Developing the site index equation using a generalized algebraic difference approach for *Pinus densiflora* in central region, Korea

Jungkee Pyo†
Seohaean Research Institute, Chungnam Institute, Hongsung 32258, Republic of Korea

**ABSTRACT**

The purpose of this study was to present practical approaches for developing site index equations and compare the results to those obtained from a generalized algebraic difference approach. Traditional estimation techniques for dominant height equations involve the random choice of observed growth intervals and base ages, where each combination results in a different set of parameter estimates. The traditional methods were affected by the choice of base age and measurement intervals. The base-age-invariant techniques produced identical results regardless of the choice of base ages or applied algorithm. The Chapman-Richards model was used in this study to develop site index equations for *Pinus densiflora* in the central region, Korea. The algebraic difference approach was used to derive an anamorphic base-age invariant site function that fit as a fixed base-age anamorphic site function (base age = 30 years). The developed site index equations apply to central region *Pinus densiflora* plantations ranging in age from 3 to 50 years. The new site functions represent an improvement over earlier site functions because the new method accounted for serial correlation in the data.

**Introduction**

Growth models have been used extensively to estimate site productivity (Clutter et al. 1983). The most effective system for estimation of site productivity needs to be based on identifying individual characteristics represented by age (Bailey and Clutter 1974; Cao 1993). The height of dominant and co-dominant trees at a given age is a critical component of growth and yield models. It is little affected by the stand densities that are usually encountered in specific plantations. The site quality estimation procedures based on height data are the most commonly used technique for evaluating site productivity (Bravo-oviedo et al. 2007). Therefore, the production of a given site is based on a dominant/co-dominant height observation for a study plot. Most of these height-based techniques for evaluation of site quality rely on the development of a site index. The general site index equation consists of constructing a single curve representing the average of heights at each age within the data. Site indices are a particular type of mathematical function that are defined by their value in time, which is called the reference point (Ciezewski 2002; Ciezewski and Strub 2008). In general, the most important data for the development of a site index come from permanent plots or from stem analysis. The repeated measurement or stem analysis data are usually combined for plots, resulting in an average height-age relationship (Garcia 2004; Dieguez-Aranda et al. 2005). Both of these data sources provide a number of observed data for a given location and allow the ultimate flexibility in model forms and estimation techniques. In order to reference the resulting curves by height at a given age, a choice of base age must then be made. This may suggest that the site index be included in the model prior to its parameter estimation. For this reason a general method for parameter estimation in site index models is based on the assumption that the site indices are equal to the observed height values at the base age. The site index, including base age, must be known prior to model fitting for existing equations. Since the observed height values contain measurement and sampling errors, such practice results in parameter estimates that are biased and unique to the pre-selected base age. The base-age-invariant is based on identification consistencies of individual trends in the data. This observation is subject to the measurement error associated with that data point, which in turn biases the fitting process because each curve is forced to pass an erroneous point of the measurement error, which in turn biases the shape of the curve. To avoid this problem, Bailey and Clutter (1974) suggested that all of the height-age data for each plot were actually estimated to site index parameters. Also, they introduced the concept of base-age invariant site index equations. The base-age parameter estimation using directly observed heights at specified base ages to restrict the predictions to values equal to the observed data. This procedure shows less bias than the base age method. Ciezewski (2002) present a detailed discussion of base-age invariant and fixed base-age site functions. The purpose of this study was to compare site index models for *Pinus densiflora* in the central region in Korea fitted using base-age invariant technique.

**Materials and methods**

**Study sites**

The study sites were concentrated in the central part of the country, located within the rectangle from 37°43′–37°45′N and 127°9′–127°11′W (Figure 1). The temperature ranges from 6.4 °C to 16.2 °C, with an annual mean of 12.2 °C.
Mean precipitation has been c. 1302 mm over the last 30 years.

A total of seven temporary sample plots were established in representative *Pinus densiflora* stands situated from 156 m to 216 m in altitude and from 5.0° to 35.0° slope. The trees were felled leaving stumps with a mean height of 0.2 m. According to Huber’s method, the logs were cut at 2 m intervals and the first and last logs were 1 m in length. The number of rings on the cross-sectional disc was counted at each point and then converted to stump age, which can be considered equal to plot age. Consequently, a total of 699 measured height-age data underwent stem analysis (Table 1).

### Fitting method

Cao (1993) initially defines base-age-invariant as a methodology for a fitting method that is ideal for serially correlated data. Cao’s procedure requires an estimate of height at a given age prior to the fitting process, which is a problem since height at a specific age is rarely measured in the field. He referred to this estimate of height and age as site index (S) at the base age (A0). Equation (2) was reformulated as a fixed base-age invariant to accommodate this change of variables.

This method refers to initial conditions of the collected data, which are obtained from sampling, measurements, re-measurements, or other type of inventory. Gieszewski et al. (2000) defines a self-referencing method as an algebraic difference approach. This method was used in many studies related to site index. It was extended to a generalized form (e.g. generalized algebraic difference approach, GADA) by Gieszewski. GADA consists of selecting a base equation, identifying the parameters that are related to the independent variable, and determining their functional relationship. Once the model is explicitly defined, the equation is solved for x and the variable H and t are substituted for their initial conditions H0 and T0. The selection of the parameters related to the x variable and the relationship between them should permit a closed form solution.

### Model development

The Chapman-Richards equation (Equation 1) has been applied by height/age model with GADA. The parameters of these methods were estimated using Model procedure in statistical analysis system:

\[
h \text{(age)} = b_0(1 - \exp(-b_1\text{Age}))^{b_2} \tag{1}
\]

where, \(h\text{(age)}\) = the height at age; \(b_0\), \(b_1\), \(b_2\) = parameter to be estimated using Chapman-Richards model. From Equation (1), the height at age may be estimated from any pair of age-height observations. The height (\(Hd\)) could be treated as at age (\(A_i\)), and from any height-age pair of observations can be calculated from (Equation 2):

\[
Hd_i = Hd_t \left[ 1 - \exp(-b_1A_2) \right]^{b_2} \left[ 1 - \exp(-b_1A_1) \right]^{-b_2} \tag{2}
\]

Similar derivations of difference forms of the Chapman-Richards equation may be made for the parameters \(b\) and \(c\), respectively (Equations 3, 4):

\[
Hd_i = b_0\left(1 - \exp(-\gamma A_2)\right)^{b_2} \tag{3}
\]
\[
Hd_i = b_0\left(1 - \exp(-b_1A_2)\right)^{\gamma} \tag{4}
\]

where, \(\gamma = \frac{\ln(1-\text{Hd}_i/b_0)}{A_1}\); \(\delta = \ln\left(\frac{\text{Hd}_i}{b_0} - (1 - e^{-b_1A_i})\right)\).

The three different formulations of the Chapman-Richards equation are either anamorphic with variable asymptotes (Equation 2) or polymorphic with fixed asymptotes (Equations 3, 4).

### Variable asymptote approach

To contain site-specific parameters, the approach to fitting a base-age-invariant site index equation was used to a dummy variable method, including a unique parameter per each plot during the fitting process (Equation 5). The following model was fitted:

\[
Hd_i = (S_1p_1 + S_2p_2 + \cdots + S_i p_i)[eq2 or eq3 or eq4] \tag{5}
\]

where, \(S_i\) = site-specific parameter for category \(i\); \(p_i\) = dummy variable; \(= 1\) for plot \(i; = 0\). The observed site index values at the specified base age are used as starting values for the \(S_i\).
parameters for each plot. This dataset consists of a single observation containing the starting values for each base age.

**Model evaluation**

The parameter of this site index equation were obtained from consecutive height-age pairs using Chapman-Richards model. The evaluation of the models involved the validation statistics. The site index equations were predicted height at age. Evaluation of this equation is based on the following criteria:

- **Fit index (FI):**
  \[ FI = 1 - \sum \frac{(y_i - \hat{y})^2}{\sum y_i^2}, \]

- **Root Mean Squared Error (RMSE):**
  \[ \text{RMSE} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n}}, \]

**Results**

Each equation provided highly accurate estimates of site index, with the fit index (FI) over 0.665 and root mean square error (RMSE) values lower than 1.657. The reliability of site index equations was not increased by FI (0.818 for Equation [2] versus 0.834 for Equation [4]; Table 2). The addition of successive characteristics into height-age data can improve precision of site index equations at the regional scale due to variations in environmental conditions.

Using three equations showed that the parameter estimates were highly significant. The height-age data and the coefficients and validation results are shown in Table 2. When considering the P-value, and standard error (SE), all parameters were significantly different from zero. A number of validation statistics were shown in the model fitness. Comparison of equation performance showed that Equation (4) provided a FI of 0.834 and RMSE of 1.657. Also, the other equation represents the FI with 0.818 and RMSE with 1.220. The box-plot of residual per age class did not reveal any unusual heteroscedasticity shape (Figure 3).

| Equation | Para. | Esti. | Std. Err. | P-value | FI | RMSE |
|----------|-------|-------|------------|---------|----|------|
| 2        | \( b_1 \) | 0.031 | 0.0021     | <0.0001 | 0.818 | 1.220 |
|          | \( b_2 \) | 2.027 | 0.0173     | <0.0001 |       |      |
| 3        | \( b_3 \) | 12.300 | 0.0641     | <0.0001 | 0.665 | 1.657 |
|          | \( b_4 \) | 2.680 | 0.0238     | <0.0001 |       |      |
| 4        | \( b_5 \) | 13.711 | 0.1035     | <0.0001 | 0.834 | 1.165 |
|          | \( b_6 \) | 0.054 | 0.0007     | <0.0001 |       |      |

Note: Para. = parameter value; Esti. = estimates; Std. Err. = standard error; FI = fit index; RMSE = root mean square error.

**Figure 2.** Predicted dominant/co-dominant height growth curve implied by observed data. Site index curves and residuals for the forms of the Chapman-Richards Equations (2), (3), and (4). The figure (a) was used for Equation (2); The figure (b) was used for Equation (3); The figure (c) was used for Equation (4). The number of refer to height at base age(=30 years).
The regression coefficients $a$ and $b$ were estimated by Model procedure in SAS. The coefficient values from Table 2 were used in Equation (1) to produce site curves for *Pinus densiflora* (Figure 3). These curves range in site index from 16 to 20 meter (base age = 30 years), and they apply to plantations that range from 5 to 80 years of age. The precision of the parameter estimates, e.g. standard errors of parameters, and overall validation statistics were higher for the Equation (4) than the other equations. Figure 2 shows that the height growth curve implied by Equation (3) with a base age. The observed average dominant/co-dominant heights are also shown. The following models were obtained with site indices defined at base ages. Figure 3 shows the height growth curves implied by Equations (2), (3) with a base age of 30.

The best results were obtained using a different model for each species with the lowest bias for both the overall height prediction (0.021) and the overall site index prediction (0.101). These results suggest that different models for each species based on the algebraic difference equation proposed by Nord-Larsen (2006) are likely to be successful for site index predictions in Denmark. Four different site indices were used to develop the site index curves for each species, shown in Figure 2 (10, 12, 14 m at reference age of 30 years). These curves were plotted over the observed data. For all the species the shape of the curves is very close to the shape of the observed data with realistic asymptotes and growth patterns.

**Discussion**

The site index for *Pinus densiflora* is presented in Table 1. The developed site index equations were applied to a matrix of site indices in the central region of Korea. By comparing site index equations from this study and other published literature we found that site indices were quite different from general site indices based on traditional techniques (Figure 2).

For *Pinus densiflora*, the shapes of the generalized algebraic difference approach site curves were similar to those based on the guide curve of Son et al. (1997) (Figure 3). Shapes were compared for site indices of 16 m, 18 m, and 20 m by taking the difference between the site index values of the generalized algebraic difference approach and those of the Korea Forest Service (2012). The largest differences were less than 3 m, and these occurred above 30 years of age. The shapes were dramatically different between the generalized algebraic difference approach curves and those of Palahi et al. (2004). Differences ranged from approximately 16 m to 20 m for ages greater than 30 years. Differences were not as great for younger ages. Thus, the generalized algebraic difference approach site functions seem to better capture the curve shape for older ages than the functions of Son et al. (1997). We attribute this improvement to the generalized algebraic difference approach functions capturing the effect of serial correlation in the data. Both this study and the Korea Forest Service (2012) used the Chapman-Richards (1981) model and the same dataset; however, the Korea Forest Service (2012) ignored the serial correlation of the data. Using dummy variable procedures with base-age site indices at 46, 47 and 49 years, the results were identical in all cases.

The site index equations of the Korea Forest Service (2012), which were based on *Pinus densiflora* stands in the central region, estimated about 14, 16, 18 index. The lower height of *Pinus densiflora* from this study might be affected by regional differences and environmental conditions. This indicates that trees experiencing an environment would have similar pattern of site index in spite of large regional differences between the two studies. On the other hand, the site index of this study was similar to research of pine species plantation in Mexico (Rivas et al. 2004).

In this study the correction factor does not correspond to the relative differences in estimated values between the two methods generating the site index equations. This approach promises the most accurate estimation for age and height relationships; however, the number of site index equations that need to be developed remains a challenge. Although the site index equations cannot be successfully applied to
different species, regions, or sites, we could not develop site-specific index equations for each situation. Developing generalized site index equations has been suggested as an alternative, not only for trees but also for large scale estimation of other species. We developed a generalized site index from the site-specific equations developed in this study. However, the generalized site index equation for coniferous species in Korea overestimated the height by up to 90% at the population scale for each species and by up to 50% at the administrative district level. It is not surprising that there was large variation in estimated height growth using the generalized site index equations between this study and Son et al. (1997); the differences in species composition and regional factors could explain the variation between the methods.

Son et al. (1997) observed that the scale parameter (b) reported from 279 studies seemed to have a normal distribution pattern with a mean of 1.25. Along with the empirical estimates, Clutter et al. (1983) theoretically proved that there should be a parameter (b), although their b value (2.67) was somewhat different from the range of empirical b values (~2.3, 2.4). In summary, this study supports the needs for developing site index equations. The current study is limited to the central part of Korea, only one species, and few comparison methods for site index equations, and these limitations still leave uncertainties in the correction factor, the generalized equation, and the possibility of a constant b value. Nevertheless, this study has developed a new approach to generating site index equations which has rarely been reported before. The approach used in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Conclusions

The generalized algebraic difference approach method of the Chapman-Richards function represents an improvement over the guide curve method, overall model precision is little upper and standard errors of regression coefficients are reduced for the approach in this study compared to two methods. These improvements were attributed to accounting for serial correlation in the data used to build the site index, which Son et al. (1997) ignored. Using the generalized algebraic difference approach on base-age-invariant site index equations developed in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Acknowledgments

This study was supported by the Research Fellowship of the National Institute of Forest Science in 2012.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Bailey RL, Clutter JL. 1974. Base-age invariant polymorphic site curves. Forest Sci. 20:155–159.

Bravo-Oviedo A, Río MD, Montero G. 2007. Geographic variation and parameter assessment in generalized algebraic difference site index modeling. Forest Ecol Manag. 247:107–119.

Cao QV. 1993. Estimating coefficients of base-age-invariant site index equations. Can J For Res. 23:2343–2347.

Cieszewski CJ. 2002. Comparing fixed- and variable-base-age-site equations having single versus multiple asymptotes. Forest Sci. 48:7–23.

Ciezewski CJ, Strub M. 2008. Generalized algebraic difference approach for serial correlation in the data used to build the site index, which Son et al. (1997) ignored. Using the generalized algebraic difference approach on base-age-invariant site index equations developed in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Conclusions

The generalized algebraic difference approach method of the Chapman-Richards function represents an improvement over the guide curve method, overall model precision is little upper and standard errors of regression coefficients are reduced for the approach in this study compared to two methods. These improvements were attributed to accounting for serial correlation in the data used to build the site index, which Son et al. (1997) ignored. Using the generalized algebraic difference approach on base-age-invariant site index equations developed in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Acknowledgments

This study was supported by the Research Fellowship of the National Institute of Forest Science in 2012.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Bailey RL, Clutter JL. 1974. Base-age invariant polymorphic site curves. Forest Sci. 20:155–159.

Bravo-Oviedo A, Río MD, Montero G. 2007. Geographic variation and parameter assessment in generalized algebraic difference site index modeling. Forest Ecol Manag. 247:107–119.

Cao QV. 1993. Estimating coefficients of base-age-invariant site index equations. Can J For Res. 23:2343–2347.

Cieszewski CJ. 2002. Comparing fixed- and variable-base-age-site equations having single versus multiple asymptotes. Forest Sci. 48:7–23.

Ciezewski CJ, Strub M. 2008. Generalized algebraic difference approach on base-age-invariant site index equations developed in this study. However, the generalized site index equation for coniferous species in Korea overestimated the height by up to 90% at the population scale for each species and by up to 50% at the administrative district level. It is not surprising that there was large variation in estimated height growth using the generalized site index equations between this study and Son et al. (1997); the differences in species composition and regional factors could explain the variation between the methods.

Son et al. (1997) observed that the scale parameter (b) reported from 279 studies seemed to have a normal distribution pattern with a mean of 1.25. Along with the empirical estimates, Clutter et al. (1983) theoretically proved that there should be a parameter (b), although their b value (2.67) was somewhat different from the range of empirical b values (~2.3, 2.4). In summary, this study supports the needs for developing site index equations. The current study is limited to the central part of Korea, only one species, and few comparison methods for site index equations, and these limitations still leave uncertainties in the correction factor, the generalized equation, and the possibility of a constant b value. Nevertheless, this study has developed a new approach to generating site index equations which has rarely been reported before. The approach used in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Conclusions

The generalized algebraic difference approach method of the Chapman-Richards function represents an improvement over the guide curve method, overall model precision is little upper and standard errors of regression coefficients are reduced for the approach in this study compared to two methods. These improvements were attributed to accounting for serial correlation in the data used to build the site index, which Son et al. (1997) ignored. Using the generalized algebraic difference approach on base-age-invariant site index equations developed in this study could be useful for further studies seeking to develop site index equations at the regional or national scale. Equation (3) was fit to data to produce the coefficients in Table 2 and the residual plot in Figure 3.

Acknowledgments

This study was supported by the Research Fellowship of the National Institute of Forest Science in 2012.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Bailey RL, Clutter JL. 1974. Base-age invariant polymorphic site curves. Forest Sci. 20:155–159.

Bravo-Oviedo A, Río MD, Montero G. 2007. Geographic variation and parameter assessment in generalized algebraic difference site index modeling. Forest Ecol Manag. 247:107–119.

Cao QV. 1993. Estimating coefficients of base-age-invariant site index equations. Can J For Res. 23:2343–2347.

Cieszewski CJ. 2002. Comparing fixed- and variable-base-age-site equations having single versus multiple asymptotes. Forest Sci. 48:7–23.

Ciezewski CJ, Strub M. 2008. Generalized algebraic difference approach on base-age-invariant site index equations. Can J For Res. 38:315–321.

Clutter JL, Fortson JCL, Pienaar V, Brister GH, Bailey RL. 1983. Timber management – a quantitative approach. Florida: John Wiley and Sons; p. 31-58.

Dieguez-Aranda U, Burkhart HE, Rodríguez-Soalleiro R. 2005. Modeling dominant height growth of radiate pine (Pinus radiata D.Don) plantations in north-western Spain. Forest Ecol Manag. 215:271–284.

García O. 2004. Comparing and combining stem analysis and permanent sample plot data in site index models. Forest Sci. 51:277–283.

Korea Forest Service. 2012. The volume/biomass and stand table. Daejeon: Korea Forest Service; p. 188–190.

Nord-Larsen T. 2006. Developing dynamic site index curves for European beech (Fagus sylvatica L.) in Denmark. Forest Sci. 52:173–181.

Palahi M, Tomé M, Pulkala T, Trasobares A, Montero G. 2004. Site index model for Pinus sylvestris in north-east Spain. Forest Ecol Manag. 187:35–47.

Rivas JC, González GÁ, González AD, Gadow KV. 2004. Compatible height and site index models for five pine species in El Salto, Durango (Mexico). Forest Ecol Manag. 201:145–160.

Son YM, Lee KH, Chung YG. 1997. Stand growth estimation using non-linear growth equations. J Korean For Soc. 86:135–145.