Jour. Korean For. Soc. Vol. 94, No. 2, pp. 121~126 (2005)

JOURNAL OF KOREAN FOREST SOCIETY

Localizing Growth Model of *Chamaecyparis obtusa* Stands Using Dummy Variables in a Single Equation

Sang-Hyun Lee* and Hyun Kim

Faculty of Forest Science, College of Agriculture & Life Sciences, Chonbuk National University, Jeonju 561-756, Korea

Abstract : This study was carried out to construct a single diameter and a single height model that could localize *Chamaecyparis obtusa* stand grown in 3 Southern regions of Korea. Dummy variables, which convert qualitative information such as geographical regions into quantitative information by means of a coding scheme (0 or 1), were used to localize growth models. In results, modified form of Gompertz equation, $Y_2=\exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al+\beta_1 k_1+\beta_2 k_2)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))$, for diameter and height was successfully disaggregated to provide different projection equation for each of the 3 regions individually. The use of dummy variables on a single equation, therefore, provides potential capabilities for testing the justification of having different models for different sub-populations, where a number of site variables such as altitude, annual rainfall and soil type can be considered as possible variables to explain growth variation across regions.

Key words **:** *dummy variables, altitude, single equation, sub-regions*

Introduction

General systems of growth projection have often been developed to cover large geographic forest areas or regions. Examples of these include PROGNOSIS (Stage, 1973), STEMS (Belcher *et al.*, 1982), and SIDFIR (Law, 1990). However because of their broad development, some potential exists that these models will not provide adequate sensitivity of estimation for sub-regions. This is because unexplained factors within sub-regions can be averaged for the whole area, but not within sub-regions, resulting in biased estimates. When intensive forest management demands growth predictions that are sensitive at regional or sub-regional levels, then general models lose their credibility and growth models will be restricted at sub-region level (Whyte *et al.*, 1992).

Some methods have been used for localizing regional models to sub-regions. The method of stratification involves modeling each different stratum individually. To justify this hypothesis, all differences between parameters of each stratum must be conduced. Burkhart and Tennent (1977) used this method to fit site index equations for radiata pine grown in New Zealand. Other methods have been used to localize growth models. Smith (1983) used the double sampling technique of Cochran (1977), to calculate an annual adjustment factor of diameter increment of the STEMS model (Shifley and Fairweather, 1983). The Bayesian method of estimating model coefficients has also been used by Berkey (1982) and Green *et al.* (1992). Green *et al.* (1992) reported a reduction of more than 50% in residual mean squares (RMS) by simultaneously estimating Honduran pine yield equation coefficients, for sub-populations with 21 different soil site groups. Gertner (1984) used a sequential Bayesian method, which adjusted the parameter estimates through time to localize a diameter increment model taken from STEMS.

Dummy variables have also been used to localize growth and yield models. These variables are used to convert qualitative information such as geographical region and season of the year into quantitative information by means of a coding scheme (0 or 1). This method involves formulation of an analysis of covariance among regions or data sets, by representing each as a dummy variable within a single equation. Gujarat (1970) has demonstrated the general approach of this method, and Ferguson (1979) used 4 dummy variables to localize a basal area increment model for 5 different forests, mainly to represent different rainfall patterns. The use of

*Corresponding author
E-mail: leesh@chonbuk.ac.kr
This work was supported by a grant No. (R05-2002-000-00524-0) from Korea Science & Engineering Foundation

dummy variables thus provides potential capabilities for testing the justification of having different models for different sub-populations, where a number of site variables such as altitude, latitude, distance from the sea, annual rainfall and soil type can be considered as possible variables to explain growth variation across regions.

Therefore, the objective of this study was to construct a single diameter and a single height model using dummy variables that could localize *Chamaecyparis obtusa* stand grown in 3 Southern regions of Korea.

Materials and Methods

In order to derive diameter and height equations, the data were used from *Chamaecyparis obtusa* temporary plots grown in Mt. Chukryeong of Jeonnam, Mt. Munsoo of Jeonbuk and Namhae of Gyeongnam provinces. All of 60 plots, which were 20 m×20 m size each plot, were used for analysis. From the each plot, 1 dominant tree was selected and cut. After cutting the sample trees diameter and heights of the certain age were measured using the stem analysis. The basic data obtained from stem analysis were transformed into projection format of intervals between measurement time T_1 and T_2 that used to build equation. A summary of relevant plot statistics is given in Table 1.

The methods used for this study were algebraic difference equation (Borders *et al.*, 1984) that has been widely used for growth and yield modeling studies. The main standard statistical procedures used were non-linear least-squares regression based on PROC NLIN in Statistical Analysis System (SAS Inc, 1990). The derivative-free method (DUD), which was found to be best in convergence, was adopted for non-linear least-squares regression among the algorithms of PROC NLIN procedures used to estimate parameters (Ralston and Jennrich, 1979).

The PROC UNIVARIATE procedure was also used to examine the residuals and provide several statistics that are valuable for making inferences about residuals patterns. The important values utilized in the analysis of this study were mean of residuals, skewness, kurtosis and extreme values. In addition, graphical charts and plots were used to check the distributions of residuals with regard to normality of errors. Residual errors were plotted against predicted values to determine goodness of fit. Because whether or not the residual patterns lay normally about the zero reference line are of the important criterion for judging the independent distribution.

The commonly adopted height projection equations are log-reciprocal (Schumacher, 1939; Woollons and Wood, 1992), Chapman-Richards (Piennar and Turnbull, 1973; Goulding, 1979), Gompertz (Whyte and Woollons, 1990), Weibull (Yang *et al.*, 1978; Goulding and Shiley, 1979) and Hossfeld (Liu Xu, 1990). The functional forms of projection equations used are presented in Table 2.

To test for a difference in slopes, an interaction term called a slope dummy variable is forms by multiplying

Table 1. Summary of each region for *Chamaecyparis obtusa* stands.

Areas	Number of Plots	Ages (years)	DBH (cm)	Height (m)	Slope (°)	Altitude (m)
Gochang	20	28 25–30	16 14–20	13.6 11.2–15.2	10-30	620
Jangseong	20	48 42–61	24 20–28	16.2 13.4–19.4	15-25	500
Namhae	20	34 31–52	18 8–26	11.4 9.4–13.8	15-25	200

Table 2. General forms of projection equations applied to data.

Equation name	Equation Forms*
Schumacher anamorphic	$Y_2 = Y_1\exp(-\beta(1/T_2^{\gamma}-1/T_1^{\gamma}))$
Hossfeld anamorphic	$Y_2 = 1/((1/Y_1)-\beta(1/T_2^{\gamma}-1/T_1^{\gamma}))$
Chapman-Richards anamorphic	$Y_2 = Y_1((1-\exp(-\beta T_1)) / (1-\exp(-\beta T_2)))^{\gamma}$
Gompertz anamorphic	$Y_2 = Y_1\exp(-\beta(\exp(\gamma T_2)-\exp(\gamma T_1)))$
Schumacher polymorphic	$Y_2 = \exp(\ln(Y_1)(T_1/T_2)^{\beta}+\alpha(1-(T_1/T_2)^{\beta}))$
Hossfeld polymorphic	$Y_2 = 1/((1/Y_1)(T_1/T_2)^{\gamma}+(1/\alpha)(1-(T_1/T_2)^{\gamma}))$
Chapman-Richards polymorphic	$Y_2 = (\alpha/\gamma)^{[1/(1-\beta)]}(1-(1-(\gamma/\alpha)Y_1^{(1-\beta)}))(T_2 - T_1)^{[-\gamma(1-\beta)]})^{(1/(1-\beta))}$
Gompertz polymorphic	$Y_2 = \exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+\alpha(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))))$

*Y_1 = diameter and height of trees at age T_1, Y_2 = diameter and height of trees at age T_2, exp = exponential function, ln = natural logarithm, and α, β, γ are coefficients to be estimated.

the dummy times measurement variable. Regressing Y on both a measurement variable X_2 and a slope dummy variable X_1X_2 yields equation with form

$$Yi = \beta_0 + \beta_2X_{i2} + \beta_3X_{i1}X_{i2} \quad (1)$$

Where,

the slope relating X_2 to Y equals β_2 when $X_1 = 0$, the slope relating X_2 to Y equals $\beta_2 + \beta_3$ when $X_1 = 1$, β_3, the coefficient on X_1X_2, equal the difference in slopes between X_1 categories 0 and 1.

A test of $H_0: \beta_3 = 0$ determines whether the two slopes differ significantly. Although the above example applies to linear regression, the same principles are applicable to non-linear models. The dummy variable thus was adapted to test whether or not different models for different sub-populations are justified.

Results and Discussion

1. Prediction of diameter growth

Most anamorphic equations generally produced biased residuals patterns, though Schumacher and Chapman-Richards anamorphic functions proved little bit superior in statistics of residuals and residuals patterns to other anamorphic equations. The statistics of residuals of the anamorphic equations fitted are presented in Table 3 with corresponding mean square error values (MSE).

Most of the polymorphic equations generally fitted well without apparent bias in residuals pattern. Comparing residual pattern and mean square error values, the Gompertz polymorphic function, equation (2), with mean square error (MSE) 2.523 was found to represent better than the other equations. The fitted coefficients and mean square error are shown in Table 4.

Table 3. Statistics of residuals with the anamorphic equations fitted to DBH data.

Equation name	MSE	Mean of residuals	Skewness	Kurtosis
Schumacher	15.462	0.90	1.08	2.95
Chapman-Richards	15.681	0.90	1.07	2.77
Gompertz	38.265	3.28	0.35	0.45
Hossfeld	42.329	3.76	0.77	1.47

Table 4. Coefficients for polymorphic equations fitted to DBH data.

Model name	Coefficient			MSE
	α	β	γ	
Schumacher	3.0653	0.909	-	5.075
Chapman-Richards	3.517	0.233	0.258	6.805
Gompertz	3.167	0.085	0.0001	2.523
Hossfeld	23.414	-	2.405	5.377

$$D_2 = \exp(\ln(D_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2 - T_1^2)) + \alpha(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2 - T_1^2)))) \quad (2)$$

PROC UNIVARIATE in SAS with equation (2) showed that residual statistics were satisfactory as it contained -0.112 values for skewness and 0.753 values for kurtosis. A Shapiro-Wilk test for normality was totally accepted as 0.98 that is much closed to 1 of normal distribution.

Modifications to equation (2), with the addition and subtraction of predictor variable namely, altitude, which reflects largely influences of temperature, soil fertility and rainfall, was tested to effect further improvements. Equation (3) represents the inclusion of altitude (Al).

$$D_2 = \exp(\ln(D_1)\exp(-\beta(T_2 - T_1) + \gamma(T_2^2 - T_1^2)) + (\alpha + \alpha_1 Al)(1-\exp(-\beta(T_2 - T_1) + \gamma(T_2^2 - T_1^2)))) \quad (3)$$

In the results of adding altitude, MSE (2.476) of equation (3) was better than MSE (2.523) of equation (2). Table 5 presents successive improvements in which additional variable was introduced to the basic form of the Gompertz polymorphic function. And a plot of residual values against predicted values is given in Figure 1. A plot of residuals against predicted values indicated that a random pattern around zero with unbiased trend. Hence, the modified Gompertz function, which is including altitude as a predictor variable, was chosen as the best representative diameter equation.

Table 5. Summary of statistics of residual values for DBH Gompertz polymorphic equation.

Statistics name	Value		
	Before adding altitude	After adding altitude	Normal distribution
Mean	0.014	0.015	
Skewness	-0.112	-0.163	0
Kurtosis	0.753	0.734	0
W:Normal	0.980	0.983	1

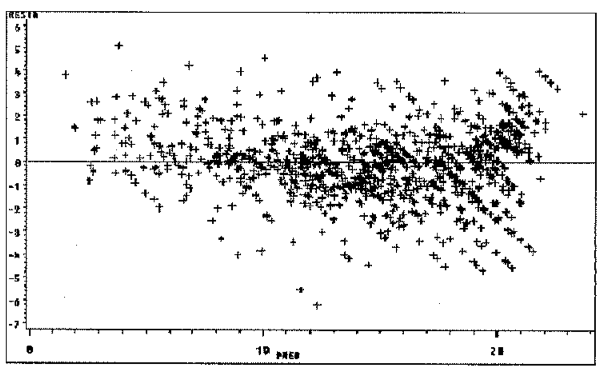

Figure 1. Plot of residual against predicted for DBH Gompertz polymorphic equation.

In order to examine the impact of locality, dummy variables (1 or 0) representing 3 separate regions, were added to equation (3). The Mt. Munsoo of Gochang in Jeonbuk region was the default locality. Equation (4) shows the form, which includes 2 dummy variables, K_1 and K_2, for Jangseong and Namhae regions, respectively.

$$D_2 = \exp(\ln(D_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al+\beta_1 k_1+\beta_2 k_2)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (4)$$

The asymptote α in equation (4) is the default dummy variable representing the Gochang region. When projections are made for Jangseong, the asymptote is $\alpha+\beta_1 K_1$, while those for Namhae is given by $\alpha+\beta_2 K_2$, respectively. The analysis proved that each of the 3 regions had substantially different asymptotes, as it showed the coefficient of dummy variables were different from zero at the $\alpha = 0.05$ provability level. Parameter estimation for altitude and dummy variables in equation (4) is given in Table 6.

2. Prediction of height growth

The same procedure, which used in modeling of diameter growth, was adopted for the height growth. Generally, the polymorphic equations represent better fitting than anamorphic with this data. The statistics of residuals of the polymorphic equations fitted are presented in Table 7 with corresponding mean square error values (MSE).

The ranking order of MSE was Gompertz > Hossfeld > Schumacher > Chapman-Richards in polymorphic equation. Thus, the Gompertz polymorphic form was selected for the best and further analysis equation with the lowest MSE of 1.563 after comparing residual statistics.

Table 6. Coefficient estimation of altitude and dummy variables for diameter equation.

Coefficient	Estimate	Standard Error	Lower 95% Confidence Level	Upper 95% Confidence Level
α_1	0.00983	0.00004	0.00975	0.00991
β_1	1.20004	0.01868	1.16338	1.2367
β_2	4.21599	0.02826	4.16053	4.27145

Table 7. Coefficients for polymorphic equation fitted to height data.

Model name	Coefficient			MSE
	α	β	γ	
Schumacher	4.262	0.507	-	3.469
Chapman-Richards	16.702	0.969	15.294	4.999
Gompertz	3.063	0.078	0.0004	1.563
Hossfeld	21.956	-	1.900	3.015

A modification to Gompertz equation, with the addition of altitude as an environmental factor, was tested to effect further improvements. Equation form (5) represents the inclusion of altitude (Al).

$$H_2 = \exp(\ln(H_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)) + (\alpha+\alpha_1 Al)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (5)$$

In the results of adding altitude, MSE(1.563) of Gompertz equation was improved to MSE(1.226) of equation (5) that had a mean residual error of 0.037 m and corresponding skewness and kurtosis values of -0.417 and 1.746, respectively. The data were evidently well balanced as 95% of residuals lay within ±1.8 m with no detectable patterns and showed goodness of fitting as shown in Figure 2.

Dummy variables were added to equation (5) to account for possible different growth patterns across regions (equation 6).

$$H_2 = \exp(\ln(H_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al+\beta_1 k_1+\beta_2 k_2)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (6)$$

The analysis showed that each of the 3 regions had substantially different asymptotes (α=0.05 provability level). Therefore, height growth projections could be made for Jangseong, the asymptote is $\alpha+\beta_1 K_1$, while those for Namhae is given by $\alpha+\beta_2 K_2$. Parameter estimation of altitude and dummy variables for height equation is given in Table 8.

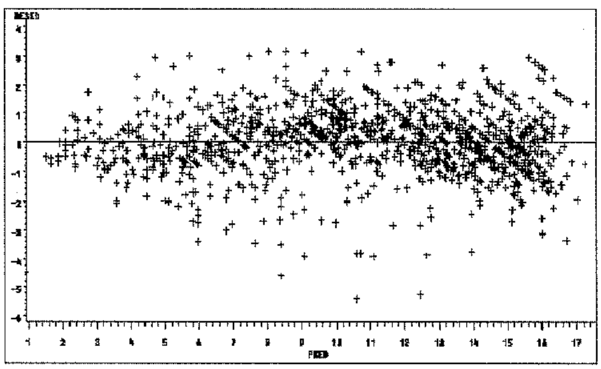

Figure 2. Plot of residual against predicted for height Gompertz polymorphic equation.

Table 8. Coefficient estimation of altitude and dummy variables for height equation.

Coefficient	Estimate	Standard Error	Lower 95% Confidence Level	Upper 95% Confidence Level
α_1	0.00983	0.00004	0.00976	0.0099
β_1	0.98806	0.01391	0.96077	1.01536
β_2	3.72689	0.0207	3.6863	3.76749

Conclusions

The dummy variable approach allowed one to use a single equation for all regions and provided that there was comparability in the growth trajectories and simply a difference in asymptotes. In this study, when modeling diameter and height predictions with modified Gompertz equations, the overall equations for diameter and height were disaggregated to provide different projection equations for each of the 3 regions individually. The following equations proved this result.

$$Y_2 = \exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al +\beta_1 k_1+\beta_2 k_2)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (7)$$

Equation (7) is applied as default and equation (8) is for projections for the Gochang region.

$$Y_2 = \exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al) (1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (8)$$

Equation (9) includes an active dummy coefficient β_1 and variable K_1 for the Jangseong region.

$$Y_2 = \exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+\alpha+\alpha_1 Al +\beta_1 k_1)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (9)$$

Similarly, equation (10) contains active β_2 and K_2 for the Namhae region.

$$Y_2 = \exp(\ln(Y_1)\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2))+(\alpha+\alpha_1 Al +\beta_2 k_2)(1-\exp(-\beta(T_2-T_1)+\gamma(T_2^2-T_1^2)))) \quad (10)$$

Site quality, genotype variability and local climatic fluctuations among different regions, can affect growth and performance of individual trees and forests at the stand level, resulting in different growth trends among regions.

Literature Cited

1. Belcher, E.M., Holdaway, M.R. and Brand, G.J. 1982. A description of STEMS: the stand tree evaluation and modelling system. USDA, Forest Service General Technical Report. NC-79:18.
2. Berkey, C.S. 1982. Bayesian approach for non-linear growth model. Biometrics 38: 953-961.
3. Burkhart, H.E. and Tennent, R.B. 1977. Site index equation for radiata pine in New Zealand. New Zealand Journal of Forest Science 7(3): 408-416.
4. Cocharn, W.G. 1977. Sampling techniques. John and Wiley and Sons, New York. 428.
5. Ferguson, I.S. and Leech, J.W. 1978. Generalized least-squares estimation of yield functions. Forest Science. 24: 27-42.
6. Gertner, G.Z. 1984. Localizing a diameter increment model with a sequential Bayesian procedure. Forest Science. 30(4): 851-864.
7. Goulding, C.J. 1979. Validation of growth models for *Pinus radiata* in New Zealand. New Zealand Journal of Forest. 24(1): 108-124.
8. Goulding, C.J. and Shirey, J.W. 1979. A method to predict the yield of log assortments for long term planning. In Elliott, D.A.(Ed). Mensuration of management planning of exotic forest plantations. New Zealand Forest Service FRI Symposium No. 20: 301-314.
9. Green, E.J., Strawdermann, W.E. and Thomas, C.E. 1992. Empirical Bayes development of Honduran pine yield models. Forest Science 38(1): 21-33.
10. Gujarat, D. 1970. Use of dummy variables in testing for equality between sets of coefficients in liner regression: a generalization. American statistician 25(4): 21-33.
11. Law, K.R.N. 1990. A growth model for Douglas fir grown in the South Island of New Zealand. FRI/Industry Research Cooperatives: Stand Growth Modelling Cooperative Report No. 18. Forest Research Institute, Rotorua.
12. Liu, Xu. 1990. Growth and yield of Douglas fir plantations in the Central North Island of New Zealand. Ph.D Thesis. School of Forestry, University of Canterbury, New Zealand. 244.
13. Pienaar, L.V. and Turnbull, K.J. 1973. The Chapman-Richards generalization of von Bertalanffy's model for basal area growth and yield in even-aged stands. Forest Science. 19: 2-22.
14. Ralston, M.L. and Jenrich, R.I. 1979. Dud: A derivative-free algorithm for nonlinear least-squares. Technometrics. 20(1): 7-14.
15. SAS Institute Inc. 1990. SAS/STAT Procedures guide. Version 6, NC.
16. SAS Institute Inc. 1990. SAS/Stat User's guide. Version 6, Cary, NC.
17. Schumacher, F.X. 1939. A new growth curve and its application to timber-yield studies. Journal of Forestry. 37: 819-820.
18. Smith, W.B. 1983. Adjusting STEMS regional forest growth model to improve local predictions. USDA, Forest Service Research. Note. NC-197.
19. Stage, A.R.1973. Prognosis model for stand development. USDA Forest Service Research Paper. INT-137: 32.
20. Whyte, A.G.D. and Woollons, R.C. 1990. Modelling stand growth of radiata pine Thinned to varying densities. Canadian Journal of Forest Research. 20: 1069-1076.
21. Whyte, A.G.D. and Woollons, R.C. 1992. Diameter distribution growth and yield modeling: recent revisions and perspectives. School of forest, University of Canterbury, New Zealand. Unpublished.
22. Whyte, A.G.D., Temu, M.J. and Woollons, R.C. 1992.

Improving yield forecasting reliability through aggregated modeling. In: Wood, G.B. and Turner (Eds), B.J. proceedings IUFRO-Integrating Information Over space and Time. Australian National University Canberra, January; 13-17, 1992: 81-88.

23. Woollons, R.C. and Wood, G.R. 1992. Improving yield forecasting reliability through aggregated modeling. In: Wood, G.B. and Turner (Eds), B.J. Proceedings IUFRO-Integrating Information Over space and Time. Australian National University Canberra, January; 13-17. 1992: 71-80.

24. Yang, R.L., Kozak, a. and Smith, H.J.G. 1978. The potential of Weibull-type functions as flexible growth functions. Canadian Journal of Forest Research. 8: 424-431.

(Received March 11, 2005; Accepted April 13, 2005)