# VOLUME EQUATIONS AND THEIR VALIDATION FOR IRRIGATED PLANTATIONS OF *EUCALYPTUS CAMALDULENSIS* IN THE HOT DESERT OF INDIA

**V. P. Tewari* & V. S. K. Kumar**

*Arid Forest Research Institute, P. O. Krishi Mandi, New Pali Road, Jodhpur 342 005, India*

*Received August 2000*

**TEWARI, V. P. & KUMAR, V. S. K. 2003. Volume equations and their validation for irrigated plantations of *Eucalyptus camaldulensis* in the hot desert of India.** Six volume equations using diameter or combining diameter with height as predictors, were compared on the basis of fit and validation statistics using data collected from *Eucalyptus camaldulensis* stands in Indira Gandhi Nahar Pariyojana (IGNP) area of Rajasthan State in India. An equation that fits very well to a data set may not necessarily be the best when applied to another data set collected from the same population. The contrasting results obtained between model fitting and validation emphasise the need for model validation as an important step in the model construction process. The combined variable equation produced the best volume estimates and hence has been recommended for use in estimating total wood volume of *E. camaldulensis* in the study area.

Key words: Fit and validation - model fitting - model construction - volume estimates

**TEWARI, V. P. & KUMAR, V. S. K. 2003. Persamaan isipadu dan pengesahan untuk ladang *Eucalyptus camaldulensis* yang diairi di gurun panas di India.** Enam persamaan isipadu menggunakan diameter atau gabungan diameter dan ketinggian sebagai peramal dibuat perbandingan berasaskan statistik padanan dan pengesahan menggunakan data yang diambil daripada dirian *Eucalyptus camaldulensis* di kawasan Indira Gandhi Nahar Pariyojana (IGNP) di Rajasthan State, India. Persamaan yang sesuai untuk satu set data mungkin tidak sesuai untuk set data lain yang diambil daripada populasi yang sama. Keputusan berlawanan ini yang diperoleh antara padanan dan pengesahan model menekankan betapa perlunya pengesahan model sebagai satu langkah penting dalam proses pembinaan model. Persamaan pemboleh ubah yang digabungkan menghasilkan anggaran isipadu terbaik dan dengan itu disyorkan untuk anggaran jumlah isipadu kayu *E. camaldulensis* di kawasan yang dikaji.

## Introduction

Volume equations play a crucial role in forest management. The importance of volume equations is indicated by the existence of numerous such equations and the constant search for their improvement. The objective of any volume equation is to provide accurate estimates with acceptable levels of local bias over the entire diameter range in the data. Equations that provide accurate predictions of volume without local bias over the entire range of diameter are one of the basic building

**E-mail: vptewari@yahoo.com*

blocks of a forest growth and yield simulation system (Bi & Hamilton 1998). Development of sound management practice is one of the major priorities of the forestry sector. In India, various volume equations and tables are constructed during forest inventories, but these equations are mostly not validated. The role of validation in examining the predictive ability of a model before its application has been stressed by various authors (Goulding 1979, Reynolds *et al.* 1981).

The aim of the present study was to develop and validate volume equations for *Eucalyptus camaldulensis* stands grown as irrigated plantations in the Indira Gandhi Nahar Pariyojana (IGNP) canal project area located in the arid parts of Rajasthan State in India.

## Materials and methods

Data used in the present study were collected from the IGNP area in Rajasthan. Plantations for the species under study covered various age groups and stand densities. Trees of different diameter classes (5 to 52 cm, class interval 5 cm) were felled and their total height (H), diameter at breast height (D) and volume (V) were recorded. Sample plots, size of 0.1 ha each, were laid out at various locations and trees within them were divided into different diameter classes. Accordingly, trees representing different diameter classes were felled on proportionate basis to have representative sample of the stand. The length of the felled tree was measured with a tape and stump height was added to get the total height. For the computation of total volume, stem and branch wood with a minimum diameter of 5 cm was considered. The volume was then calculated by dividing the stem and branches into logs of 3-m length, measuring the mid-diameters and applying Huber's formula to estimate individual log volumes. A total of 91 trees was measured from the plantations.

The data were divided into two sets by random sampling. The first data set contained 70% of the observations and was used for fitting the volume equations while the latter contained the remaining data and was used for validation. These data sets will henceforth be referred to as the fitting and validating data sets respectively. The summary statistics of these two data sets are presented in Table 1.

**Table 1** Summary statistics of the fitting and validating data sets

| Variable | Range | Mean | SD | Kurtosis | Skewness |
|---|---|---|---|---|---|
| | | | Fitting data set | | |
| D (cm) | 5.0–51.9 | 15.5 | 9.2 | 6.0325 | 2.2179 |
| H (m) | 6.6–26.6 | 14.9 | 4.7 | - 0.2667 | 0.5353 |
| Age (years) | 3.0–28.5 | 11.6 | 6.2 | 0.7742 | 1.0731 |
| Stand density (stems $ha^{-1}$) | 486–3257 | 1750.6 | 692.1 | 0.0261 | 0.3015 |
| | | | Validating data set | | |
| D (cm) | 5.7–47.0 | 16.4 | 11.0 | 1.0501 | 1.2575 |
| H (m) | 6.8–27.8 | 14.9 | 6.6 | - 0.9370 | 0.6944 |
| Age (years) | 3.0–28.5 | 13.9 | 7.3 | - 0.9745 | 0.4429 |
| Stand density (stems $ha^{-1}$) | 486–3257 | 1424.4 | 656.0 | 0.6657 | 0.4700 |

SD = standard deviation

### *Model fitting*

This study compared six volume equations (Table 2) selected from forestry literature, based on their wide applications (Spurr 1952, Loetsch *et al.* 1973, Clutter *et al.* 1983). Each model was fitted to the fitting data set. To reduce heteroscedasticity in the error structure of volume estimation, and to avoid the consequences of violating the distributional assumptions, weighted least squares regression was applied for fitting equations 1 to 4. It was not necessary for equations 5 and 6 as they were fitted with non-linear technique. The non-linear equations were fitted using SPSS software through Levenberg-Marquardt minimisation method. The convergence criterion for accepting the values of parameter estimates was taken as 1.00E-08. The weight applied for equations 1 and 4 was $1/(D^2H)^k$ while for equations 2 and 3 it was $1/(D^2)^k$. Furnival's (Furnival 1961) index of fit was used to select the best weight function value for k, ranging from 0 to 3 with an even interval of 0.05. This index (I) is based on transformed maximum likelihood values and takes the following form:

$$I = \left[\text{antilog}\left\{\sum_{i=1}^{n}\left(\log\sqrt{(X_i)^{-k}}\right)/n\right\}\right]^{-1} S$$

where

$X_i$ is $D_i^2H_i$ or $D_i^2$ as the case may be,
S is the least squares estimate of the standard error of the weighted error term, and
n = number of trees in the sample.

It provides a relative measure of the departures from linearity, normality and homoscedasticity of residuals.

The coefficient of determination ($R^2$) and the root mean square error (RMSE) were used to determine the quality of fit. For the non-linear regressions, a fit index (FI) analogous to $R^2$ in linear regression (Cornell & Berger 1987) was used which was computed as:

$$FI = 1 - \left[\sum_{i=1}^{n}\left(V_i - \hat{V}_i\right)^2 / \sum_{i=1}^{n}\left(V_i - \bar{V}\right)^2\right]$$

**Table 2** Volume equations compared in the study

| Equation type | Designation |
|---|---|
| $V = a + bD^2H$ | 1 |
| $V = a + bD^2$ | 2 |
| $V = a + bD + cD^2$ | 3 |
| $V = a + bH + cD + dD^2 + eD^2H + fDH$ | 4 |
| $V = aD^b$ | 5 |
| $V = aD^bH^c$ | 6 |

where,

$V_i$ = observed volume for tree i
$\hat{V}_i$ = predicted volume for tree i
n = number of trees in the sample, and
$\bar{V}$ = mean observed volume of n trees.

Residuals were also graphically examined to check for any trend. A rank was assigned to each equation based on each criterion (Cao *et al.* 1980). The smaller the rank the better the performance of the model. The ranks were then summed up to arrive at the final fit rank for each model, which is indicative of its performance with respect to all the criteria considered. The $R^2$ statistics were adjusted for the number of parameters used in the regression models to make the rankings more independent.

*Model validation*

For validation purpose, models can be tested in various ways. Firstly, the resampling approach may be suggested (Bi & Hamilton 1998, Bi 1999, 2000). Secondly, iterative validation procedure may be considered (Williams 1997). Thirdly, if one has two independent data sets from the same area or population, one set can be used for fitting and the other for validation or, alternatively, if one has only one data set, it can be divided into two through random procedure for the purpose. Here the last approach has been adopted. The model validation was assessed on the basis of following evaluation criteria.

Average residual or prediction bias (B):

$$B = \sum_{i=1}^{n} r_i / n$$

where $r_i$ represents the difference between the observed and predicted volume for the ith tree in the validating data set. Wilcoxon's sign and rank test (Philip 1994) was used to test bias produced by the six volume equations.

The variances:

$$\text{Var}(B) = \sum_{i=1}^{n} (r_i - B)^2 / (n-1)$$

RMSE which combines prediction bias and precision:

$$\text{RMSE} = \sqrt{(B^2 + \text{Var}(B))}$$

Ranking for each model was then assigned as described above.

## Results and discussion

### *Model fitting*

The values of model coefficients obtained by applying various equations to the fitting data set are given in Table 3. The standard errors given in the table show that all the partial regression coefficients were significant except for equation 4 wherein only the coefficient for $D^2H$ was significant. It is to be pointed out here that the standard errors for parameter estimates are not exact due to heteroscedasticity of the error terms and multicollinaerity among variables in equations 3 and 4. Also, the standard errors for the parameter estimates for equations 5 and 6 are asymptotic as these functions were fitted through nonlinear technique. The values of the power k, estimated for the weights applied on equations 1 to 4, were 1.70, 2.80, 2.15 and 1.75 respectively.

Table 4 compares the fit statistics for each of the equations used. The $R^2$ values were generally high and acceptable for all the equations while RMSE values were very low except for equations 5 and 6. Final ranking showed that equations 1 and 4 ranked first followed by equation 6. Equation 3 ranked last among the six equations used. It must be emphasised that equation 4 involved six variables, and standard errors for the partial regression coefficients of this equation (Table 3) showed that only one coefficient relating to $D^2H$ was significant. Therefore, equation 1 was preferred over equation 4.

**Table 3** Values of coefficients for different equations obtained for fitting data set

| Equation | a | b | c | d | e | f |
|---|---|---|---|---|---|---|
| 1 | - 0.00514 | 3.31E-05 | | | | |
| | (0.00043) | (4.10E-07) | | | | |
| 2 | - 0.01172 | 5.37E-04 | | | | |
| | (0.00099) | (1.75E-05) | | | | |
| 3 | 0.01718 | - 0.00716 | 0.00092 | | | |
| | (0.00475) | (0.00102) | (4.91E-05) | | | |
| 4 | - 0.00750 | 6.63E-05 | 0.00138 | 5.15E-05 | 3.46E-05 | - 4.41E-05 |
| | (0.00623) | (0.00068) | (0.00181) | - (0.00013) | (4.97E-06) | (9.65E-05) |
| 5 | 1.69E-04 | 2.41298 | | | | |
| | (1.72E-05) | (0.02675) | | | | |
| 6 | 3.16E-05 | 2.10241 | 0.89456 | | | |
| | (5.55E-06) | (0.03149) | (0.08374) | | | |

Values in parentheses give the standard error of the parameter estimates.

**Table 4** Fit statistics for volume equations for *Eucalyptus camaldulensis*

| Equation | df | $R^2$ | RMSE ($m^3$) | ΣRank | Final rank ($R_f$) |
|---|---|---|---|---|---|
| 1 | 62 | 0.991 (3) | 0.00001 (1) | 4 | 1 |
| 2 | 62 | 0.938 (6) | 0.00001 (1) | 7 | 4 |
| 3 | 61 | 0.966 (5) | 0.00005 (4) | 9 | 6 |
| 4 | 58 | 0.991 (3) | 0.00001 (1) | 4 | 1 |
| 5 | 62 | 0.995 (2) | 0.02922 (6) | 8 | 5 |
| 6 | 61 | 0.998 (1) | 0.01720 (5) | 6 | 3 |

Values in the parentheses give the ranks.

Figure 1 shows the plots of residuals (observed-predicted) against the predicted total volumes. These indicate non-randomness of residuals and residuals show some trends, at least for equations 2 and 3, which imply heteroscedasticity of the data. This is also indicative of the fact that the constant variance assumption in regression analysis is not verified (Fonweban *et al.* 1995). The plots also revealed that the least dispersion was for equations 1, 4 and 6, which is in conformity with the rankings given in Table 4. The dispersion for the larger trees was also lowest for equations 1, 4 and 6 though it was very high in the case of the other three equations. Thus, from the analysis of the fitting data, equations 1 and 4 seemed to provide the best fit to the *E. camaldulensis* data.

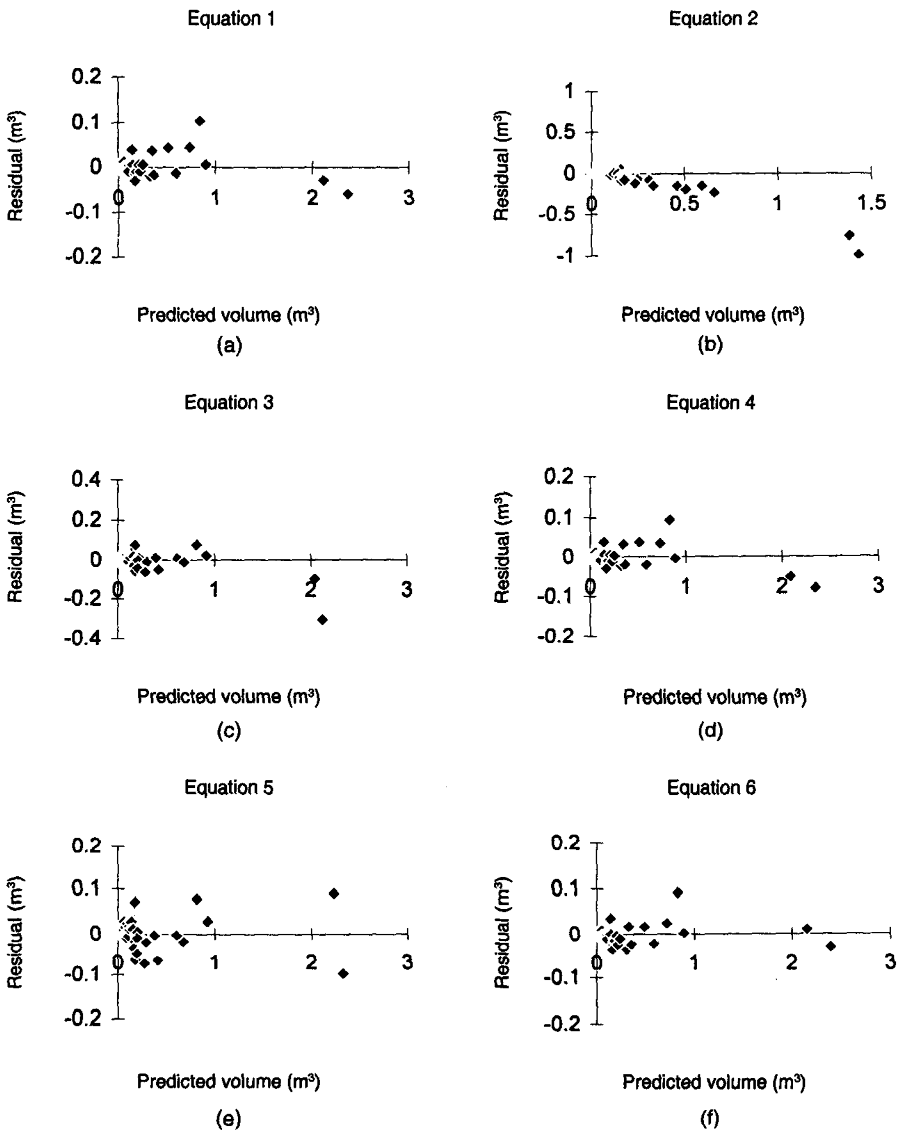


**Figure 1** Plots of residuals (observed-predicted) against predicted volume of *Eucalyptus camaldulensis* for fitting data set

### *Model validation*

The predictive ability of the different equations was assessed using an independent data set (validating data set) for model validation. The volume equations obtained from the fitting data set were applied to the validating data set. The bias gives the accuracy of prediction while the variance provides information regarding precision of the prediction. The root mean square error provides a composite measure (combining bias and precision) of the overall accuracy of prediction. The smaller these values the better the prediction. All these statistics were considered to assess the overall performance of each equation. Table 5 compares the validation statistics for the six equations used.

Equation 1 had the lowest bias, variance and RMSE while equation 2 had the maximum bias, variance and RMSE. The final ranking showed that equation 1 was the best predictor while equation 2 was the worst. Equations 2, 3 and 5 occupied the last positions for volume prediction in overall ranking. Wilcoxon's sign and rank test was used to test the significance of the bias produced by the volume equations used and results are given in Table 6. The asymptotic significance for all the equations showed that the null hypothesis of the test, i.e. the difference between sum of the positive and negative rank is zero, is accepted. Hence it may be concluded that the volume predictions by the equations were unbiased. However, in the case of equation 2, we may say that it was biased to some extent ($p = 0.08$).

**Table 5** Validation statistics for volume equations for *Eucalyptus camaldulensis*

| Equation | Bias ($\pm m^3$) | Var (B) ($\pm m^3$) | RMSE ($m^3$) | ΣRank | Final rank ($R_v$) | $R_f$ | $R_f+R_v$ | Overall Rank |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.00460 (1) | 0.0007 (1) | 0.02671 (1) | 3 | 1 | 1 | 2 | 1 |
| 2 | 0.08000 (6) | 0.0250 (6) | 0.17368 (6) | 18 | 6 | 4 | 10 | 4 |
| 3 | 0.02104 (5) | 0.0053 (4) | 0.07431 (4) | 13 | 4 | 6 | 10 | 4 |
| 4 | 0.00693 (2) | 0.0008 (2) | 0.02777 (2) | 6 | 2 | 1 | 3 | 2 |
| 5 | 0.01505 (4) | 0.0070 (5) | 0.08239 (5) | 14 | 5 | 5 | 10 | 4 |
| 6 | 0.00948 (3) | 0.0011 (3) | 0.03426 (3) | 9 | 3 | 3 | 6 | 3 |

Values in parentheses give ranks.

**Table 6** Wilcoxon's sign and rank test for validating data set

| Equation | Z | Asymptotic significance |
|---|---|---|
| 1 | $-0.120^a$ | 0.904 |
| 2 | $-1.754^b$ | 0.079 |
| 3 | $-0.288^b$ | 0.773 |
| 4 | $-0.072^a$ | 0.943 |
| 5 | $-0.769^a$ | 0.442 |
| 6 | $-1.321^b$ | 0.186 |

$^a$: based on negative rank; $^b$: based on positive rank

Figure 2 shows the plots of predicted volumes against the actual volumes of the validating data set fitted to a linear model. Predictions across the data range were generally good for all of the six equations. The intercept values vary from - 0.00550 (equation 6, Figure (2f)) to 0.01757 (equation 2, Figure (2b)). This is an indication of some prediction bias. Equation 1 produced the least dispersion of the data points followed by equations 4 and 6; this is indicative of good precision. Considering the values of the intercept and the coefficient, it can be inferred that all the equations, except equation 2, will result in underestimated volume because they have negative intercept and coefficient less than one. The value of coefficient (0.645) in equation 2 indicated 35% underestimation, which is very high, though by taking positive intercept into consideration, it was compensated to some extent. Overall, the figure indicated that volume prediction by equation 1 was closer to the observed volume in comparison with predictions by the other equations. Figure 3 shows the plot of residuals against the predicted total volumes. No clear trends were seen in the residuals. Hence, it can be said that they were randomly distributed. Equations 1 and 4 showed least dispersion; this result is in conformity with the ranking given in Table 5. Equations 5, 3 and 2 produced high residuals in the higher range of volume. This shows that predictions from these equations were less accurate in this range. In the lower range of volume, all the equations produced almost similar values of residuals. Based on these validation analyses, we conclude that equation 1 is preferred for prediction of total volume.

The above analysis shows that equation 4, which performed best in the fitting phase along with equation 1, dropped to second place during model validation. On the other hand, equation 1 remained at the first place both in fitting and validation. Equation 2, which ranked fourth in fitting, came last during model validation while equation 3, which ranked last in fitting, went up to the fourth place during validation. On the basis of quality of fit alone, equation 4 along with equation 1 can be recommended for use, which may result in less accurate volume predictions when applied to an independent data set. This emphasises the importance and need of validating a model prior to its use. The validation process is necessary so that the model can be used with some confidence (Goulding 1979, Reynolds & Chung 1986).

Equation 1, the combined variable equation, has been well recognised in volume predictions of many tree species, with $R^2$ usually above 95% (Avery & Burkhart 1994). In the present study we can also recommend the same equation (equation 1) on the grounds of both fit and validation in comparison with equation 4. Equation 4 involves more parameters than equation 1. Moreover, all the parameters in equation 4, except for $D^2H$, were not significant. Thus, equation 4 will essentially reduce to equation 1. The final equation based on pooling the fitting and validating data set (obtained through weighted least squares analysis, power $k = 1.75$) is given below:

$$\text{Volume} = -0.00226 + 0.0000333\, D^2H; \quad df = 89; \ R^2 = 0.990; \ \text{RMSE} = 0.00001$$

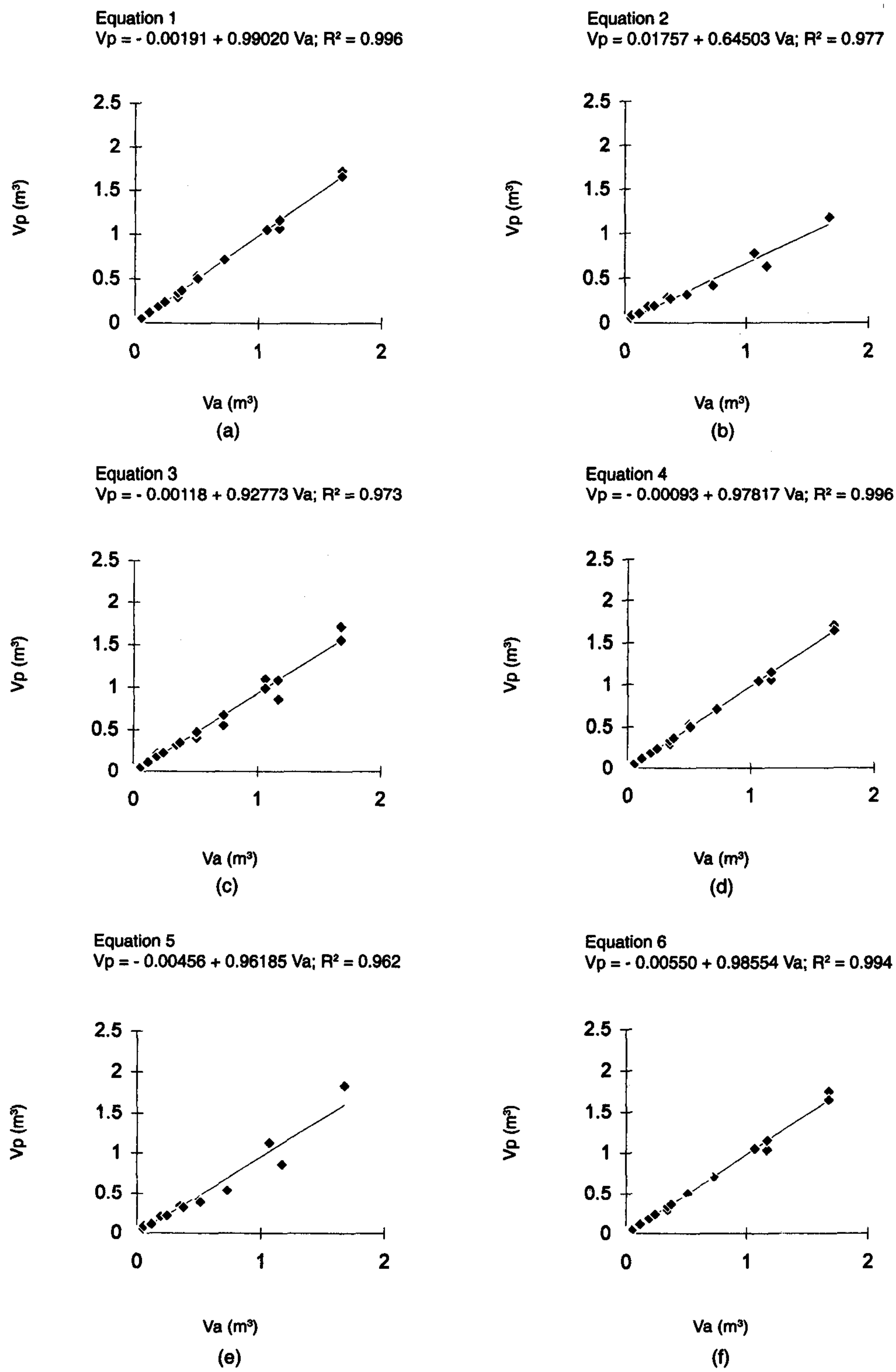


**Figure 2** Scatter data and fit of total predicted volume (Vp) against actual volume (Va) for *Eucalyptus camaldulensis*

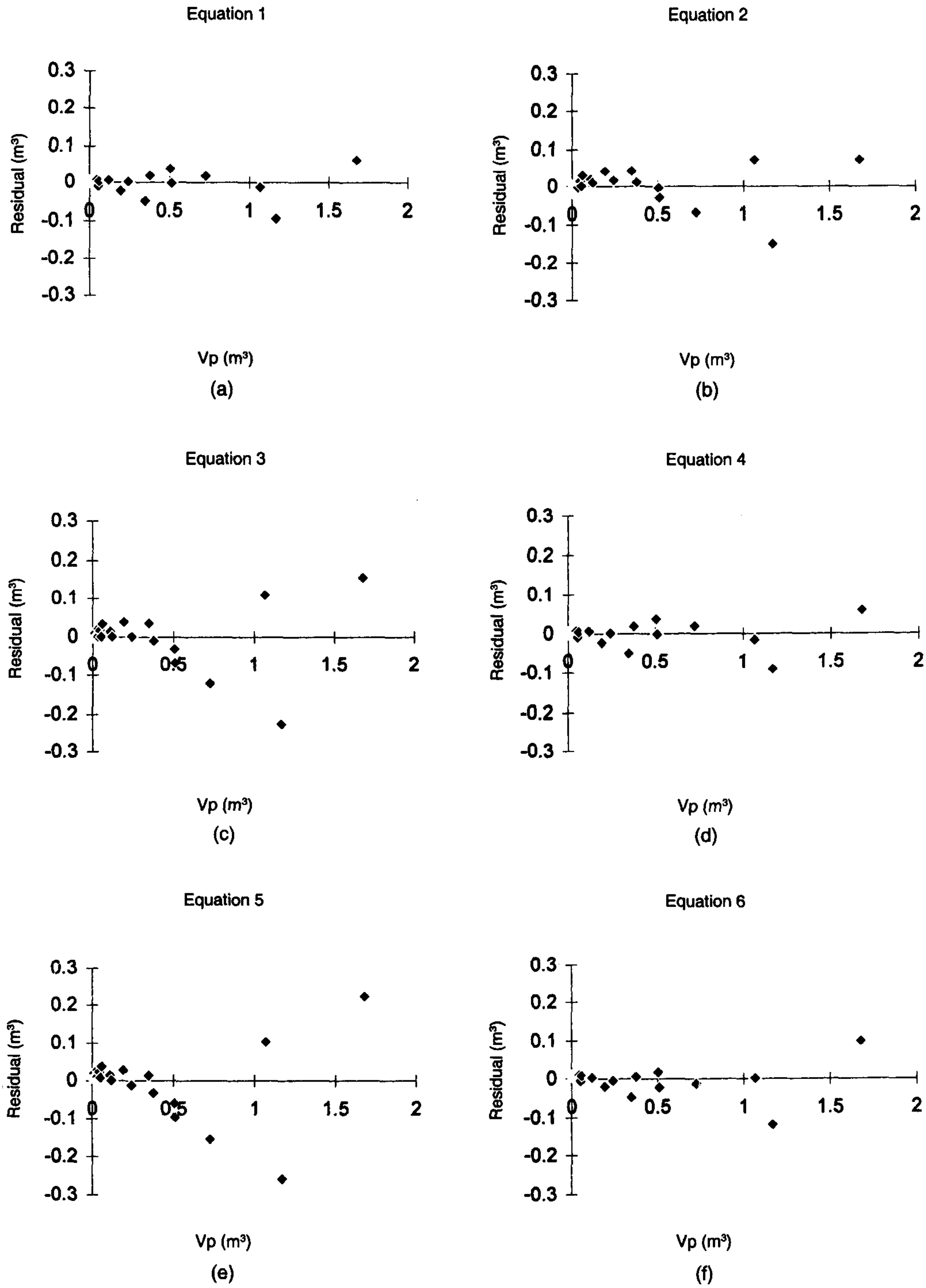


**Figure 3** Plots of residuals (observed-predicted) against predicted volume of *Eucalyptus camaldulensis* for validating data set

## Conclusions

It can be concluded from the study that the combined variable equation (model 1) performed well in both the fitting and validation process. Therefore, it can be used to predict volume for *E. camaldulensis* in the study area. The contrasting results obtained between model fitting and validation emphasise the need for model validation as an important step in the model construction process in order to get the best choices.

## Acknowledgements

The authors wish to thank the Director, AFRI for providing facilities during data collection. The guidance provided by R. C. Jain during the course of this study is gratefully acknowledged. The help rendered by B. Singh, S. L. Chouhan and J. P. Dadhich in data collection is highly appreciated.

## References

AVERY, T. E. & BUKHART, H. E. 1994. *Forest Measurements.* McGraw-Hill, New York. 408 pp.

BI, H. 1999. Predicting stem volume to any limit for native tree species in southern New South Wales and Victoria. *New Zealand Journal of Forestry Science* 20: 318–331.

BI, H. 2000. Trigonometric variable-form taper equations for Australian eucalypts. *Forest Science* 46: 397–409.

BI, H. & HAMILTON, F. 1998. Stem volume equations for native tree species in southern New South Wales and Victoria. *Australian Forestry* 61: 275–286.

CAO, Q. V., BURKHART, H. E. & MAX, T. A. 1980. Evaluation of two methods for cubic volume prediction for loblolly pine to any merchantable limit. *Forest Science* 26: 71–80.

CLUTTER, J. L., FORTSON, J. C., PIENAAR, L. V., BRISTER, G. H. & BAILEY, R. L. 1983. *Timber Management: A Quantitative Approach.* John Wiley and Sons, New York. 333 pp.

CORNELL, J. A. & BERGER, R. D. 1987. Factors that influence the value of the coefficient of determination in simple linear and nonlinear regression models. *Phytopathology* 77: 63–70.

FONWEBAN, J. N., MAYAKA, T. B. & SEUKEP, J. B. 1995. Construction and validation of tree volume equations for *Eucalyptus saligna* in Cameroon. *Commonwealth Forestry Review* 74: 355–360.

FURNIVAL, G.M. 1961. An index for comparing equations in constructing volume tables. *Forest Science* 7: 337–341.

GOULDING, C. J. 1979. Validation of growth models used in forest management. *New Zealand Journal of Forestry* 24: 108–124.

LOETSCH, F., ZOHRER, F. & HALLER, K. E. 1973. *Forest Inventory.* Volume II. BLV Verlagsgesellschaft, Munchen. 469 pp.

PHILIP, M. S. 1994. *Measuring Trees and Forests.* CAB International, Wallingford. 310 pp.

REYNOLDS, M. R. & CHUNG, J. 1986. Regression methodology for estimating model prediction error. *Canadian Journal of Forest Research* 16: 931–938.

REYNOLDS, M. R., BURKHART, H. E. & DANIELS, R. F. 1981. Procedures for statistical validation of stochastic simulation models. *Forest Science* 27: 349–364.

SPURR, S. H. 1952. *Forest Inventory.* John Wiley and Sons, New York. 476 pp.

WILLIAMS, M. S. 1997. A regression technique accounting for heteroscedastic and asymmetric errors. *Journal of Agricultural, Biological and Environmental Statistics* 2: 108–129.