Development of ingrowth models for forest types in South Korea

Ga Hyun Moon\textsuperscript{a}, Jong Su Yim\textsuperscript{a}, Na Hyun Moon\textsuperscript{b} and Man Yong Shin\textsuperscript{b}

\textsuperscript{a}Division of Forest Industry Research, National Institute of Forest Science, Seoul, South Korea; \textsuperscript{b}Department of Forest, Environment, and System, Kookmin University, Seoul, South Korea

**ABSTRACT**
Understanding of stand growth information is necessary for establishing forest management plans, but accurate models for estimating ingrowth are currently lacking in Korea. This research aims to develop an ingrowth estimation equation according to various forest types using nationwide forest monitoring data by the National Forest Inventory (NFI). A two-stage approach was developed based on the ingrowth database using permanent sample plots from the 5th (2006–2010) and 6th (2011–2015) NFI. In the first stage, the ingrowth probability was estimated using a logistic function. In the second stage, the ingrowth amount was estimated using a conditional function by regression analysis. In results, a logistic regression model based on the number of sampling plot which did not result in ingrowth (Model VI), was selected for an ingrowth probability estimation equation. After performing three types of statistical test to evaluate the ingrowth estimation equation suitability, three optimal models were selected based on their respective estimation ability: Coniferous Forest (Model IV), Broad-leaved Forest (Model VII), and Mixed Forest (Model VI). The estimation ability of the proposed estimation equation was statistically verified and showed no problems of suitability or applicability. If high-quality data are continuously accumulated for comparison and contrast with the present sampling plot data through the ongoing NFI system, this research can present a new direction in ingrowth modeling for Korean forests.

1. Introduction

Studies on forest stand growth and harvest model development are required to establish a forest management plan that conforms to the forest policy paradigm. Prediction of the future stand structure greatly vary, even within the same species, according to location and stand management practices (Sung et al. 2014). Moreover, the longer the estimation period, the lower the model accuracy, which further complicates the process (Lootens et al. 1999). However, more accurate estimates of the future stand state are possible by accruing information on stand conditions, such as current stand status, ingrowth development, or measurements that would enable ingrowth estimation. In particular, there is a need for the measurement and consistent monitoring of ingrowth which functions as a significant mechanism in forming secondary growth forests. This accurate assessment and estimation of forest resources facilitate the sustainable forest management.

In other countries, experiments on ingrowth are more common as part of the growth and harvest model development. Many studies have attempted to quantify the variations of ingrowth caused by environmental changes using models (Hamilton and Brickell 1983; Vanclay 1989; Adame et al. 2010; Li et al. 2011; Klopcic and Boncina 2012). A previous research aimed to produce highly accurate estimates of ingrowth by analyzing the environmental factors affecting different species and developing a model equation with diverse stand variables. In this process, various models based on two-stage modeling technique proposed by previous researchers were developed. For example, Vanclay (1994) conducted a study on the forest ingrowth and development of harvest models in the Queensland rainforests on regeneration and ingrowth. Vanclay noted that the concept of ingrowth, which refers to individual trees reaching a particular size through ingrowth, and the concept of regeneration, which refers to the process of reforestation through natural or anthropogenic methods are closely related.

In reality, however, regeneration can be highly variable depending on conditions such as forest stand characteristics, survey period and climate; when researching a particular tree species, regeneration may occur for a specific period of time; in other cases, it may not. Ingrowth is a result of past patterns of natural regeneration and is then correlated (Shifley 1990; Qin 1998), and the ingrowth is also likely either to occur or not to do during a specific growth period. Not least because using a single equation makes it hard to identify the ingrowth occurrence, it should be...
preceded to examine the occurrence probability during a particular period. That is, not until the result from the initial processing stage is used does the next stage proceed in which the amount of occurrence only for those stands with ingrowth is estimated. In such cases, ingrowth model can be developed in two stages: in the first stage, a logistic can be used to estimate the probability that the ingrowth of the target species will occur, and in the second stage, a regression equation can be used to develop a model for estimating the ingrowth rate under the conditions in which the ingrowth occurs. Various studies have been carried out so far by considering the ingrowth as a stochastic event for the two outcomes and applying a two-stage model development approach (Qin 1998; Brovo et al. 2008; Adame et al. 2010; Li et al. 2011; Klopcic and Boncina 2012; Yang and Huang 2015). Previous researches on the development of the ingrowth model were mainly based on permanent sample plots or National Forest Inventory (NFI) sample plots where the survey data were accumulated over a long period of time. Ingrowth model, in particular, is part of the individual tree growth model, where the ingrowth of each individual tree is estimated in various ways depending on its unique ingrowth conditions (Martin 1982); therefore, in such cases, research data from permanent sample plots that are surveyed repetitively over a given period of time is required to determine the location of each individual tree.

In Korea, research in ingrowth estimation involves various factors, such as ingrowth amount, mortality rate, and cutting amount yield, to determine stand growth (Forestry Research Institute 1996); however, research on ingrowth modeling is relatively insufficient. An immense amount of data collection and an analysis of permanent sampling plots are required to establish ingrowth models and estimation equations (Allen 1993). Unfortunately, both concrete research mechanisms and permanent sampling plot data are lacking. With the completion of the 5th (2006–2010) and 6th (2011–2015) NFI in Korea, the basic infrastructure enabling the estimation of future stand growth has been established through the collection of forest resource monitoring data. The NFI data enabling the analysis of future stand growth using permanent sampling plots at regular intervals are managed in the form of a database, and can be applied to different forms of data for diverse purposes. Furthermore, monitoring data on temporal changes in the forest ecosystem have become the basis for developing ingrowth estimation equations for accurately determining changes in stand growth (Gillis et al. 2005).

This research aims to determine the number of ingrowth trees and stock growth by forest types as well as the environmental factors affecting ingrowth to develop an estimation equation for calculating ingrowth according to forest type. Research by forest type has been the norm in Korea regarding development of a growth model at national level. In accordance with this practice, in this paper, the classification by forest type was firstly made concerning the study on ingrowth based on the NFI data examined at national level. This study employs the 5th and 6th NFI data of permanent sampling plots. This objective is achieved by applying a two-stage approach: we estimate the ingrowth probability using the logistic function, then estimate the ingrowth amount using the conditional function through a regular regression analysis.

2. Materials and methods

2.1. Data collection

This study limits its spatial scope to 4270 permanent forest sampling plots arranged by systematic sampling at 4 km intervals during the 5th NFI, which was conducted for forests throughout Korea (Figure 1). We divide the research period into 5-year interval using the re-analyzed monitoring data of the 6th NFI and aim to develop an equation that estimates ingrowth in the stand throughout the research period (National Institute of Forest Science 2016; Yim et al. 2016). By analyzing the status of the permanent sampling plots from the 5th and 6th research sites with the acquired annual NFI data, the total number of sampling plots of all forest types is categorized as follows: 2585 Coniferous Forest, 1364 Broad-leaved Forest, and 864 Mixed Forest (Table 1). The classification of on-site forest types in the NFI to select sample trees is as follows: the coniferous forest is where the basal area of a coniferous tree species is equal to or greater than 75%, where that of a bread-leaved tree species is equal to or greater than 75% is categorized as the bread-leaved forests, and the mixed forest indicates where each of the basal areas of the two kinds of trees species is more than 25% and less than 75%.

This study aims to determine the ingrowth (i.e. volume of trees) in Korea, which has not been previously observed because it was considered to not have exceeded some measures of size limit (e.g. 6 cmdbh), through stand growth data. That is, the 6th NFI data did consider, as the ingrowth, trees which were newly included through growth (e.g. those with 6–8 cmdbh) while they had not been contained in the 5th one because of their dbh lower than 6 cm at the time of measurements. However, trees have been shown to periodically grow to the minimum diameter class during the forest inventory period every 5 or 10 years (Beers 1962; Shifley 1993; Vanclay 2003). Assuming that the inverse relationship between the stand density and ingrowth amount is maintained, we exclude outlier data on the same sampling plots by forest type.

2.2. Model development

Based on the ingrowth database established using the NFI data, this study developed an estimation equation for ingrowth generation according to forest type using a two-stage approach. For this purpose, we examined the major ingrowth models from the previous research to which the Korean NFI data can be applied. We then selected the final candidate models for the ingrowth estimation. Reflecting both possibilities of ingrowth
generated or ingrowth not generated (Klopcic and Boncina 2012), each candidate model was developed through the two-stage approach that satisfied the statistical premise.

2.2.1. Ingrowth probability

The six candidate models used in the first stage of developing the ingrowth probability estimation equation are in the form of a logistic regression model that estimates the probability through a combination of various stand variables. The dependent variables of each candidate model show the occurrence and nonoccurrence of ingrowth in binomial form (i.e. between 0 and 1). Equation 1 is the result of converting this into a probability function through logit transformation.

\[
 p_i = \logit^{-1}(\beta X) = \frac{1}{1 + e^{-\beta X}} \tag{1}
\]

where \( p_i \) is the ingrowth occurrence probability, and \( \beta \) is the regression coefficient to estimate.

In the first stage, each model used the same form of logistic regression for the common goal of ingrowth probability estimation (Table 2); however, different independent variables were applied based on the data characteristics to develop the estimation equation (Vanclay 1989; Qin 1998; Brovo et al. 2008; Adame et al. 2010; Li et al. 2011; Yang and Huang 2015). We used the Nlin Procedure module of SAS program for this purpose (SAS Institute Inc 2004).

The model I employed such independent variables as the site index (SI) and the basal area ha\(^{-1}\) (BA) of stands, while the model II was deployed to estimate the ingrowth probability by analyzing a relatively simple stand variable of quadratic mean diameter (\( D_q \)) and stand average height (\( H \)). The model III was then devised for the ingrowth modelling which is compatible with the Spanish NFI data, and the model IV was developed to project the ingrowth for the boreal mixed species stands in Alberta, Canada. Each model was exploited to explain the patterns by which the amount of ingrowth had naturally occurred through consideration of target trees and selection of relevant variables. In the model V the fact was taken into account that, when the equation was devised, the higher the measurement interval length of the sampling plot (\( L \)) and the number of samples are, the larger the ingrowth probability is. Accordingly, the index (\( \omega \)) was utilized to indicate what impacts the two variables have as time passed. Lastly, the model VI included the independent variables which show stand density in order to develop a model to estimate the frequency and annual occurrence of ingrowth.
The model’s suitability of the logistic regression analysis can be evaluated herein using the root mean square error (RMSE) examining the difference between the values of the actual data and the estimated data gathered from the estimation model (Burnham and Anderson 2004; Bewick et al. 2005). The RMSE that can quantitatively calculate the error of the estimated value of the ingrowth occurrence probability can be determined as follows using Equation (2):

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} \]  

(2)

where \( n \) is the number of materials used; \( Y_i \) is the measured value from the permanent sampling plot; and \( \hat{Y}_i \) is the value estimated from the estimation equation. We compared the calculated RMSE values from each candidate model’s estimation equation to develop the final ingrowth probability estimation equation. The ingrowth probability estimation with the lowest value was finally selected.

### 2.2.2. Ingrowth amount

In the second stage, the ingrowth amount was estimated through a regression analysis of only the sampling plots where ingrowth occurred (Table 3). This study modified a previous ingrowth amount estimation equation by using the number of trees \( ha^{-1} \) as the dependent variable and obtained (Qin 1998; Shin et al. 2003; Adame et al. 2010; Yang and Huang 2015) a total of seven candidate models.

In the model I, the number of recruits was estimated by the linear function of the SI and BA. The model II did then use the equation whose independent variables include the reciprocal of a squared value stand average diameter at breast height (DBH) and the number of trees \( ha^{-1} \) (N) in the form of the natural logarithm. The model III and IV were selected as the candidates because both are found to be appropriate by the study of Qin (1998) estimating the annual amount of ingrowth of each species for the trees with more than 9.1 cm of diameter at breast height. The other models contain such independent variables as BA, \( D_p \), N, SI, SC1, all of which were used either respectively or as interactive forms.

After estimating the ingrowth probability, we selected a candidate model and proposed an equation for estimating the ingrowth amount. In this process, the fit index (FI) was applied, which has the advantage of evaluating model fitness while not being sensitive to the sample size (McCullage and Nelder 1986; Gillis et al. 2005). The explanation ability of each candidate model was preliminarily evaluated based on the value calculated from Equation (3):

\[ FI = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} \]  

(3)

where \( Y_i \) is the observed value used in modeling; \( Y_i \) is the value estimated from the ingrowth candidate model; and \( \bar{Y} \) is the average of the values estimated from the ingrowth candidate models. The FI used to verify the model suitability cannot easily determine the explanation ability for the difference between the experimental and estimated values because the independent variables included in the candidate models are all different from each other and calculated in a modified form during the model development (Shin et al. 2003). Thus, considering the calculation results of the three types of evaluation statistics, we prioritized the candidate models and selected the one with the smallest sum of ranks in the test statistics as the optimal model (Shin 1990; Aratabtzis and Burkhart 1992; Lee 2003). The first test statistics, the model’s mean difference (MD), demonstrates the bias of the calculated estimated value in comparison with the experimental data (Equation (4)), whereas the second test statistics, the model’s standard deviation of difference (SDD), evaluates the precision of the ingrowth estimation equation (Equation (5)). The last test statistics is the standard error of difference (SED) of the model in the form of the square root of the mean squared error (MSE), as shown in Equation (6).

\[ MD = \frac{1}{n} \sum_{i=1}^{n} c_i \]  

(4)
Table 4. Estimated parameters and root mean square errors (RMSEs) for ingrowth probability models by forest type.

| Forest type       | Model no. | \(b_0\)  | \(b_1\)  | \(b_2\)  | \(b_3\)  | \(b_4\)  | \(b_5\)  | RMSE  |
|-------------------|-----------|---------|---------|---------|---------|---------|---------|-------|
| Coniferous Forest | I         | -2.2557 | 0.0016  | 0.3977  | -       | -       | -       | 0.34  |
|                   | II        | -1.7217 | 0.1077  | 0.1140  | -       | -       | -       | 0.34  |
|                   | III       | -1.3530 | 0.0412  | 0.0236  | -0.2236 | -0.0005 | -       | 0.33  |
|                   | IV        | -0.8575 | 0.1423  | -0.0024 | -0.0482 | -0.0181 | 1.0125  | 0.34  |
|                   | V         | 10.2682 | 0.0010  | 0.1262  | 0.0032  | 0.0664  | -3.6200 | 0.34  |
|                   | VI        | -5.5621 | 0.0036  | -5.8090 | 0.0002  | 0.0164  | 0.6384  | 0.31  |
| Broad-leaved Forest | I         | -1.3661 | 0.0256  | 0.2108  | -       | -       | -       | 0.39  |
|                   | II        | -1.7254 | 0.0287  | 0.2098  | -       | -       | -       | 0.39  |
|                   | III       | -1.0904 | 0.0352  | 0.0275  | 0.1444  | -0.0003 | -       | 0.39  |
|                   | IV        | 0.6880  | 0.1029  | -0.0013 | -1.4441 | 0.0907  | 0.5239  | 0.39  |
|                   | V         | 10.1598 | 0.0101  | 0.0656  | -0.7113 | 0.5211  | -3.2894 | 0.39  |
|                   | VI        | -9.7031 | -0.0013 | -0.4574 | 0.0006  | 0.0898  | 1.6965  | 0.34  |
| Mixed Forest      | I         | -0.7254 | 0.0305  | 0.0180  | -       | -       | -       | 0.49  |
|                   | II        | -1.8141 | 0.2693  | -0.2722 | -       | -       | -       | 0.41  |
|                   | III       | -5.8583 | 0.3562  | -0.0679 | -0.1180 | 0.0015  | -       | 0.40  |
|                   | IV        | 12.0020 | 0.0165  | -0.0013 | -17.8740 | 0.0315  | 0.5663  | 0.34  |
|                   | V         | 2.4360  | -0.0105 | 0.1020  | -2.0260 | -2.7752 | -0.0597 | 0.35  |
|                   | VI        | -5.7086 | -0.0089 | -4.3589 | 0.0005  | 0.1391  | 0.7503  | 0.31  |

\[
SDD = \sqrt{\frac{e_n}{n-1}} \quad (5) \\
SE_D = \sqrt{\frac{S_n}{n} + M^2} \quad (6)
\]

\(e_n\) is the difference between the experimental and estimated values from the model, while \(n\) denotes the number of trees researched. The SDD of the model applies formula \(V_{e} = \left(\sum e_i^2 - \sum e_i^2 \right) / (n-1)\), which takes the form of the square root of the model variance \(V_{e}\).

### 2.3. Model evaluation

We evaluated the estimation ability of the equation within diverse stands to examine the suitability of the finalized ingrowth estimation equation. First, we estimated the ingrowth probability per sampling plot to examine the suitability of the final ingrowth probability estimation equation, applying the NFI data to the sampling plots used in modeling. Second, we referred to the yield table per species (National Institute of Forest Science 2018) to suggest three stand conditions. We finally verified the suitability of the ingrowth amount estimation equation developed for each stand condition. The stand conditions were categorized into stand ages of 15, 30, and 50 years and by a site index of good, fair, and poor. The first stand condition is the stand variable with stand age of 15 years and the lowest site status ("poor"). Likewise the stand condition "fair" is stand age of 30 years and stand condition "good" is stand age of 50 years.

### 3. Results

#### 3.1. Modeling the ingrowth probability

Table 4 presents the results of estimating the regression coefficient by logistic regression after applying the NFI data from the permanent sampling plots to six candidate models. These results signified that for research with multiple candidate models, the smaller the RMSE value of a model, the higher the estimation ability of the model. Thus, the ingrowth probability estimation equation developed from Model VI, which has the smallest RMSE value, is the optimal equation. Model VI, which is based on the model suggested by Li et al. (2011), is a logistic regression model for estimating the ingrowth occurrence probability \(P_i\) using the independent variables of basal area \(\text{ha}^{-1} (BA)\), number of trees \(\text{ha}^{-1} (N)\), ratio of target species basal area \(\text{SC}_t\), site index \(\text{SI}\), and the minimum diameter at breast height \(D_{\text{min}}\). Model VI requires information regarding the research period of the permanent sampling plots to estimate the ingrowth; hence, we used a research period of 5 years, which is the inventory period of the NFI sampling plots, to finalize the ingrowth probability estimation equation according to the forest type.

Furthermore, we estimated the ingrowth probability under different stand conditions by applying the finalized ingrowth probability estimation equation to the actual sampling plot data (Table 5). When applying the ingrowth probability estimation equation to the actual sampling plots, the estimation ability is considered to be excellent when the sampling plots with ingrowth occurrence provide an estimated value close to 1 and when the sampling plots without ingrowth occurrence provide a value close to 0. This proves the suitability of the estimation equation. In conclusion, the proposed ingrowth probability estimation equation effectively reflects the characteristics of the stands and provides relatively accurate data on ingrowth probability.

#### 3.2. Modeling the ingrowth amount

After estimating the ingrowth probability to develop the ingrowth estimation equation, we developed the ingrowth amount estimation equation only for the sampling plots exhibiting ingrowth, with the number of ingrowth trees as the dependent variable. Table 6 shows the results of estimating the regression coefficients per model, applying the sampling plot data per forest type to the seven candidate models. The estimation equation developed herein included stand variables signifying the stand density, such as \(D_{\text{min}}, \text{BA}, \text{SI}\), but excludes explanatory variables reflecting characteristics, such as climate, soil, and disturbance. Thus, it
has limits for modeling the interaction between the environmental factors related to climate and soil and trees with the ingrowth that newly occurred (Fourcaud et al. 2007). This weakens the model explanation ability. Climate change affects forests and ecological patterns through disturbances caused by changes in vegetation regions (Sousa 1984; Dale et al. 2000; Dawson et al. 2011; Seidl et al. 2011); hence, such environmental factors must be considered and included in estimation equations as the explanatory variable.

Moreover, the FI used to clarify the relationship between the experimental and estimated values in the final estimation equation according to forest type indicates that higher values represent a higher explanation ability of the model for the data used (Son et al. 2002; Gerhard et al. 2017). It shows a satisfactory quality overall but should further improve the explanation ability. The result of calculating three types of test statistics to evaluate the suitability of the estimation equation for the number of ingrowth trees per forest type showed that Model IV for Coniferous Forest, Model VII for Broad-leaved Forest, and Model VI for Mixed Forest have the best explanation ability.

Regarding the coniferous forest, it was the model VII that showed the best results in the model’s mean difference (MD), whereas the model IV revealed the lowest value in the model’s standard deviation of difference (SDD) and the standard error of difference (SED). As for the broad-leaved forest, it is found out that the model I did indicate the smallest errors in the model’s mean difference (MD), while the model’s standard deviation of difference (SDD) saw the model VII best fit the data. With regard to the mixed forest, firstly, the model V was the best in the model’s mean difference (MD) but the model VI turned out to be superior to the others in the model’s deviation of difference, thereby identifying no consistent trends by forest type or evaluation statistics.

Finally, we evaluated the explanation ability of each estimation equation for the number of ingrowth trees to evaluate the suitability of the ingrowth amount estimation equation by applying three different stand conditions (Table 7). The results showed that, regardless of forest type, ingrowth is observed more often in stand conditions characterized by a young tree age and a low site index, but observed less often in conditions of older tree age and high site index, such as Stand condition 3.

Therefore, the estimation equation of the number of ingrowth trees per forest type has a relatively satisfactory explanation ability because it accurately reflects the stand conditions. Moreover, according to the define of site index, stands with a higher site index grow faster because they have more space for trees to grow and less space for ingrowth to occur (Byun et al. 1996), which agrees with the results of this research.

| Table 5. Statistics of ingrowth probability estimated using ingrowth probability equations by forest type. |
| --- |
| Forest type | Stand variables | No. of plots | Mean | SD | Min | Max |
| Coniferous Forest With ingrowth | 227.9786 | 0.6746 | 0.0305 | 0.0014 | 0.0045 | 0.042 |
| Coniferous Forest Without ingrowth | 280.8 | 0.5165 | 1.5071 | 0.0370 | 0.0967 | 0.1012 |
| Broad-leaved Forest With ingrowth | 4.0709 | 1.0865 | 0.0353 | 0.0030 | 0.0994 | 0.064 |
| Broad-leaved Forest Without ingrowth | 4.2188 | 0.0363 | 0.0163 | 0.0000 | 0.0967 | 0.064 |
| Mixed Forest With ingrowth | 3.8021 | 0.0515 | 0.0036 | 0.0052 | 0.0288 | 0.066 |
| Mixed Forest Without ingrowth | 6.6854 | 0.0709 | 0.0050 | 0.0088 | 0.0810 | 0.0061 |

| Table 6. Estimated parameters and evaluation statistics for the ingrowth amount models by forest type. |
| --- |
| Forest type | Model No. | Estimates of the regression coefficients | Evaluation Statistics |
| Coniferous Forest | I | 227.9786 | 0.6746 | 0.0305 | 0.0014 | 0.0045 | 0.042 | 1.76 | 41.59 | 41.61 |
| Coniferous Forest | II | 280.8 | 0.5165 | 1.5071 | 0.0370 | 0.0967 | 0.1012 | 0.064 | 0.51 | 37.49 | 37.50 |
| Coniferous Forest | III | 4.0709 | 1.0865 | 0.0353 | 0.0030 | 0.0994 | 0.064 | 0.51 | 37.49 | 37.50 |
| Coniferous Forest | IV | 4.2188 | 0.0363 | 0.0163 | 0.0000 | 0.0967 | 0.064 | 0.51 | 37.49 | 37.50 |
| Coniferous Forest | V | 3.8021 | 0.0515 | 0.0036 | 0.0052 | 0.0288 | 0.066 | 0.51 | 37.49 | 37.50 |
| Coniferous Forest | VI | 6.6854 | 0.0709 | 0.0050 | 0.0088 | 0.0810 | 0.0061 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | I | 289.5019 | 0.6967 | 0.0364 | 0.0051 | 0.0247 | 0.062 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | II | –0.2741 | 0.5486 | 0.0860 | 0.0002 | 0.0002 | 0.062 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | III | 217.8000 | 0.0000 | 0.0312 | 0.0309 | 0.0046 | 0.067 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | IV | 0.0000 | 1.6133 | 0.0360 | 1.8766 | 0.0300 | 0.067 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | V | 2.5806 | 0.0490 | 0.0546 | 0.1297 | 0.0900 | 0.067 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | VI | 5.1459 | 0.0088 | 0.1229 | 2.4998 | 0.0414 | 0.067 | 0.51 | 37.49 | 37.50 |
| Broad-leaved Forest | VII | 3.3037 | 0.0391 | 0.0310 | 5.2399 | 0.0215 | 0.068 | 0.51 | 37.49 | 37.50 |
| Mixed Forest | I | 176.6465 | 0.8467 | 3.6772 | – | – | – | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | II | 1.6843 | 0.3629 | 4.0872 | – | – | – | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | III | 28.5452 | 0.0005 | 0.0329 | 0.0607 | 0.0027 | 0.062 | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | IV | 0.0052 | 1.1786 | 0.0444 | 2.6638 | 0.0002 | 0.067 | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | V | 1.6363 | 0.0594 | 0.0660 | 0.0605 | 0.0810 | 0.067 | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | VI | 4.3994 | 0.0291 | 0.0217 | 2.2310 | 0.0440 | 0.068 | 0.41 | 0.28 | 44.86 | 44.88 |
| Mixed Forest | VII | 4.6187 | 0.0050 | 0.0989 | 1.7531 | – | – | 0.41 | 0.28 | 44.86 | 44.88 |

4. Discussion
The changes in ingrowth conditional on stand for ingrowth modeling is due to the possibility that
The two-stage approach exploits two-stage techniques to estimate ingrowth probability in the first step and estimates the ingrowth amount in the second step under the assumption that recruitment happened. Model VI turned out to be the best model for the first step analysis on ingrowth probability without dividing for the three forest types. This may be because Model VI consists of variables, such as basal area $\text{ha}^{-1}$, stand density, number of tree $\text{ha}^{-1}$, site index, minimum diameter, that have great impacts on ingrowth. Among those, ingrowth modeling is considered very important in terms of stand density (Fortin and DeBlois 2007; Yang and Huang 2015), the finally selected estimation equation also includes it as an important determinant. Especially, the mixed-species stand, as a measure of stand density to show the intensity of competition among standing trees, can be considered stand variable to explain Species Composition. For a similar study, Fortin and DeBlois (2007) estimated ingrowth with two-stage approach using ZIP (Zero-Inflated Poisson) model for the broad-leaved forest in the region of Quebec, southern Canada. Poisson distribution, which gives a basis for the model, means the distribution of the frequency distribution within a certain time and location (Jung 2010) and it is used as an alternative for binomial distribution in the case of low frequency of a certain event (Barry and Welsh 2002). Although this study included regression using ZIP distribution proposed by Lambert (1992) based on Poisson distribution and Zero-Inflated Negative Binomial (ZINB), it has some limitations as it included some restricted species.

Enough spatial within stand is required for a sapling, young tree, grows to be a tree with more than 6 cm at diameter of breast height (dbh). Here, “space” means the area for growth of saplings. Generally, it is necessary to know the interactions of environmental factors such as light, temperature, soil characteristics, and moisture for estimating growth. If space for growth and environmental requirements are met within stand, growths in diameter and height would happen over time (Lundqvist and Elfving 2010). To study the mechanism, through which recruitment occurs, the various environmental factors are needed as independent variables for ingrowth estimation equation based on the knowledge of physiological process. All of the estimation equation, developed by forest types, included an independent variable that represents stand density, and this gives an idea that the growth space within the stand is a major predictor for ingrowth. It is expected that an advanced model of more significance will be facilitated by further research on ingrowth models including forest origin, age-class, site index, although the ingrowth models at national levels have been devised by forest type.

The estimation equation developed in the current study includes variable for stand density, but not explanatory variables such as climate, soil, and some factors reflecting disturbance. Thus, the limitation is inevitable in the modeling the interaction between environmental factors related to climate and soil and trees newly being recruited, and this is the reason for low explanatory power of the model. Since climate change affects forests through disturbance and biological pattern by changing vegetation (Sousa 1984; Dale et al. 2000; Dawson et al. 2011; Seidl et al. 2011), the estimation equation needs these environmental factors for explanatory variables. A research (Klopcic and Boncina 2012), which developed two-stage approach using 21 environment variables for Fagus sylvatica L., included annual average temperature, precipitation, and ungulate density index for independent variables for the estimation equation. If current NFI field data are used, it is practically impossible to get such variables, but climate factors can be obtained from the materials offered by meteorological administration, thus they should be taken into account.

As for the ingrowth occurrence, naturally since a natural forest consists of a variety of species and is multilayer, it is likely that the ingrowth consistently occurs as time passes. By contrast, regarding artificial forests, the ingrowth tends to arise in the relatively low age-class only, as they are composed of simple even-aged stands by artificial planting whose purpose is to maximize the economic utility and productivity of land, thereby promoting the intensive forest management (Lee et al. 2010). It is essential to understand ecological characteristics of stands to expound the pattern by which the ingrowth is naturally generated. Thus, in follow-up research, there is a need for an estimation equation that reflects both the environmental conditions and the ecological characteristics of stands.
charateristics of ingrowth and the classification of stand origin.

Model VII was the finally selected equation to estimate ingrowth for broad-leaved forest, and it includes predictors such as growing stock, site index, and stand density. Site index is included only in the model for broad-leaved forest among the ingrowth estimation equations for each forest type. Site index is one of the basic information which plays an important role in forest management and provides quantitative information on forest productivity determined by geographical conditions, soil conditions, and climate factors (Elfväng and Kiviste 1997; Brovo and Montero 2001; Skovsgaard and Vanclay 2008). This research tested the estimation ability with measured data to determine the goodness of fit of ingrowth estimation equation, and set up the stand conditions according to site productivity based on yield table for each species. Ek (1974) analyzed the relation between the number of ingrowth trees and site productivity and reported no significant difference. Hann (1980) also estimated the tendency between higher site index and increase in ingrowth, but later mentioned the need of correction for the interpretation of the result of site index. Also, there was a related report on that site index has little impact on stand density, since it is obtained from average height of dominance tree and co-dominance tree in stand for specified reference age. Therefore, it should be noted that site index might not be a proper explanatory variable to reflect stand density. In addition, as Qin (1998) suggested, further research is needed on a methodology to test site productivity in uneven-aged forest and mixed-species forest.

In the current study, NFI data used to build an estimation equation have fixed sample size and measurement period. Tables 2 and 3 also used the data gathered at permanent sample plots and conducted analysis using two-stage approach. For example, forest resource inventory is conducted for detailed examination of various types of stand structure with different sample plot types (square plot, rectangular plot, circular plot, etc.) and different sample plot sizes. Then, in the case of investigating ingrowth in stand, there is higher possibility of ingrowth occurrence for greater sample size (Vanclay 1994; Li et al. 2011; Yang and Huang 2015). In a similar way, because the longer sample plot measurement interval, the higher possibility of ingrowth, this should be noted for an estimation equation to be used. Ingrowth estimation equation built in this study is able to provide estimates of ingrowth by forest types, which ables calculating the precise statistical information for forest inventory (National Institute of Forest Science 2018). For instance, the area of stand age class I in Korea is 203 ha (3.3%) out of total stocked land 6073 ha. However, although there exists I stand age class standing trees, the information on the growing stock of that class is provided being assumed to be “0”. Thus, the way to calculate the volume of standing trees with dbh of less than 6 cm should be reviewed in addition to exploiting the materials of ingrowth to calculate the growing stock of I stand age class (Develjak et al. 2014; National Institute of Forest Science 2018). It would be helpful for more precise statistical calculation to use the ingrowth estimation equation developed in this study and the data collected by NFI on species, diameter at root, volume, etc. To sum, the result of the current study has its contribution as it can provides the informational basis for forest management plan and forest statistical estimation. However, further study is needed to develop an estimation equation for greater suitability and to raise its reliability because the improvement in research on ingrowth modeling seems desirable in Korea.

5. Conclusion

This study employed the NFI monitoring data from throughout Korea to analyze ingrowth, which is not commonly studied because of the difficulties of time-series data collection. This study aimed to develop an ingrowth estimation equation according to forest type.

By applying a two-stage approach to develop the estimation equation, Model VI was evaluated to be optimal during the first stage of the ingrowth probability estimation model development. Our selection of the optimal model was processed through RMSE statistics calculated from each candidate model. The ingrowth probability estimation equation was then developed based on the finalized model. The experimental data of the sampling plots used in the modeling were applied to the estimation equation to estimate the ingrowth probability and evaluate the suitability of the proposed estimation equation, revealing no specific issues.

Moreover, the final ingrowth amount estimation equation for ingrowth number of forest types were selected based on seven candidate models through a statistical evaluation based on the fit index of the estimation equation. For forest types, Model IV, VI and VII were best fit. The equation for the ingrowth amount proved its estimation ability under various stand conditions and showed no issues in terms of suitability and applicability.

The results of this research could be combined with high-quality materials accumulated through the ongoing NFI and comparison data to develop an ingrowth model with a relatively advanced explanation ability. Furthermore, if more systematic calculations and verification processes are achieved using time-series monitoring data for ingrowth estimation, this research could represent a new direction in the field of ingrowth modeling for Korean forests.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adame P, Del Rio M, Cañellas I. 2010. Ingrowth model for Pyrenean oak stands in north-western Spain using continuous forest inventory data. Eur J Forest Res. 129(4):669–678.
Arabatzis AA, Burkhart HE. 1992. An evaluation of sampling methods and model forms for estimating height–diameter relationship in lobolly pine. Forest Sci. 38(1):192–198.

Allen RB. 1993. A permanent plot method for monitoring changes in indigenous forests: a field manual. Christchurch (New Zealand): Manaaki Whenua- Landcare Research.

Barry SC, Welsh AH. 2002. Generalized additive modelling and zero inflated count data. Ecol Model. 157(2–3):179–188.

Beers TW. 1962. Components of forest growth. J For. 60(4):245–248.

Bewick V, Cheek L, Ball J. 2005. Statistics review 14: logistic regression. Crit Care. 9(1):112–118.

Brovo F, Montero G. 2001. Site index estimation in Scots pine (Pinus sylvestris L) stands in the High Ebro Basin (northern Spain) using soil attributes. Forestry. 74(4):395–406.

Brovo F, Pando V, Ordóñez C, Lizarralde I. 2008. Modelling ingrowth in Mediterranean pine forests: a case study from Scots pine (Pinus sylvestris L) and Mediterranean maritime pine (Pinus pinaster Ait.) stands in Spain. Invest Agrist: Sist Recur For. 17(3):250–260.

Burnham KP, Anderson DR. 2004. Multimodel inference. Sociol Methods Res. 33(2):261–304.

Byun WH, Lee WK, Bae SW. 1996. Forest growth yield. Seoul (Korea): Yoochuno Media. (in Korean).

dale VH, Joyce LA, McNulty S, Neilson RP. 2000. The interplay between climate change, forests, and disturbances. Sci Total Environ. 262(3):201–204.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM. 2011. Beyond predictions: biodiversity conservation in a changing climate. Science. 332(6025):53–58.

Develjak M, Poljanse A, Zenko B. 2014. Modelling forest growing stock from inventory data: a data mining approach. Ecol Indic. 41:30–39.

Ek AR. 1974. Nonlinear models for stand table projection in mixed selection and irregular shelterwood forest stands. Ann For Sci. 28(4):118–127.

Elfving E, Kiviste A. 1997. Construction of site index equations for Pinus sylvestris L. using permanent plot data in Sweden. Forest Ecol Manag. 98(2):125–134.

Forestry Research Institute. 1996. Forest growth and yield prediction model. Seoul (Korea): Forest Research Institute (in Korean).

Fortin M, DeBlois J. 2007. Modelling tree recruitment with zero-inflated models: the example of hardwood stands in southern Quebec. Canada Forest Sci. 53(4):529–539.

Fournier T, Zhang X, Stokes A, Lambers H, Körner C. 2007. Plant growth modelling and applications: the increasing importance of plant architecture in growth models. Ann Bot. 101(8):1053–1063.

Gerhard C, Büchner RD, Klein AG, Schermelleh-Engel K. 2017. A fit index to assess model fit and detect omitted terms in nonlinear SEM. Struct Eqn Model Multidiscip J. 24(3):414–427. Available online.

Gillis MD, Omule AY, Brierley T. 2005. Monitoring Canada’s forests: the National Forest Inventory. For Chron. 81(2):214–221. Available online.

Hamilton DA, Jr, Brickell JE. 1983. Modeling methods for a two-state system with continuous responses. J For Res. 13(6):1117–1121.

Hann DV. 1980. Development and evaluation of an even- and uneven-aged ponderosa pine/Arizona fescues stand simulator. Research Paper INT-267. USDA Forest Servive, Ogden, UT.

Jung YH. 2010. Bayesian inference for repeated count data using mixed-effects zero-inflated models and permanent plots in the Acadian Forest Region of North America. Can J for Res. 41(10):2077–2089.

Leontius JR, Larsen DR, Loewenstein EF. 1999. A matrix transition model for an uneven-aged, oak-hickory forest in the Missouri Ozark highlands. Proceeding of the 10th Biennial Southern Silvilcultural Research Conference; Shreveport (LA).

Lundqvist L, Elving B. 2010. Influence of biomechanics and growing space on tree growth in young Pinus sylvestris stands. Forest Ecol Manag. 260(12):2143–2147.

Martin GL. 1982. A method for estimating ingrowth on permanent horizontal sample points. Forest Sci. 28(1):110–114.

McCollage P, Nelder JA. 1986. Generalized linear models. London (UK): Chapman & Hall.

National Institute of Forest Science. 2016. Development of national forest resources interpretation and assessment. Seoul (Korea): National Institute of Forest Science (in Korean).

National Institute of Forest Science. 2018. Tree volume biomass and yield table. Admin Pub. Seoul (Korea): National Institute of Forest Science (In Korean).

Qin Y. 1998. Ingrowth models and juvenile mixed woods stands dynamics [master’s thesis]. Alberta (Canada): University of Alberta.

SAS Institute Inc 2004. SAS/STAT 9.1 User’s guide. Cary (NC): USA.

Seidl R, Fernandes PM, Fonseca TF, Gillet F, Jönsson AM, Merganičová K, Netherer S, Arpacci A, Bontemps J, Bugmann H, et al. 2011. Modelling natural disturbances in forest ecosystems: a review. Ecol Modell. 222(4):903–924.

Shifley SR, Ek AR, Burk TE. 1993. A generalized methodology for estimating forest ingrowth at multiple threshold diameters. Forest Sci. 39(4):776–798.

Shifley SR. 1990. Analysis and modeling of patterns of forest ingrowth in the North Central United States [Ph.D dissertation]. Minneapolis (MI): University of Minnesota.

Shin MY. 1990. The use of ridge regression for yield prediction models with multicollinearity problems. J Kor for Soc. 79(3):260–268.

Shin MY, Lee MJ, Hong SA. 2003. Development of individual tree mortality equations for four coniferous species by using distance-independent competition index (in Korean with abstract in English). J Kor for Soc. 92(6):581–589.

Skovsgaard JP, Vanclay JK. 2008. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry. 81(1):13–31.

Son YM, Lee KH, Lee WK, Kwon SD. 2002. Stem taper equations for six major tree species in Korea. J Kor for Soc. 91(2):213–218.

Sousa WP. 1984. The role of disturbance in natural communities. Annu Rev Ecol Syst. 15(1):353–391. Available online.

Sung JH, Lee YG, Park KE, Shin MY. 2014. Assessment of stand diversity change by different silvicultural treatments for natural deciduous forests in Mt. Gariwang. J Kor for Soc. 103(4):613–621. Available online.

Vanclay JK. 1989. A growth model for north Queensland rainforests. Forest Ecol Manag. 27(3–4):245–271. Available online.

Vanclay JK. 1994. Modelling forest growth and yield: applications to mixed tropical forests. Wallingford (UK): CAB International.

Vanclay JK. 2003. Growth modelling and yield prediction for sustainable forest management. Malasy Forestor. 66(1):58–69.

Yang Y, Huang S. 2015. Two-stage ingrowth models for four major tree species in Alberta. Eur J Forest Res. 134(6):991–1004. Available online.

Yim JS, Kim CM, Kim ES, Yoo BO, Ryu JH, Kim SH, Kim JC, Jung IB. 2016. Development of national forest resources interpretation and assessment. Res Rep. Seoul (Korea): National Institute of Forest Science. (in Korean).