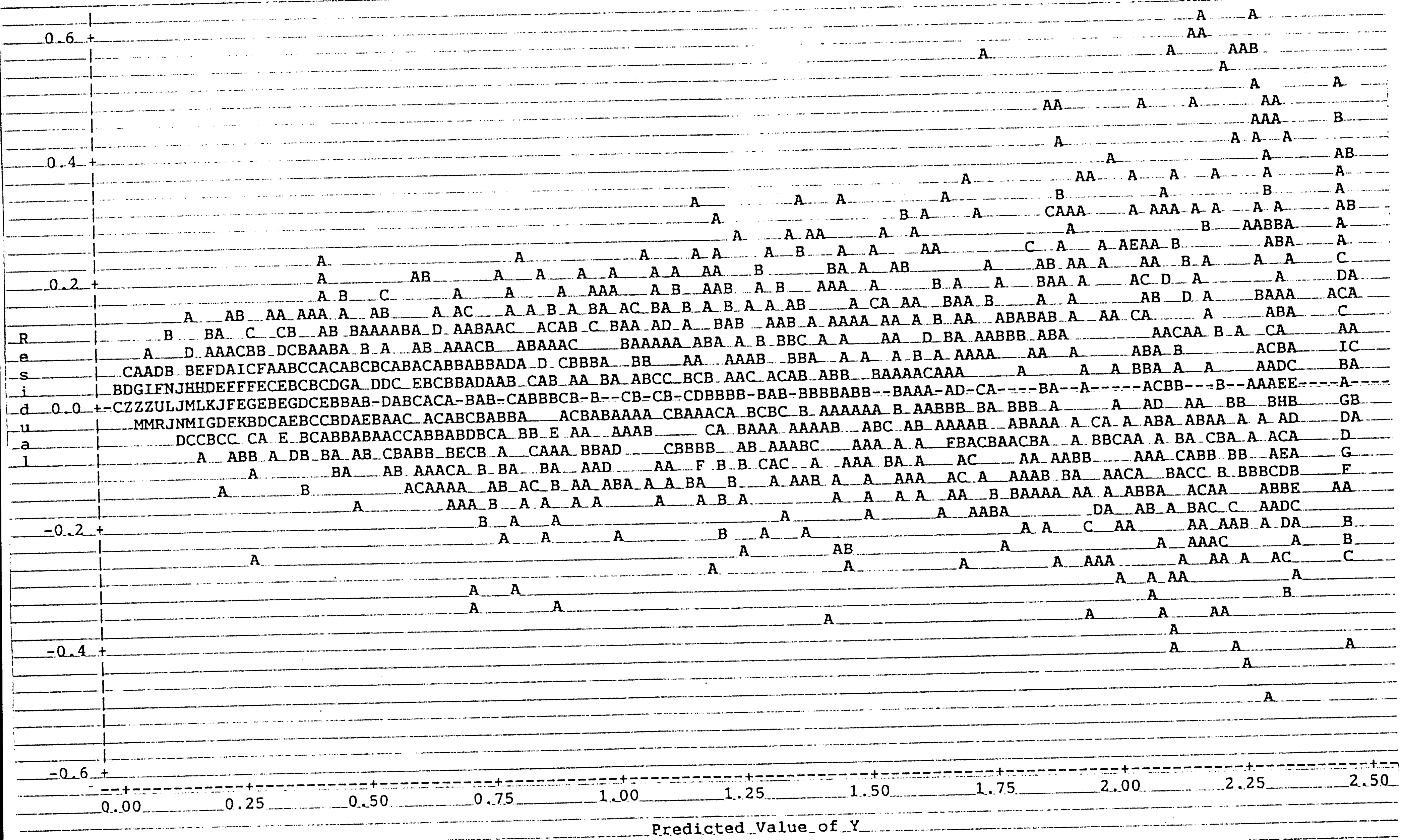
Root MSE	0.12821	R-square	0.9912
Dep Mean	1.09141	Adj R-sq	0.9912
C.V.	11.74732		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
Z1	1	0.193827	0.03043906	6.368	0.0001
Z2	1	1.467445	0.07408902	19.807	0.0001
Z4	1	-1.776997	0.11446549	-15.524	0.0001
Z5	1	1.115725	0.07042972	15.842	0.0001
RESTRICT	-1	10.428036	6.01391714	1.734	0.0831

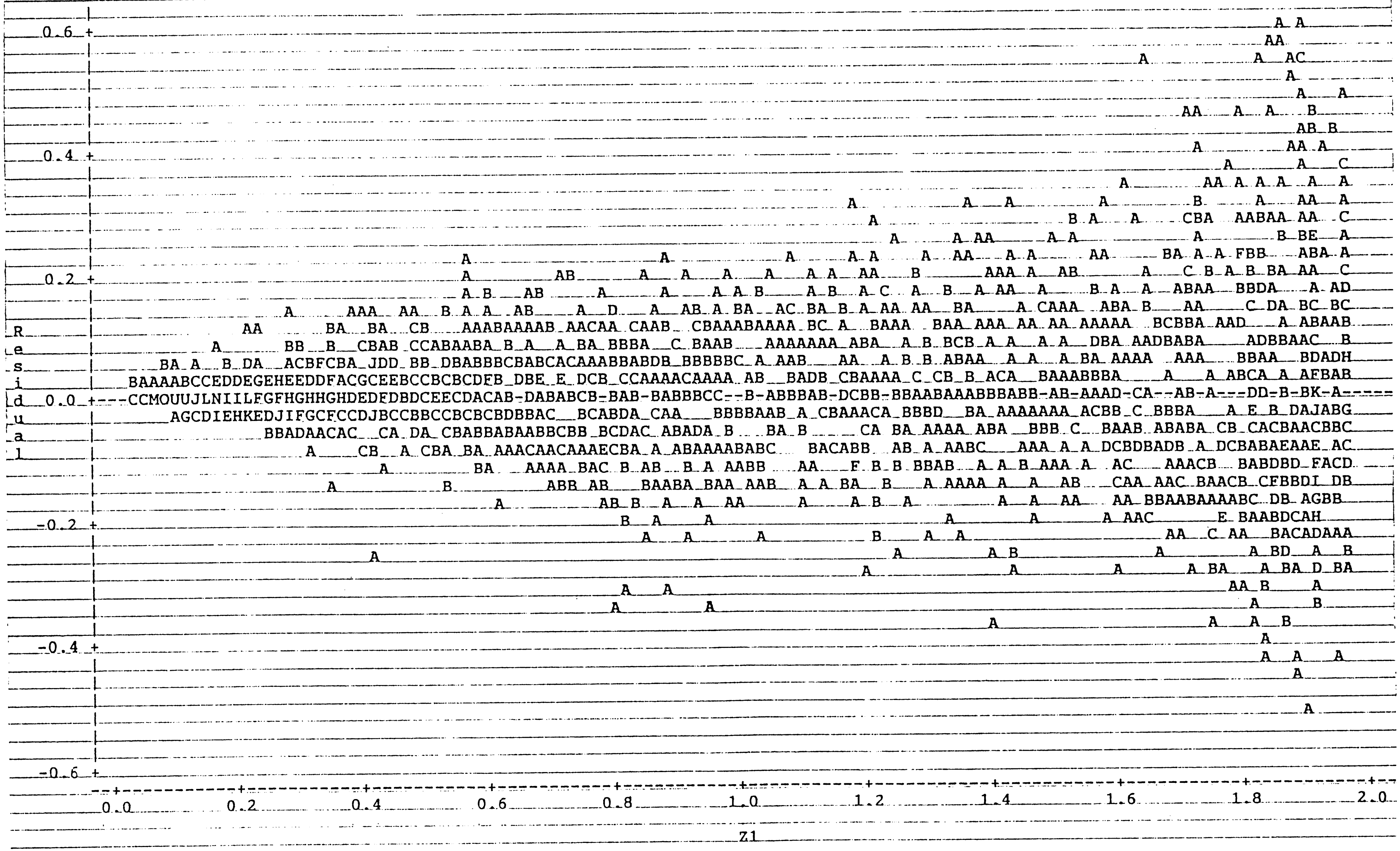
RESTRICTED MODEL
NB:COEFFS ARE *2 ,*3 ,*5,*6

Plot of RESID*PRED. Legend: A = 1 obs, B = 2 obs, etc.



RESTRICTED MODEL
NB:COEFFS ARE *2 ,*3 ,*5,*6

Plot of RESID*Z1. Legend: A = 1 obs, B = 2 obs, etc.



```
                                   RESTRICTED MODEL
                              NB:COEFFS ARE *2 ,*3 ,*5,*6

                                 UNIVARIATE PROCEDURE

Variable=RESID        Residual

               Moments                          Quantiles(Def=5)                          Extremes

 N               2216  Sum Wgts      2216   100% Max  0.589118     99%  0.417478      Lowest    Obs    Highest    Obs
 Mean        0.004593  Sum       10.17912    75% Q3   0.059377     95%  0.224949    -0.50726(   1306)  0.549774(   2040)
 Std Dev     0.128071  Variance  0.016402    50% Med  -0.00076     90%  0.151481    -0.44506(    200)  0.577987(   2088)
 Skewness    0.551834  Kurtosis  2.542134    25% Q1   -0.06082     10%  -0.14311    -0.43656(    876)  0.580715(   2208)
 USS         36.37747  CSS       36.33071     0% Min  -0.50726      5%  -0.19485    -0.43481(    597)  0.587778(   2121)
 CV          2788.108  Std Mean  0.002721                           1%  -0.30407    -0.41809(    877)  0.589118(   1845)
 T:Mean=0      1.6884  Prob>|T|    0.0915   Range     1.09638
 Sgn Rank        5113  Prob>|S|    0.8653   Q3-Q1      0.1202
 Num ^= 0        2216                       Mode     -0.00329
 D:Normal    0.086217  Prob>D        <.01

                  Histogram                        #  Boxplot                    Normal Probability Plot
                                                                    0.55+                                                *
    0.55+*                                         11     *              |                                             ***
         .*                                        16     0              |                                        ***  ++
         .**                                       24     0              |                                   ******+++
         .******                                   91     0              |                              +********
         .****************                        247     |              |                        ***********
         .*********************************************   707  +--+--+   |                ************
         .***********************************************  750  *-----*  |          *********+
         .*****************                       265     |              |    ********+
         .******                                   82     0              |***
         .**                                       18     0              |*
         .*                                         4     0        -0.55+*
   -0.55+*                                          1     *               +----+----+----+----+----+----+----+----+----+----+
          ----+----+----+----+----+----+----+----+--                          -2        -1         0        +1        +2
          * may represent up to 16 counts
```

restriction is not significant and in fact a set of non-compatible model coefficients summed to 1.0041.

Thus, the evidence in this preliminary analysis showed that these two data sets, which had no overlap in size-class distribution, can very properly be combined to provide one compatible set of volume and taper equations. Attempts to include the effects of age and stocking as predictors to improve the precision were to no avail.

The analysis was then expanded to include all five tree populations. Initially there was some problem in sorting out and making allowance for different procedures in sectional measurement. These confounding effects have been countered as far as can be practically possible, but there could well still be some inherent variation not accounted for.

The combined variable equation was found to be a successful common form to adopt for all five forest populations. In some cases in which there was a wider spread of tree size class, the inclusion of an additional variable, h, was also statistically significant, but in the interest of comparative analyses, the form $v = b_0 + b_1 d^2h$ was retained as the basis for comparison. Indications are that this standardisation will have only a small impact on best precision for individual forest equations.

The equations for each of the individual forests is shown in Table 5.

Table 5 Combined-Variable Equations for each Individual Forest and for All Forests Combined

	b_0	b_1	r^2	Root MSE
Woodlaw	0.008 98 ± 0.002 015	0.000 027 195 ± 0.000 000 41	0.9802	0.000 19
Pukerau	0.073 71 ± 0.005 641	0.000 026 530 ± 0.000 000 69	0.9426	0.000 51
Herbert, H74	0.001 90 ± 0.001 793	0.000 020 966 ± 0.000 000 15	0.9863	0.000 29
Herbert, H54	0.035 82 ± 0.002 518	0.000 026 690 ± 0.000 000 20	0.9854	0.000 40
Longwood	0.057 31 ± 0.031 883	0.000 027 319 ± 0.000 000 44	0.9755	0.000 48
All Forests	0.008 78 ± 0.001 338	0.000 027 429 ± 0.000 000 08	0.9925	0.000 37

The similarities in the b_1 slope coefficients for individual forests and the improvements in precision evident in the overall equation representing all forests combined relative to that for the equations for each forest on its own suggested that there was likely to be little difference among regressions for individual forests. An analysis was carried out, nevertheless, testing all combinations of dummy variables to represent forests. The full equation of the form,

$$v = b_0 + b_1 d^2h + b_2 \text{ dummy1} + b_3 \text{ dummy2} + b_3 \text{ dummy 4} + b_5 \text{ dummy 5}$$

was finally narrowed down to

$$v = b_0 + b_1 d^2h + b_2 \text{ dummy 1}$$

where dummy1 represents allowance made for Longwood Forest, there being insufficient gain in precision to reflect real gains from any of the others.

Table 6 Overall Volume Equation with and without Dummy Variable

Statistic	With	Without
Intercept	0.009 55 ± 0.001 348	0.008 78 ± 0.001 338
d2h	0.000 027 272 ± 0.000 000 09	0.000 027 429 ± 0.000 000 08
dummy (= for Longwood)	0.039 172 ± 0.011 285	-
Root MSE	0.000 36	0.000 37
R2	0.9926	0.9925

The gain in statistical precision through using a dummy variable is small, but there may be some improvement in the accuracy of estimation through its use, which needed to be tested. An assessment of its likely impact in this regard is shown in Table 7 for the same tree dimensions as in Table 3. This assessment reveals that the statistically justified dummy variable equation is not appropriate for estimation in actual practice.

Table 7 Volumes in m^3 estimated by four equations

Tree Dimensions d (cm)	h (m)	Volumes (m^3) Equation 1	Equation 2	Equation 3	Equation 4
12	8	0.0401	0.0403	0.0406	0.0798
37	28	1.0789	1.0602	1.0545	1.0937
50	33	2.2366	2.2717	2.2591	2.2983
65	38	4.2171	4.4125	4.3877	4.4269

Equation 1	Existing T13 volume equation
Equation 2	v = 0.008 78 + 0.000 027 429 d^{2}h, the recommended new overall equation
Equation 3	v = 0.009 55 + 0.000 027 272 d2h, the equation for all forests except Longwood, with a dummy variable for Longwood
Equation 4	v = 0.009 55 + 0.000 027 272 d2h + 0.039 172 dummy 1, the equation for Longwood when all other data are also involved.

There is reasonable consistency in all these estimates except for the small tree size for equation 4. But this anomalous estimate is not surprising when the range of tree size collected in Longwood from a single age class was 30 to 67 cm for diameter at breast height and 26 to 39 m for height. This evaluation suggests that the actual dummy variable adjustment is not advisable.

Taper equations, compatible with the volume equations both with and without dummy variables, were prepared for all data combined. Corresponding incompatible taper equations, where the restriction that $\Sigma b_i = 1$ was lifted, were also obtained. There was no apparent gain from including polynomials of higher than order 5 in the analysis, most likely because the diameters at 0.15 m above ground were excluded. Profiles of trees from individual forests were identified separately so that tests of sub-population bias could be conducted.

The taper equations used as a basis for later discussion of the findings are shown in Tables 8 and 9. They are of the same form as shown earlier for the Woodlaw and Pukerau preliminary analysis.

Table 8 Taper Equations Compatible with Overall Volume Equation with and without Dummy Variables

	Volume Equation without Dummy				Volume Equation with Dummy			
Statistic	Coefficient	p > F	Coefficient	p > F	Coefficient	p > F	Coefficient	p > F
z_1	0.279 658	0.0001	0.368 644	0.0001	0.248 423	0.0001	0.368 894	0.0001
z_2	0.519 217	0.0437	-	-	0.586 233	0.0229	-	-
z_3	2.333 585	0.0001	3.433 495	0.0001	2.184 274	0.0001	3.426 149	0.0001
z_4	-3.997 355	0.0001	-5.002 808	0.0001	-3.855 769	0.0001	-4.990 995	0.0001
z_5	1.864 895	0.0001	2.200 669	0.0001	1.816 838	0.0001	2.195 951	0.0001
Restriction	56.296	0.0001	62.348 351	0.0001	58.655	0.0001	65.488 860	0.0001
R^2	0.9920		0.9920		0.9920		0.9919	
Root MSE	0.1335		0.1335		0.1336		0.1336	
Σb_i	1.0		1.0		1.0		1.0	

Table 8 shows that the effect of pairing with either of the two above volume equations has little effect, but that the overall equation without including a dummy variable for Longwood is slightly better. The exclusion of z_2 from the form of equation appeared to give better consistency of estimation among forests, based on various tests not shown above, and so should be preferred. But the greatest concern was the highly significant, undesirable inclusion of the compatibility restriction, which appears to be disturbing the lower profiles of Pukerau forest data in particular.

Consequently, a corresponding set of non-compatible equations was worked out and note taken of the degree of departure from compatibility in terms of $\Sigma b_i = 1$.

Table 9 Taper Equations Not Forced to be Compatible with Overall Volume Equations With and Without Dummy Variables

	Volume Equation without Dummy				Volume Equation with Dummy			
Statistic	Coefficient	p > F	Coefficient	p > F	Coefficient	p > F	Coefficient	p > F
z_1	0.358 198	0.0001	0.393 891	0.0001	0.350 225	0.0011	0.395 414	0.0001
z_2	0.202 684	0.4450	-	-	0.256 433	0.3342	-	-
z_3	2.923 383	0.0001	3.349 568	0.0001	2.798 790	0.0001	3.337 994	0.0001
z_4	-4.501 781	0.0001	-4.889 999	0.0001	-4.381 335	0.0001	-4.872 503	0.0001
z_5	2.024 816	0.0001	2.154 126	0.0001	1.983 462	0.0001	2.147 064	0.0001
R^2	0.9920		0.9920		0.9920		0.9920	
Root MSE	0.1333		0.1333		0.1334		0.1334	
Σb_i	1.0073		1.0076		1.0076		1.0080	

In all cases the sum of the coefficients lies between 1.00 and 1.01, so that the accuracy loss is less than 1 per cent through not forcing compatibility. Moreover, the residual error trends in the profiles between 2 and 5 metres appeared to be much more satisfactory for all forests, including Pukerau, as compared with the corresponding compatible equations. The evidence in Table 9 confirms also the justification for not including the z_2 term and also for not distinguishing much between the two paired volume equations.

Compatible and non-compatible taper equations have also been derived for each of the four forest regions separately. Their coefficients are shown in Tables 10 and 11 together with the recommended overall equation by way of comparison. The volume equation used in each case was the overall one without dummy variables.

Table 10 Compatible Taper Equations for Separate Forests

STATISTIC	PUKERAU	LONGWOOD	WOODLAW	OTAGO COAST	OVERALL
z_1	0.411 033	0.292 495	0.427 134	0.337 975	0.368 644
z_3	4.086 653	4.444 796	2.403 112	3.479 882	3.433 495
z_4	-6.396 116	-6.708 130	-3.357 734	-4.997 685	-5.002 808
z_5	2.898 431	2.970 840	1.527 488	2.179 828	2.200 669
Restriction	-15.264	26.497	4.248	16.617	56.295
R^2	0.9949	0.9936	0.9936	0.9918	0.9920
Root MSE	0.0900	0.1144	0.1245	0.1414	0.1335

Table 11 Taper Equations for Separate Forests, Not Compatible

STATISTIC	PUKERAU	LONGWOOD	WOODLAW	OTAGO COAST	OVERALL
z_1	0.388 499	0.371 474	0.450 732	0.351 155	0.393 891
z_3	4.149 561	4.095 650	2.346 202	3.433 693	3.349 568
z_4	-6.488 558	-6.198 429	-3.300 049	-4.935 718	-4.889 999
z_5	2.940 945	2.752 185	1.510 033	2.154 487	2.154 126
R^2	0.9950	0.9938	0.9936	0.9919	0.9920
Root MSE	0.0895	0.1124	0.1244	0.1414	0.1333
Σb_i	0.9904	1.0209	1.0069	1.0036	1.0076

Again various informal tests of prediction bias would suggest that, despite the slight gains in precision in fitting individual equations for Pukerau, Longwood and Woodlaw forests, the fit of the overall taper equation is virtually as good as the individual ones and better than the Otago Coast one. Moreover, the one and two per cent errors in not forcing compatibility for the Pukerau and Longwood forests respectively suggest that the overall equation would be more appropriate to use until further refinements to the modelling could be and need to be made.

DISCUSSION

There are obvious disadvantages in preparing separate tree volume and taper equations for a small population, particularly if there is a limited range of tree size. The inconsistencies in T232 attest to that. Precision is impaired and users are unaware of the full extent of the applicability of the models. In this 5 population study reported here, the slopes of the individual volume equations were so close to one another and the intercepts were not far enough apart, that an overall equation achieved an insignificant loss of accuracy of estimation but a very substantial gain in precision. There was little evidence, however, that further improvements could be obtained through introducing further explanatory variables such as altitude, rainfall, stocking, age and the like or even dummy variables, as shown here, to explain the residual variation. There was, in fact, little variation left to explain!

That is not to say that in other regions or for other species, major improvements could not be effected through the use of covariates. Temu (1992), for example, improved the accuracy and precision of South Island Douglas fir tree volume and taper estimation in this very way. He was able to demonstrate an improvement with a volume equation of the form:

$$v = \alpha_0 d^{(\beta_1 + Z_1\beta_2 + Z_2\beta_3)} h^{\gamma}$$

where

v = volume inside bark in m3
d = diameter at breast height outside bark in cm
h - height in m
Z_i - dummy variables for the ith region
α, β_i, γ = non-linear least-squares coefficients

The full equation tested for three regional sets of data were:

$$v = (\alpha_1 + \alpha_2 Z_1 + \alpha_3 Z_2) d^{(\beta_1 + Z_1\beta_2 + Z_2\beta_3)} h^{(\gamma_1 + Z_1\gamma_2 + Z_2\gamma_3)}$$

where $Z_1 = 1$ for Nelson, $Z_2 = 1$ for Southland, otherwise $Z_i = 0$.

Percentage biases were much reduced for all except the 5.0 cm dbhob class compared with the separate equations in T13, T120, T136 and T228 developed by FRI, while the precision through using 600 trees altogether was also greatly improved, a gain of more than 50 per cent compared with an overall equation $v = \alpha d^{\beta} h^{\gamma}$ without the dummy variables or weight, $1/d^2h$. The greatest benefit was a greatly improved balance in the distribution of residuals by region and an acceptably narrow range of -0.17 to + 0.20 m^3 in the actual residuals for a mean tree volume of around 0.52 m^3.

Temu also formulated a segmented polynomial of order 2 with two join points rather than trying to seek a single higher order polynomial to describe the tree profile. This taper equation:

$$d'^2 = \frac{v}{Kh}(\beta_1 z^2 + \beta_2 z + d\beta_3(z-\alpha_1)^2 I_1 + d\beta_4(z-\alpha_2)^2 I_2)$$

was compatible with the above volume equation. In this equation z is the relative tree height from the tip, equal to (h - h')/h, α_i are the join points such that $I_i = 1$ for $z \geq \alpha_i$ and $I_i = 0$ for $z < \alpha_i$, and d, h, v and β_i are as before. No locality or other covariate was able to explain residual variation in taper. The maximum and minimum biases in predicting mean diameter inside bark for all the data were 0.99 and -1.00 cm respectively.

Temu was able to show that major improvements over the single volume equation approach were able to be produced. As there were only two taper equations, F136 and F223, and none for either T15 or T120 volume functions, the value of an overall taper equation for South Island Douglas fir is apparent. Unfortunately, it is in a form that is not utilisable at present for use in FRI packages like MARVL, AVIS and PROD.

The Otago/Southland radiata pine study has also confirmed the advantages of pooling data to construct regional volume and taper equations. In this case, there was no evidence to recommend employing variables to account for different localities or sub-populations. The statistical significance of incorporating a dummy variable for Longwood forest represented a quirk arising directly from the restricted tree sizes included in that sample. Thus the recommendation in this case would be to use the overall volume and taper equations, respectively:

$$v = 0.008\ 779 + 0.000\ 027\ 429\ d^2h$$

where

v is volume in m^3
d is diameter at breast height outside bark in cm
h is total height in m;

$$d_k^2 = \frac{v}{(\pi / 40\ 000)h}\left[2\beta_1(\frac{h-h_k}{h}) + 4\beta_3(\frac{h-h_k}{h})^3 + 5\beta_4(\frac{h-h_k}{h})^4 + 6\beta_5\frac{h-h_k}{h})^5\right]$$

where

d_k is diameter inside bark in cm at height hk above ground
β_i are least-squares regressions for the non-compatible overall equation
$\beta_1 = 0.393\ 891$
$\beta_3 = 3.349\ 568$
$\beta_4 = -4.889\ 999$
$\beta_5 = 2.154\ 126$

The non-compatible taper equation is recommended because the residuals along the lower stem were better distributed than for the compatible ones. Nevertheless, practitioners should, if predictions with the above pair are unsatisfactory, choose which of the 16 sets of paired equations in Tables 8, 9, 10 and 11 suit their circumstances best. Whichever is chosen will definitely provide better predictions than equations for individual sub-populations as at present.

GENERAL CONCLUSIONS AND RECOMMENDATIONS

Evidence presented here suggests that pooling of existing tree sectional measurement data for the Otago and Southland regions could result in better tree volume and taper estimation for the five sub-populations. There is only a small amount of residual variation about overall equations to justify the use of one dummy variable for Longwood, but estimation in practice would not be assisted through its adoption because of the very restricted sample for that sub-population.

The analyses appear to indicate that it seems advisable to construct and utilise in practice overall regional equations, and to conduct simple statistical tests of difference between predicted and actual values in future samples within this region, before building any new separate equations. That recommendation may well apply also to other regions.

Other considerations of routine practice that appear to warrant more critical examination in the near future but which were not examined in depth here, include:

(1) recommendations on the spread and replication of tree sizes in the sample;

(2) more flexible sectional measurement procedures, particularly a choice of butt end diameter that can be representatively measured, as diameters at 0.15 m can rarely be so described;

(3) revision of bark measurement procedures;

(4) use of outside as well as inside bark volume and taper equations;

(5) catering for a wider range of volume and taper equation forms to be acceptable for use in modelling packages, particularly segmented polynomials and equations with covariates;

(6) study of other regions where revision of the consistency of volume and taper equations is of practical concern.

ACKNOWLEDGEMENTS

The assistance of James Makoni, Fraser Whyte and Richard Woollons in conducting this project is gratefully acknowledged.

REFERENCES

Eggleston, N. 1992. Comparing the range of regional tree volume and taper tables. Stand Management Cooperative Report No 26

Temu, M.J. 1992. Forecasting yield of Douglas fir in the South Island of New Zealand. Ph.D. thesis. University of Canterbury