Dissertation of the Degree of
Master of Landscape Architecture

Estimating Korean Pine(*Pinus koraiensis*) Habitat Distribution Considering Climate Change Uncertainty
- Using Species Distribution Models and RCP Scenarios -

불확실성을 고려한 잣나무의 서식 적지 분포 예측
- 종 분포 모형과 RCP시나리오를 중심으로 -

August 2015

Graduate School of Seoul National University
Department of Landscape Architecture
Yoonjung Ahn

Abstract

Estimating Korean Pine(*Pinus koraiensis*) Habitat Distribution Considering Climate Change Uncertainty

- Using Species Distribution Models and RCP Scenarios -

Yoonjung Ahn
Landscape Architecture and Rural Systems Engineering
Graduate School of Agriculture and Life Science
Seoul National University
Advising Professor : DongKun, Lee

Climate change can significantly affect tree species distribution in forests. Therefore, adaptation planning is needed to obtain maximum returns on tree growth. *Pinus koraiensis*, the common name is Korean pine, is a major afforestation species in Korea and is normally distributed in frigid zones. For this reason, global warming could affect the distribution of the Korean pine. Therefore, this study aimed to predict the distribution of the Korean pine and its suitable habitat area considering uncertainty

by applying climate change scenarios in an ensemble model.

Species distribution points and environmental variables data were used for the input data in the model. First, a site index was considered when selecting present and absent points by using the stratified method. Secondly, environmental and climate variables were chosen by literature review and then correlation analysis was performed to select variables that were not correlated. Subsequently, the selected variables were confirmed with experts. Those variables were then used as input data of BIOMOD2 (BIOdiversity MODelling 2). Next, the present distribution model was made and the result was validated with data splitting and Receiver Operating Characteristic (ROC). Next, Representative Concentration Pathways (RCPs) scenarios were applied to the models to create the future distribution model. Finally, the ensemble models were built and consensus maps were created using model committee averaging (MCA). In addition, overlay maps and uncertainty maps were used to quantify the uncertainties of the results.

The estimated results of the individual models showed significant variation. Among the eight models, Random Forest (RF) had the highest accuracy. The Artificial Neural Network (ANN) model tended to overestimate results, and the Maximum Entropy Algorithm (Maxent) results were distinct from those of the other models. These differences can be explained by the algorithms of each model, the interaction of input data, and the verification

methodology. The uncertain area from individual models was excluded from the ensemble model results.

In the midterm future (2040s), the models themselves created the major differences observed in Korean pine distribution. In contrast, both the models and RCPs scenarios caused variation in the long-term future (2090s). Results of ensemble models were calculated using uncertainty and overlay maps, with the uncertainty of one overlay map close to 17%. The uncertainty of the five times overlaid area was around 8% in both the midterm and long-term futures.

Suitable habitat for the Korean pine in the midterm future is mainly distributed in the central part of Korea, Gangwon province, and the southern part of Korea. In the long-term future, this preferred area will disappear from the southern part of Korea as well as some areas of Gangwon province. Generally, most model and ensemble results predicted that the suitable habitat area would decrease in the mid- and long-term future.

As the Korean pine is an afforestation species, it cannot be planted in protected areas. Therefore, protected areas were eliminated from the results of the ensemble model.

The ratios of protected area were 25%, 25%, 19%, and 22% in RCPs 2.6, 4.5, 6.0, and 8.5, respectively, in the midterm future. There was no significant difference among the results. The protected area ratios were 24%, 40%, 31%, and 24% in the long-term future, indicating that available areas to plant Korean pine

will be reduced in the future.

In conclusion, climate change scenarios and species distribution models (SDMs) create uncertainties in the evaluation of the future distribution of the Korean pine. Therefore, when estimating species distribution under climate change, uncertainties should be considered. In addition, the models show that the suitable habitat area for the Korean pine will decrease in the future, making it important for the climate change adaptation plan to reduce this impact.

This study is significant in that it considered uncertainties in the SDMs and RCPs scenarios. The results of this study could be important considerations in the process of plantation planning.

- ☐ *Keywords: Ensemble models, BIOMOD2, Machine learning models, Statistical models, Species distribution, Forest ecosystem*
- ☐ *Student number: 2013-23255*

Table of contents

List of tables

List of Figures

I. Introduction

Thirty percent of the earth is covered with forest, making forest management important for maintaining the major functions of trees and plants, including carbon absorption, timber production, and air cleaning (Muraoka, 2015). However, climate change has been implicated as a factor in the alteration of the distribution of forest species, prompting the need for the development of adaptation plans (Jaeschke et al., 2013). Rising average temperatures have led to predictions of species extinction and shifting distribution ranges (Parmesan and Yohe, 2003).

Timber production is an important part of regional and national economies in the world(MEA, 2005), and a quarter of the Korean forest is used for this purpose (Statistics Annual Report, 2015). Thus, studies are needed to determine the most suitable areas to plant economically important trees in order to ensure optimal timber returns (Kim et al., 2012; Statistics Annual Report, 2014).

Estimating potential or suitable habitats has been an active area of study, and in the late 1900s, site index[1)] and overlay methods were used to investigate suitable habitats as influenced by climate change (Shin et al., 2006; Lee et al., 2012). However, overlay analysis was not considered the optimal method for this type of study because it tends to be subjective (Kwon et al., 2012).

---

1) Site index: the estimation of site quality, estimate growth. The standard year of Korea site index is 30 years(Shin et al., 2006; Korea forest research Institute, 2014).

Recently, logistic regression models and SDMs are being used for estimating suitable habitat area (Kim et al., 2012; Choi et al., 2015). However, these methods have some inherent uncertainty, and therefore have not been actively incorporated into decision-making processes for timber plantation areas (Spittlehouse and Stewart, 2003; Kwon et al., 2012). Afforestation businesses are closely linked to the economy, and therefore, uncertainties are considered an important concern.

BIOMOD2, which is based on R, provides 10 different models (eight of which are specified here): Generalized Linear Models (GLM), Generalized Additive Models (GAM), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), Artificial Neural Network(ANN), Generalized Boosted Models (GBM), Random Forest(RF), and Maximum Entropy Algorithm(Maxent). BIOMOD2 was used for estimating suitable habitat area for the Korean pine in this study. In addition, RCPs 2.6, 4.5, 6.0, and 8.5 scenarios were applied in the models.

The Korean pine is one of the major afforestation species in Korea, and it is not only a popular species for landscaping in urban areas, but also provides pine nuts and timber (Kim et al., 2001; Lee et al., 2009; Jo et al., 2013). However, since the species prefers the frigid zone, it could be vulnerable to alterations in temperature, and hence studies estimating suitable habitat area for the Korean pine under the effects of climate change are needed (Kim et al., 2011).

To estimate Korean pine distribution, this study will consider uncertainties of climate change scenarios and SDMs. The results of this study could prove useful as decisions regarding where to plan this valuable tree species are made.

II. Literature reviews

1. Uncertainties

Recent studies are trying to reducing uncertainties for increasing reliability of the results(Thuiller, 2003; Pearson et al., 2006; Kwon et al., 2014). First of all, recent studies try to reduce uncertainties in two different aspects. The first is the uncertainties in input data(Pearson et al., 2006). To reduce uncertainties, sensitivity analysis and Bayesian theory were suggested(Convertino, 2014). Especially, many studies were focus on reducing uncertainties in climate change scenarios. Uncertainties of climate change scenarios were quantified by using multi scenarios and comparing the results(Higa, 2013).

The second is uncertainties in species distribution models(Martins et al., 2014). To quantified the uncertainties in species distribution models multi models and ensemble models were used(Thuiller, 2003; Higa, 2013).

Comparing the multi models' results are mainly focus on clarifying the differences caused by models' algorithm (Pliscoff et al., 2014), However, ensemble models make a consensus result which is eliminated uncertainties from each individual models(Thuiller et al. 2004; Arau´jo et al., 2005; Pearson et al., 2006; Elith and Graham, 2009; Miller, 2014).

Especially ensemble models provide better results when there is significant variations among individual models(Sollich and Krogh,

1996; Kuncheva and Whitaker, 2003). Ensemble methods have become more popular because it provides various verification methods and overcome the shortages of individual models(Thuiller et al., 2009).

Overseas countries have applied ensemble models from early 2000(Thuiller, 2003; Thuiller et al., 2004; Pearson et al., 2006) but there are not many cases in Korea(Kwon et al., 2014). Uncertainties should be considered to support decision making(Beale and Lennon, 2012).

2. Species distribution models(SDMs)

Table 1. Characteristics of SDMs (P: Presence, PA: Presence and Absence)

Types	Models	Characteristics	Required data
Statistical models	GLM (Gneralized Linear Model)	This is applicable for categorial explanatory variable and needed to be defined for ecological theory function. However, it could be hard to interpretate the results if categorial variables were used because categorial variables are treated as dummy.	PA
	GAM (Gneralized Addictive Models)	It is an outstanding method when the explanatory variables and the dependent variable are nonlinear relation. It is widely used for conservation planning or analogizing with a population scale though it depends on degree of freedom and could not follow ecological niche theory and .	PA
	MARS (Multivariate Adaptive Regression Splines)	This model can handel a large dataset and faster compare to other SDMs. This model considers interaction between variables.	PA

Table 1 continue (P: Presence, PA: Presence and Absence)

Types	Models	Characteristics	Required data
Statistical models	FDA (Flexible Discriminant and Mixture Models)	This model was developed to over come the linear regression model. The large dataset are grouped by differentiation between the variables.	PA
Machine learning models	GBM (Generalized Boosted Model or Boosted Regression Trees)	This model is consisted with regression trees. Variables are divided into binary groups. GBM has great potential in faster and great performance. This model only deducts the best model and this is the difference with the other regression tree models.	PA
	GARP (The Genetic Algorithm for Rule-Set Prediction)	This model finds statistical relation with species distribution and climate variables. This model is very sophisticated and runs several times to find the best model.	PA
	SRE (Surface Range Envelop or BIOCLIM)	Hyper box method was introduce to develop this model. This model is suitable for analysis large scale.	P

Table 1 continue (P: Presence, PA: Presence and Absence)

Types	Models	Characteristics	Required data
Machine learning models	Maxent (Maximum Entropy Model)	Maximum entropy approach was used for estimating species distribution. This model is one of machine learning models estimating with variables and limited factors. This model generated great interest because it has higher predictive accuracy with only presence data.	P
	CTA (Classification and Regression Tree)	This model is useful when there are many categorial variables. It performs well with finding the relation among nonlinear and nonrepeated variables.	PA
	ANN (Artificial Neural Networks)	This model is based on artificial neural network algorithms. It has great analysis ability but It does not always show the better performance than other statistical models. Also, It is hard to understand the process of the model because of many hidden networks.	PA
	RF (Random Forest)	This model is based on making decision trees with great dataset and average them. It could analyze with mean square error or misclassification error rate.	PA

3. Climate change and vegetation distribution

Vegetation distribution change impacts on other land animal many studies have been focus on this subject(Cailleret et al., 2014).

In Korea, monitering method have been a major method for measuring the impact(Heo et al., 2005; Lee et al., 2010). Last 30 years the average temperature has been increased and the habitat has changed a lot(Ministry of Environment, 2013). Hence, there is the number of studies focusing on estimating forest species distribution under climate change have rapidly increased.

This type of studies could be divided into three groups. First is the target species which lives in a specific climate region. Secondly, protected or endanger species and lastly afforestation species(Gibson et al., 2014; Bede-Fazekas, 2014; Choi et al., 2014; Choi et al., 2015).

Bede-Fazekas(2014) studied about four different Pinus which is native species in mediterranean region. Shin et al(2012) estimated the distribution of forest species in Korea. Kim et al(2009) did the similar study with Shin et al(2012) but the study focuses on Gangwon province. However, those studies could not reflect the characteristics of individual species. Therefore, studies about individual species is needed.

There are some studies which focus on endangered species.

Yu et al.(2014) studied about seven endangered species and Park et al(2014) researched about Dwarf Stone Pine(Pinus pumila) which is one of red list species in IUCN.

Meanwhile, regression analysis and overlay method were used for finding suitable habitat area of afforestation(Shin et al., 2006; Lee et al., 2012). These methodologies are easily used but tend to be subjective. Because of this reason the result could not actively imply on policy(Kwon et al., 2012).

Lately, species distribution models were used for estimating suitable habitat area(Kim et al., 2012; Choi et al., 2015). However, afforestation accompanies economic investment and profit creation the result should be accurate. So the uncertainties in the input data and species distribution model should be consider to output the reliable results(Thuiller et al, 2004).

III. Methodology

1. Scope of study

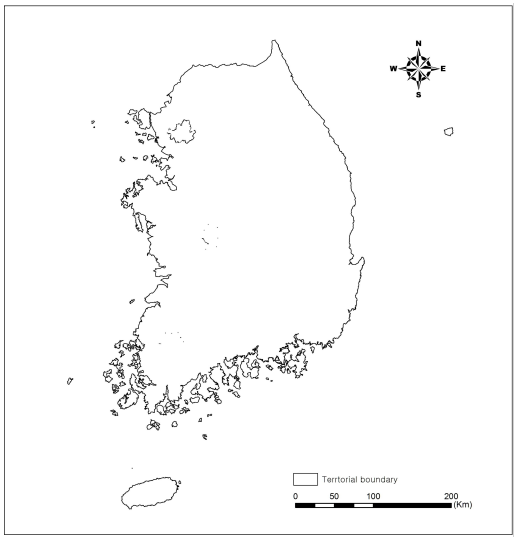


Figure 1. Study area

The research site for this study was the forest in Korea (Figure 1). The time scale of this study extended from the present (The average of 2001 to 2010), to the midterm future (The average of 2041 to 2050), and the long-term future (The average of 2091 to 2100).

The target species *P. koraiensis* is sensitive to climate change, making it a suitable candidate for this study. Climate variables were considered to find the tree's suitable habitat area throughout these time-periods (Han and Park, 1988).

2. Material and methods

2.1 Overview of the study

This study was divided into three steps. First, a literature review was done to select environmental variables that affect Korean pine distribution. Correlation analysis was then performed to select variables for the study, and variables that were not correlated were used for the input data of the models. The variables were then confirmed via an interview with experts. Second, data splitting and ROC were used for the verification of the models. Lastly, RCPs scenarios (2.6, 4.5, 6.0, and 8.5) were applied to estimate the future distribution of the Korean pine.

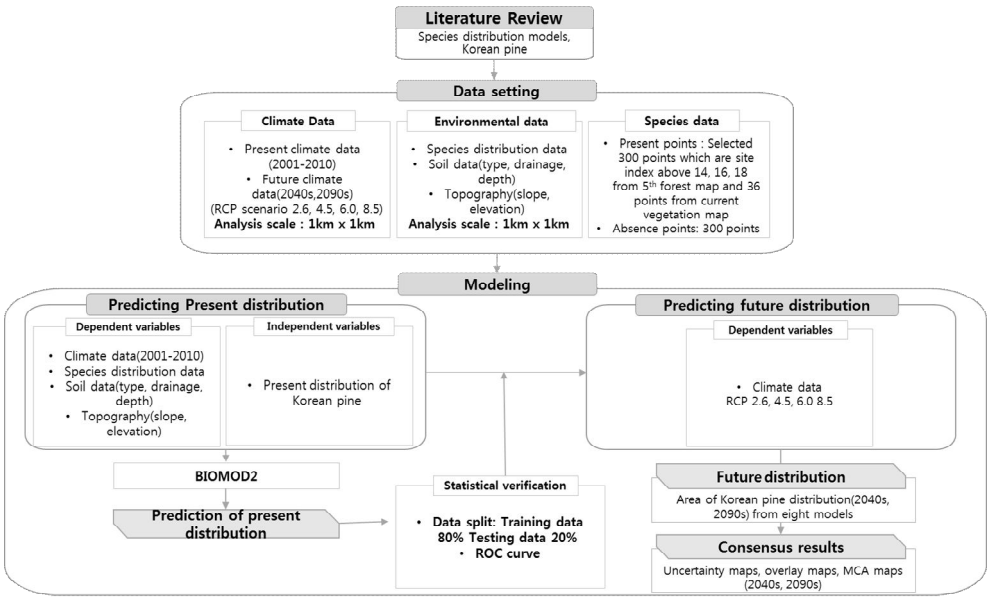


Figure 2. Flow of Study

2.2 Data setting

2.2.1 Climate variable

To evaluate climate variables, a literature review was done to determine the climate variables that affect Korean pine distribution.

Before creating the models themselves, however, a correlation analysis was performed (Fortin and Dale, 2014). Correlation among the variables of greater than ±0.6 is considered a strong correlation, and in this case, one of the associated variables should be eliminated from the model (Lee and Noh, 2013).

Climate observation data were used for the current climate data (2001 to 2010) and RCPs scenarios (2.6, 4.5, 6.0, and 8.5) were used for future climate data. Climate information was provided by the Korean Meteorological Administration.

2.2.2 Environmental variables

Five environmental variables affect Korean pine distribution. A digital elevation map was used to determine the altitude and slope. Information about soil depth, type, and drainage was provided by the National Academy of Agricultural Science.

According to Pearson and Dawson (2003), the analysis scale

should be local (approximately 1 km to 10 km resolution) to cover topographic variability. For this reason, the analysis scale was set to 1 km resolution.

2.2.3 Species occurrence and absence points

To select the occurrence data, the 5th forest map and current vegetation map were used. Because the Korean pine is naturally distributed in the southern part of Korea and the artificial forest is distributed in the central part of Korea, these two maps together identified the suitable distribution area of the Korean pine.

The vegetation map includes only vegetation that is in the age-class above 2. For this reason, this study assumes that the naturally distributed Korean pine is healthy. In addition, the site index was calculated using information from the 5th forest map, with an average site index of 14. Therefore, Korean pine with a site index above 14 is considered to be in good condition.

This study made two assumptions. The first is that the natural distribution area is actually a suitable habitat for the Korean pine. Second, it was assumed that the artificial forest was planted in the recommended area from Korea Forest Service and that the same management was applied to the entire forest.

Equal-proportional stratified sampling was used to select occurrence points from the 5th forest map (Elith and Graham, 2009), which involved selecting points with the same ratio as a portion of the stratum (Green, 1979; Hirzel and Guisan, 2002).

Site index was used as the standard for dividing the strata. The site index is different for each species, and for the Korean pine, its average is 14 at a standard age of 30 years. Therefore, 14, 16, and 18, which are above the average, were used for the standard of strata. The ratio of points in the group of 14, 16, and 18 was 7:2:1, respectively. Therefore, 300 points (210, 60, and 30 points) were selected. Each point was 3 km away from the other. The natural occurrence points were selected from each polygon, with a total of 36.

Pseudo-absence has been used widely because it can analyze using only occurrence points. However, pseudo-absence can sometimes reduce the accuracy of models, and hence absence points were selected from the forest, except from the habitat of the Korean pine (Elith et al., 2006; Elith and Graham, 2009; Phillips et al., 2009).

Table 2. Input data details and data reference

Category	Subcategory	Details	Data	Data Reference
Sampling points	Present points	300 points	5th Forest Map	Korea Forest Service
		36 points	Vegetation map	Ministry of environment
	Absence points	300 points	5th Forest Map	Korea Forest Service
Environmental data	Soil	Depth	Soil survey data	National Academy of Agricultural Science
		Drainage	Soil survey data	National Academy of Agricultural Science
		Types	Soil survey data	National Academy of Agricultural Science
	Topography	Altitude	DEM	Ministry of Environment
		Slope	DEM	Ministry of Environment
Climate data	Precipitation	Year precipitation	Observed data & RCP scenarios	Meteorological Administration
	Temperature	Average temperature in January	Observed data & RCP scenarios	Meteorological Administration
		Coldness index	Observed data & RCP scenarios	Meteorological Administration

2.2.4 Application to Models

BIOMOD2, one of the packages in R, was used for estimation of Korean pine distribution(Thuiller et al., 2009).

Among 10 available models [GLM, GAM, CTA, FDA, ANN, GBM, RF, Maxent, Multivariate Adaptive Regression Splines (MARS), and Rectilinear Envelope Similar to BIOCLIM (SRE)], only eight models were used to quantify the uncertainty of the SDMs. The MARS and SRE models were excluded because they cannot analyze categorical variables; this study included three categorical variables.

Individual models were run five times and binary maps were deduced by using thresholds. Ensemble models were created with seven different standards: mean of probability, median of probabilities, coefficient of variation of probabilities, coefficient of variation of probabilities, MCA, and weighted mean of probabilities.

To verify the models, data were split, with 80% for training data and 20% for test data, and each model was run five times (Thuiller, 2003; Kwon, 2014). ROC was used to provide the standard for selecting the optimal model. The area under the ROC curve, also called the Area under Curve (AUC), can range from 0 to 1. An AUC of 0.5 to 0.7 implies better-than-random prediction, an AUC of 0.7 to 0.9 means moderate predictive value, and an AUC of 0.9 is considered

to be of high predictive accuracy (Franklin, 2006; Pearson et al., 2006).

AUC is not affected by species prevalence, and hence it is dependable for model comparisons (Thuiller, 2003; Franklin, 2006). Therefore, it was chosen as the verification method in this study.

To create the binary maps, maximum sensitivity[2] and specificity[3] were used as thresholds (Hu and Jiang, 2011; Heibl and Renner, 2012; Kim et al., 2012; Kim et al., 2014).

2) Sensitivity: proportion of actual presences that are accurately predicted(Franklin, 2006)

3) Specificity: proportion of actual absences that are accurately predicted(Franklin, 2006)

IV. Results

1. Selecting variables

Only the variables that did not show correlation were selected for use in the models. A few soil and climate variables showed correlation, but since soil variables were considered important, they were included in the models.

As a result, the environmental variables of altitude, slope, soil type, drainage, and depth were selected. For the climate variables, the average temperature, yearly precipitation, and coldness index were chosen (Table 3).

Table 3. Names of variables and data format

Category	Name	Data Format	Reference
Environmental variables	Altitude	Continuous	Korea Forest Service, 2008
	Slope	Continuous	Korea Forest Service, 2008
	Soil type	Categorial	Korea Forest Service, 2008
	Soil drainage	Categorial	Korea Forest Service, 2008
	Soil depth	Categorial	Korea Forest Service, 2008
Climate variables	Yearly Precipitation[4)]	Continuous	Lee et al., 2009
	Average temperature in January	Continuous	Lee et al., 2009
	Coldness Index[5)]	Continuous	Lee et al., 2012

4) Yearly Precipitation: average precipitation in a year

5) Coldness Index: $CI = -\sum(5 - MT)$ (MT: mean temperature) (Chiu et al.,

2. The results of estimating distribution of Korean pine

2.1 Individual models' results

2.1.1 Estimation of present distribution

The result of verification by dividing the data into 80% for training data and 20% for test data showed an AUC above 0.7 in all models. According to the verification standard, this result is considered well established.

Among the models that were run five times, RF showed the highest ROC of 0.990 and CT had an lowest ROC of 0.822. The ROC of all models was above 0.8, which means that the accuracy of the models is high (Table 4).

Table 4. Result of ROC, sensitivity, and specificity of each model which were runs five times

Models	ROC	Sensitivity	Specificity
Maxent	0.881	70.482	91.696
GAM	0.828	71.386	80.623
GLM	0.825	81.024	69.55
GBM	0.923	79.819	88.235
CT	0.822	82.831	78.201
FDA	0.831	75.301	78.201
RF	0.995	96.084	96.194
ANN	0.881	77.108	85.467

2014)

Eight binary maps were constructed from the results of running each model five times (Figure 3). Generally, estimated results showed that the suitable habitat of the Korean pine is in the central district of Korea. The results of GAM showed that area to be around 13,000 ㎢ which is the smallest area, and GLM was approximately 19,000 ㎢ which is the biggest area. The remaining models' estimated results ranged from 13,000 ㎢ to 16,000 ㎢ .

In the 5th forest map, the artificial forest area was approximately 2,300 ㎢ and the natural area was approximately 5 ㎢ . There was a considerable difference between the results of the models and current distribution maps (Figure 3), likely because the models estimated potential habitat area based on the current establishment of the Korean pine (Figure 4).

Figure 3. The result of each model's prediction

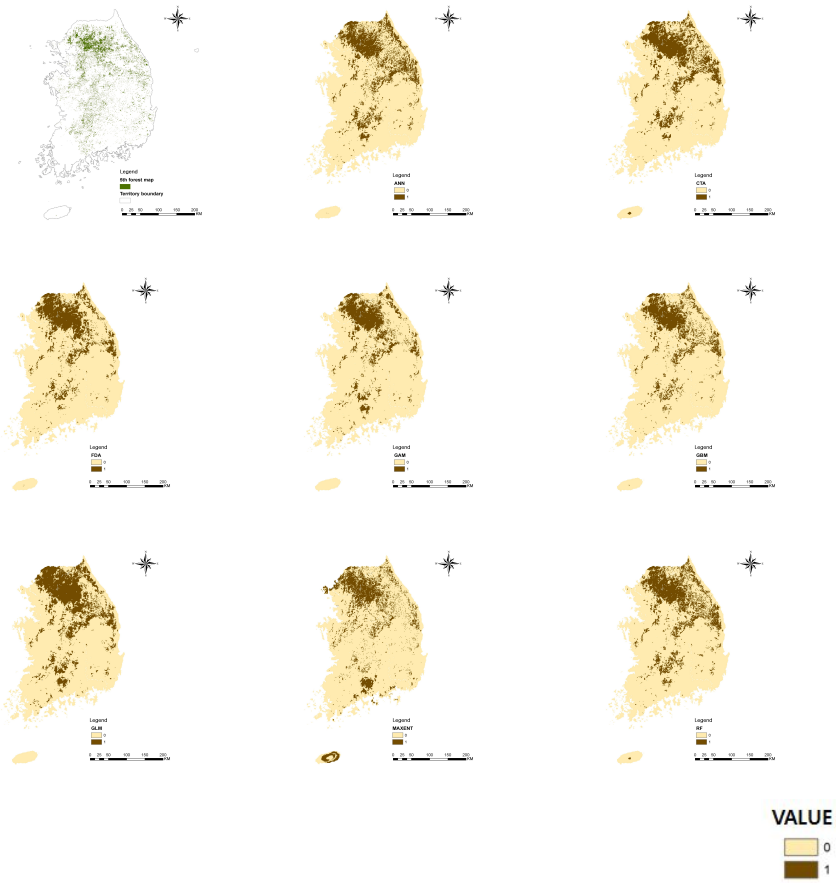


Figure 4. 5th Forest map(2006~2010) and individual models' prediction results(From left to right 5th Forest map, ANN, CT, FDA, GAM, GBM, GLM, Maxent, and RF)

2.1.2 Estimation of future distribution

① Midterm future

Figures 5 and 6 show the results of models in which RCPs scenarios were applied. Figure 5 shows the distribution area and Figure 6 shows the distribution pattern.

The ANN model overestimated results by factors of three to five compared to the other models. The distribution area estimated by ANN was appropriately 60,000 ㎢ to 70,000 ㎢. In contrast, the estimated results of CT, GBM, Maxent, and RF ranged from 2,000 ㎢ to 31,000 ㎢ (Figure 6).

Overall, there was no significant variation when the results of different RCPs scenarios were applied. However, when RCPs 4.5 was applied, the estimated distribution area was larger than the other scenarios, likely because future annual precipitation would correlate positively with an increase in Korean pine distribution in the southern province (Lee et al., 2009; IPCC, 2014).

To summarize, based on the overall estimated distribution results in the future, the Korean pine will be distributed in the central part of Korea, Gangwon province, and some parts in the south, including the Jiri mountain. The estimated area in Gangwon province where a natural habitat of the Korean pine will remain in the midterm future, however, decreased

in some models (Figure 6).

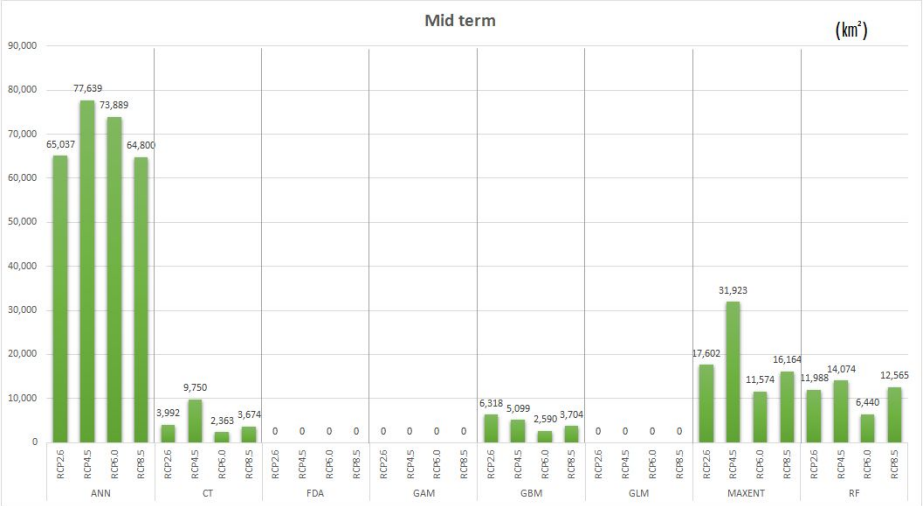


Figure 5. The result of 8 different models' prediction applying RCP 2.6, 4.5, 6.0, and 8.5

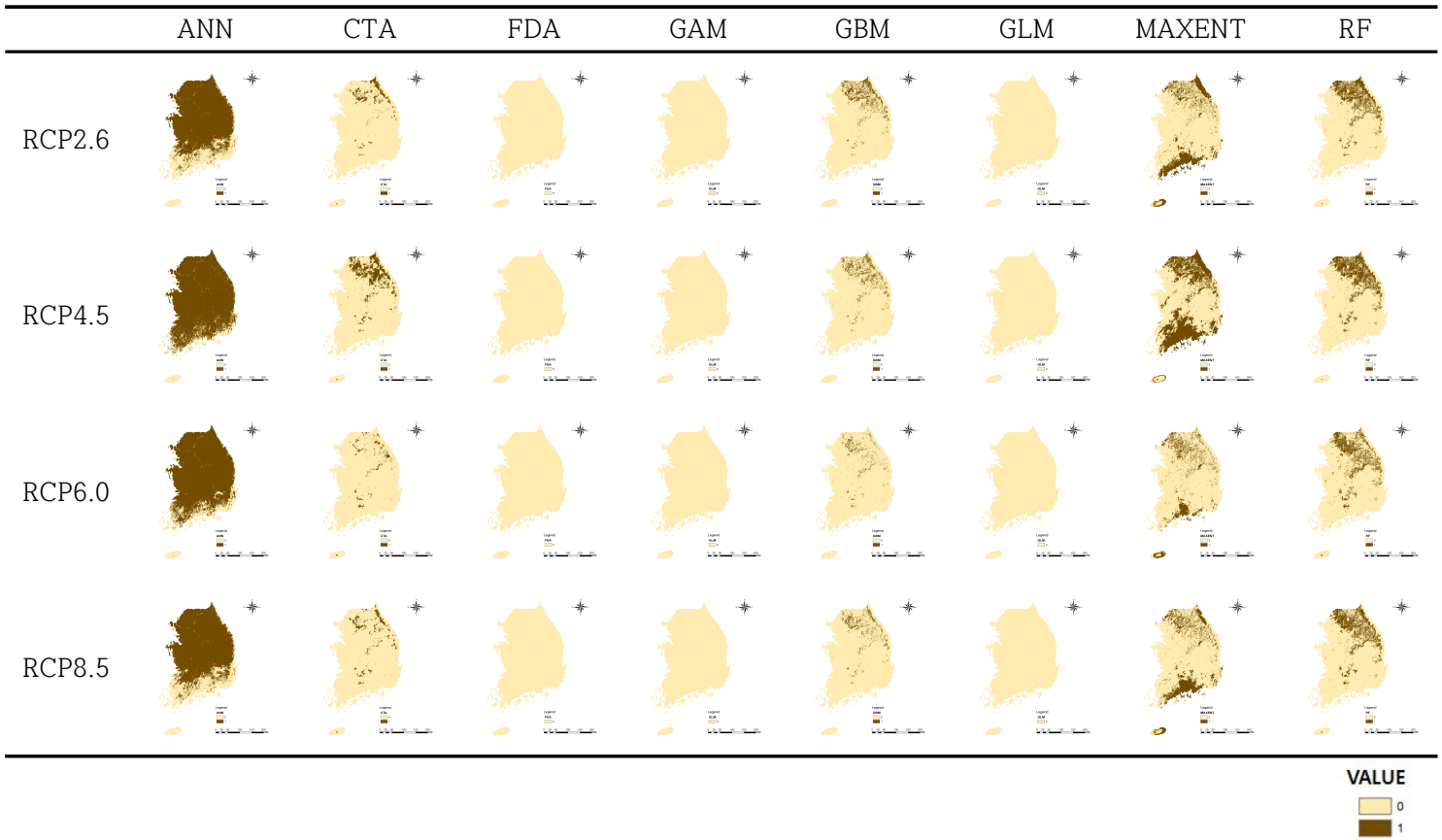


Figure 6. The results from eight SDMs(ANN, CT, FDA, GAM, GBM, GLM, Maxent, and RF) which were applied RCP 2.6, 6.0, 4.5, and 8.5 scenarios

② Long-term future

Figures 7 and 8 show the results of models in which RCPs scenarios were applied. Figure 7 shows the distribution area and Figure 8 shows the distribution pattern.

The results were identical to the results of the midterm future, with ANN overestimating compared to the other models. The ANN model estimated around 23,000 ㎢ to 63,000 ㎢, whereas the result of FDA, GAM, and GLM were 0 ㎢. CT, GBM, and RF showed similar estimated results to each other.

Results were also affected by the importance of input data. Similar to the results of the midterm future, Maxent estimated that the suitable habitat area will increase in the southern area of Korea owing to the influence of yearly precipitation (Choi et al., 2015), as shown in Figure 8.

The GBM model estimated that the suitable habitat area for Korean pine decreased from RCPs 2.6 to RCPs 8.5. In contrast, ANN, CT, and RF showed a fluctuation with RCPs 4.5. The results of Maxent increased with RCPs 8.5, likely from the influence of precipitation.

After considering the individual models' results, the machine learning models showed similar trends and the statistic models had the same result as zero. When comparing the results from each scenario, there were no significant

differences in the midterm future. However, in the long-term future, there was some disparity among the scenarios, with most models estimating that the suitable habitat area will decrease from RCPs 2.6 to RCPs 8.5.

To summarize, in the midterm future, the major differences were a result of the models, but in the long-term future, both the models and scenarios created variation (Jose et al., 2009). In addition, in the midterm future, suitable habitat area will predominantly be distributed in the central and southern parts of Korea and Gangwon province. In the long-term future, current areas of natural habitat will disappear in the southern part of Korea and in some areas of Gangwon province.

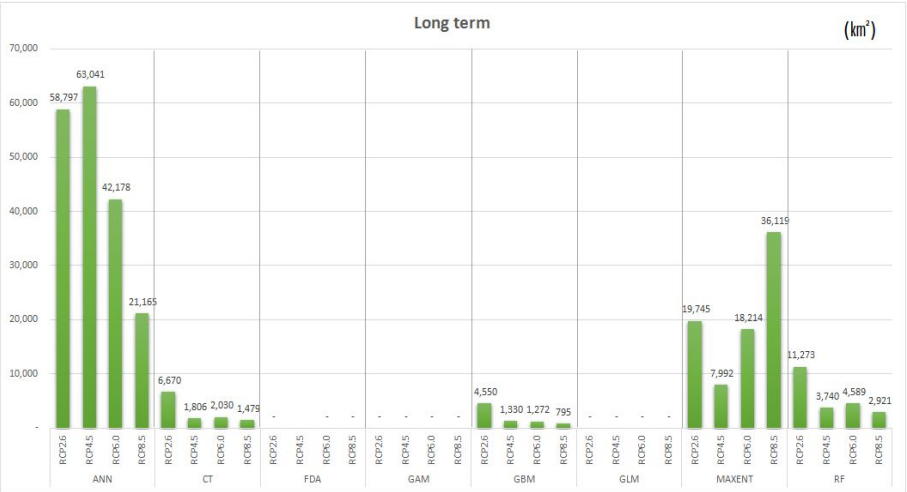


Figure 7. The suitable habitat area which are applied RCP 2.6, 4.5, 6.0, and 8.5

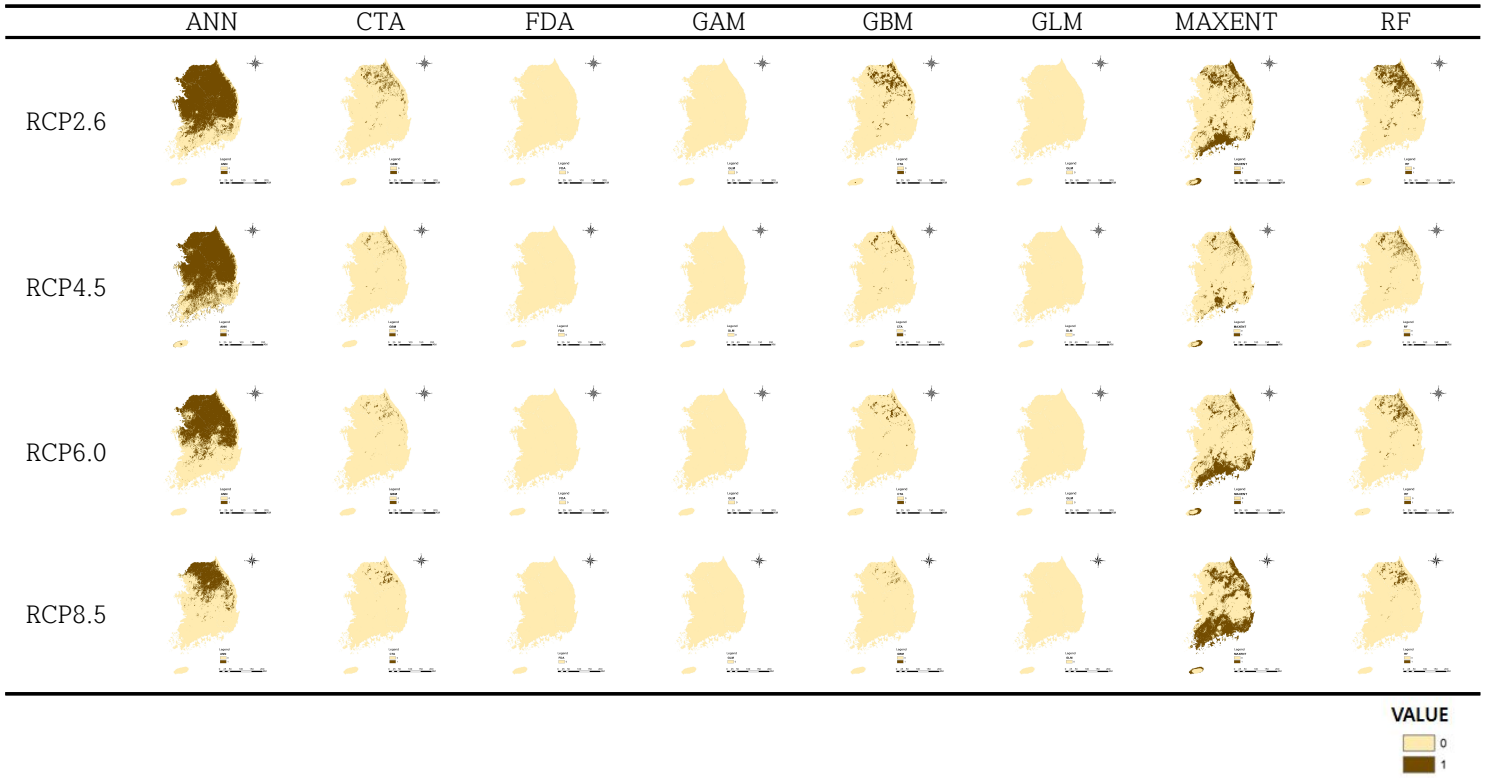


Figure 8. The results from eight SDMs(ANN, CT, FDA, GAM, GBM, GLM, Maxent, and RF) which were applied RCP 2.6, 6.0, 4.5, and 8.5 scenarios

2.2. Results of the ensemble models

2.2.1. Estimation of current distribution

The ensemble models were built using individual models that had an ROC above 0.7, and since all of the models met this criterion, they were all included in the ensemble models. In this study, all of the ensemble models had ROC above 0.9, making them well-established models (Table 5).

Table 5. The ROC of ensemble models

Model(abbreviation)	Cut off	ROC
Mean of probability(prob.mean)	569	0.917
Coefficient of variation of probabilities(prob.cv)	**NA**	**NA**
Confidence interval(prob.ci)	509	0.916
Confidence interval(rob.ci.alpha)	614.5	0.918
Median of probabilities(prob.median)	485.5	0.888
Model committee averaging(committee.averaging)	**393.5**	**0.913**
Weighted mean of probabilities(prob.mean.weight.decay)	571.5	0.919

Weighted mean of probabilities had the highest ROC of 0.919; however, MCA is capable of incorporating the results of individual models (Arau´jo, 2007; Coetzee et al., 2009; Marmion et al., 2009). In addition, when considering the results of the ensemble models, MCA represented the individual models' results well, and for this reason, MCA was chosen to represent the final outcome. The coefficient of

variation of probabilities measures the probability of uncertainties, and hence this value was used to show the uncertainties of the results.

The binary maps were constructed by using a cut off for the results of MCA. An uncertainty map was constructed from the coefficient of variation of probabilities, and frequency maps were constructed by overlaying the individual models' results (Figure 9).

The estimated suitable habitat area was approximately 16,700 ㎢ by MCA, but this overestimated the current Korean pine distribution. In addition, the area estimated to be a suitable habitat area by one model was approximately 7,100 ㎢ , and the same area was estimated to be 6,900 ㎢ when eight models were overlaid.

The overestimated figures were eliminated from the MCA and considered an uncertain area in the uncertainty map (Figure 9).

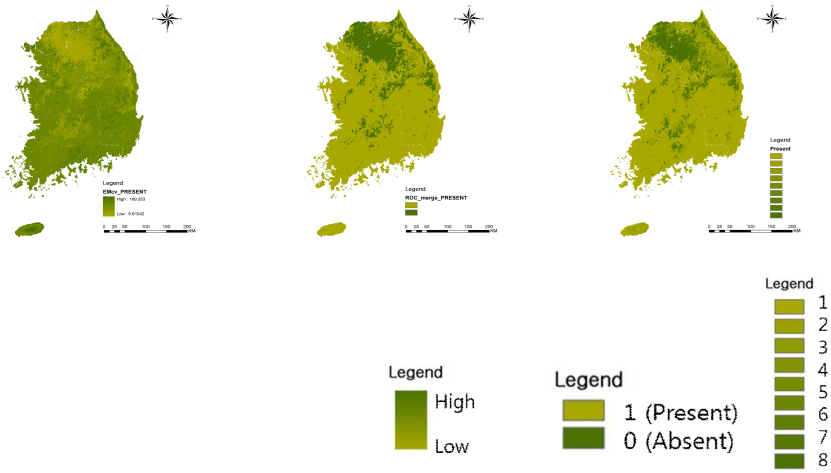


Figure 9. Result of estimating current distribution (Left: coefficient of variation of probabilities(prob.cv), Middle: model committee averaging, Right: loverly map)

2.2.2. Estimation of future results

① Midterm future

Binary maps were made using the results of MCA. Uncertainty maps were created using the coefficient of variation of probabilities, and frequency maps were built by the overlay of individual models' results (Figure 10).

Similar to the current distribution results, some parts of Gangwon province and central areas of Korea were identified as suitable habitat for the Korean pine. In addition, when RCPs scenarios were applied, the results showed a trend identical to that of the midterm future. Suitable habitat decreased 78%, 68%, 78%, and 77% when RCPs of 2.6, 4.5, 6.0, and 8.5, respectively, were applied. The once overlaid area ranged from 17,000 ㎢ to 5,000 ㎢ and the five times overlaid area ranged from 4,000 ㎢ to 700 ㎢ .

Moreover, the suitable habitat areas for the Korean pine in the central part of Korea and Gangwon province were reduced. When RCPs 2.6 was applied, only 2,000 ㎢ remained as suitable habitat area, and with RCPs 8.5, the area was only 1,800 ㎢ . In addition, the suitable habitat in the central part of Korea decreased from 9,000 ㎢ to 870 ㎢ when RCPs 2.6 and 8.5 were applied, respectively.

All of the ensemble models' results were comparable when

RCPs 2.6, 4.5, 6.0, and 8.5 scenarios were applied to the midterm future. Considerable variation among the RCPs scenarios may not occur until the long-term future (IPCC, 2014).

Figure 10. The result of prediction midterm future(2040s) applying RCP 2.6, 4.5, 6.0, and 8.5
(Left: coefficient of variation of probabilities(prob.cv), Middle: model committee averaging, Right: loverly map)

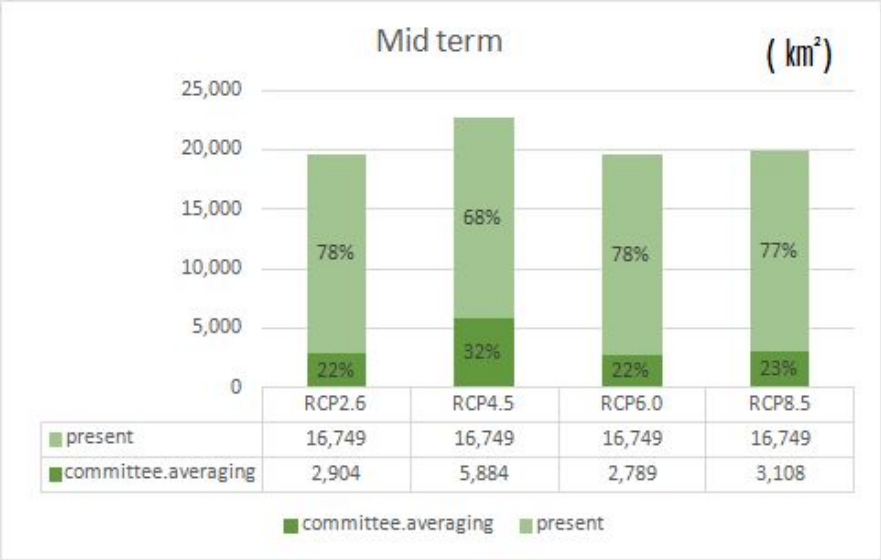


Figure 11. The estimated results of Korean pine habitat when applied RCP 2.6, 4.5, 6.0, and 8.5 in the midterm future(2040s)

② Long-term future

Binary maps were constructed from the results of MCA. Uncertainty maps were constructed using the coefficient of variation of probabilities, and frequency maps were created by the overlay of individual models' results (Figure 12).

Similar to the results of the midterm future, suitable habitat was distributed in the central part of Korea and the Gangwon area. However, with the exception of the RCPs 2.6 scenario, the habitat area dramatically decreased when RCPs 4.5, 6.0, and 8.5 were applied. Specifically, the habitat area decreased from 79% to 98% (Figure 16).

The overlay map shows that the once overlaid area ranged from 7,000 ㎢ to 57,000 ㎢ , and the five times overlaid area was approximately 800 ㎢ to 32,000 ㎢ . Suitable habitat decreased from 2,000 ㎢ to 200 ㎢ in the Gangwon area and from 600 ㎢ to 60 ㎢ when RCPs 2.6 and 8.5 were applied, respectively.

The results of the ensemble models showed a trend similar to that of the individual models. The suitable habitat area decreased when RCPs 2.6, 4.5, 6.0, and 8.5 were applied.

To summarize, the uncertain areas from the individual models were excluded from the results of the ensemble models. In the midterm future, there was no significant difference among the results when different RCPs scenarios were applied, but there were some variability in the long-term future.

Figure 12. The result of prediction long-term future(2090s) applying RCP 2.6, 4.5, 6.0, and 8.5
(Left: coefficient of variation of probabilities(prob.cv), Middle: model committee averaging, Right: loverly map)

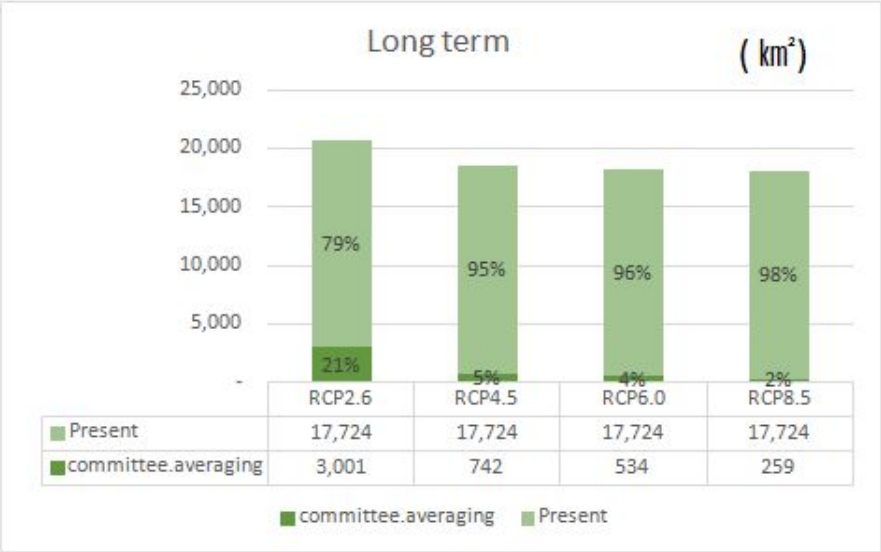


Figure 13. The estimated results of Korean pine habitat when applied RCP 2.6, 4.5, 6.0, 8.5 in the long-term future(2090s)

V. Discussion

1. Uncertainties of climate change scenarios and SDMs

The results indicated a range of uncertainties, and significant differences were made by the SDMs rather than the RCPs scenarios, similar to results described by Joes (2009).

In contrast, in the research of Higa (2013), variation was mainly created by climate change scenarios rather than SDMs. This difference could be explained by the fact that this study created situations using one climate model, but Higa (2013) used several models to make climate change scenarios.

When comparing the results of four different climate change scenarios (RCPs 2.6, 4.5, 6.0, and 8.5), there were some variations in the long-term future, likely because RCPs effects peak after the 2050s (IPCC, 2014).

There were some differences in the results of individual models, an outcome consistent with the results obtained by Martins et al. (2004), Pearson et al. (2006), and Jose et al. (2009). The differences are explained by the algorithms of each model, the interaction of input data, and verification methodology (Thuiller et al., 2004; Arau´jo et al., 2005;

Pearson et al., 2006; Elith and Graham, 2009; Miller, 2014).

Statistical models estimate the future using a data sample from the past. In contrast, machine learning models that are based on classification methods use decision trees to estimate the future. Maxent has a different algorithm compared to the other models because it considers absent points as background information (Elith et al., 2011), and for this reason, the results could be different in each model (Zukerman et al., 2001). The machine learning models showed better performance than the statistical models in this study, as also shown by Kellie and Kimes (2008). In contrast, Pliscoff et al. (2004) found that statistical models performed better.

Previous studies suggest that the formation of input data will affect the performance of the models. Three of eight categorical variables were used in this study, and they were treated as dummy variables with a value of zero in the statistical models, making the impact of these variables insignificant and difficult to interpret (Franklin, 2010). On the other hand, decision tree based models can handle categorical variables very well. Similarly, RF and ANN can identify the interaction among variables.

Statistical models are limited in their ability to handle skewed or multimodal responses. In contrast, machine learning models can predict the complex relations and

interactions among the variables (Muñoz, 2004) and for this reason, the machine learning models performed better in this study.

Input data was affected by the importance of the variables (Elith and Graham, 2009). The soil variables were not important variables in GAM, GBM, and FDA, which were statistical models because the soil variables are categorial variables. Hence, the climate variables had a greater effect on the models, and as a result, the habitat area was estimated to be 0 ㎢ in the future.

In contrast, in the machine learning models, especially ANN, the climatic variables were the important variables in the model and they tended to be overestimated. In the case of Maxent, annual precipitation was the most important variable in the model, and as precipitation—which correlated positively with Korean pine distribution—increased in southern parts of Korea in the future, the habitat area also increased (Choi et al., 2015).

As the results of individual models showed considerable variation, the ensemble method was used to reduce the uncertainties of the models (Jose, 2009). The ensemble model tended to have higher ROC compared to the individual models, as demonstrated by Arau´jo (2005). The ensemble model's accuracy could have been higher if individual models with lower ROC were excluded, but in this study, every model

was included in the structure of the ensemble model.

The ensemble model result, which created an average of the individual models' results, showed the same result as the mean of the binary maps of the individual models. Thus, the result of the overlay of the individual models was similar to that of the ensemble model (Figure 14).

The results of three models(FDA, GAM, and GLM) were 0 ㎢ in the midterm and long-term futures, and the results could be overlaid one to five times. When the results were overlaid three times, the outcome was the same as that of the ensemble model. The area with one overlay was the largest, which means that the model had uncertainty and its result was unreliable. In addition, when the results were overlaid five times, this led to the smallest area.

The overlaid maps' uncertainties were calculated using uncertainty maps. The uncertainties were calculated by dividing the standard deviation by the average, and the model is not scaled. However, if the number of uncertainties in the maps increases, then the results are dissimilar to the average of the models (Thuiller et al., 2014).

The uncertainty of the map with one overlay was approximately 17. With a three times overlay area, the uncertainty was near 8 to 10 in both the midterm and long-term future (Table 6).

This finding implies that the ensemble model reduces seven

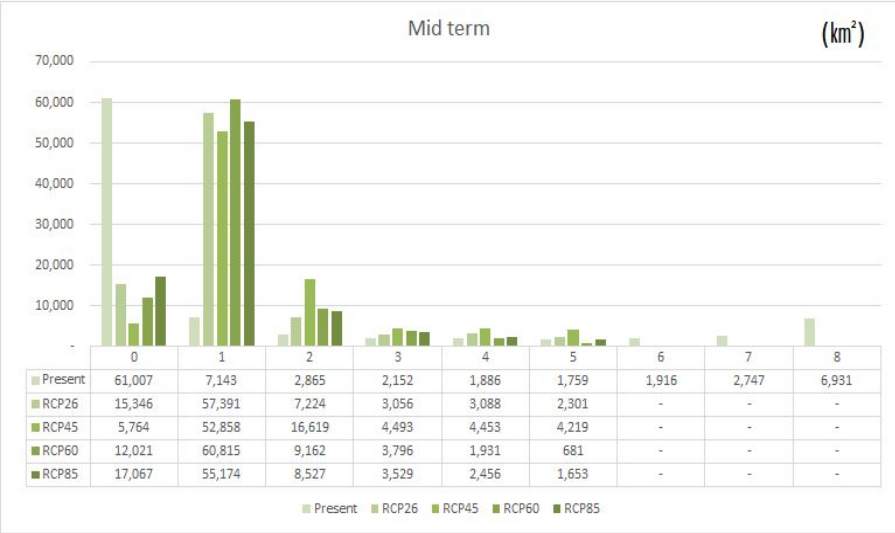
of the uncertainties in the individual models' results.

This study inspected the range of uncertainties by using ensemble model results and overlay maps, and also discussed the uncertainties of RCPs scenarios and SDMs. Most studies rely on the result of one model (Beale and Lennon, 2012) but it is important to investigate the uncertainties in the SDMs (Arau´jo et al., 2005b, 2006; Thuiller et al., 2005).

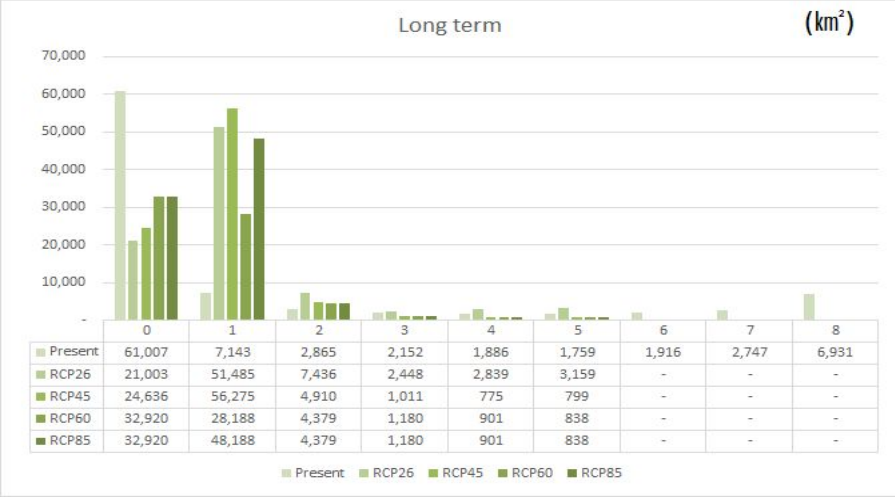
Table 6. The possibility of uncertainties in mid term future and long term future

(scale: standard deviation/ average of the result of models)

Mid term Future				
VALUE	RCP2.6	RCP4.5	RCP6.0	RCP8.5
0	18	17	18	18
1	17	17	17	17
2	13	15	12	12
3	10	10	10	10
4	9	9	10	9
5	9	8	9	9
Long term Future				
VALUE	RCP2.6	RCP4.5	RCP6.0	RCP8.5
0	18	16	18	17
1	17	15	17	17
2	13	13	15	15
3	11	9	11	13
4	10	9	11	11
5	9	8	10	10



	0	1	2	3	4	5	6	7	8
Present	61,007	7,143	2,865	2,152	1,886	1,759	1,916	2,747	6,931
RCP26	15,346	57,391	7,224	3,056	3,088	2,301	-	-	-
RCP45	5,764	52,858	16,619	4,493	4,453	4,219	-	-	-
RCP60	12,021	60,815	9,162	3,796	1,931	681	-	-	-
RCP85	17,067	55,174	8,527	3,529	2,456	1,653	-	-	-



	0	1	2	3	4	5	6	7	8
Present	61,007	7,143	2,865	2,152	1,886	1,759	1,916	2,747	6,931
RCP26	21,003	51,485	7,436	2,448	2,839	3,159	-	-	-
RCP45	24,636	56,275	4,910	1,011	775	799	-	-	-
RCP60	32,920	28,188	4,379	1,180	901	838	-	-	-
RCP85	32,920	48,188	4,379	1,180	901	838	-	-	-

Figure 14. The result of overlaid in the mid and long term future when applied RCP2.6, 4.5, 6.0, and 8.5

2. Estimated Korean pine distribution

This study shows that in general, the distribution of the Korean pine will decrease in the future. In particular, the natural distribution area will nearly disappear in the long-term future. Choi et al. (2015) also found a similar result because the Korean pine lives in the frigid zone and temperatures will increase in the future (Kim et al., 2011; IPCC, 2015).

Specifically, in the long-term future, the distribution area in the southern part of Korea where the natural habitat area is currently distributed will disappear, largely because the Korean pine is found in the high altitudes of the Jiri mountain, and as temperatures rise in the future, this habitat will disappear.

In the same way, in some parts of the Gangwon area where both natural habitat and artificial forest exist currently, the habitat will decrease in the long-term future because of rising temperatures. Meanwhile, the central part of Korea will remain a suitable habitat because this region's temperatures will decrease and precipitation will increase in the future.

These climate change factors are important considerations when making Korean pine plantation plans.

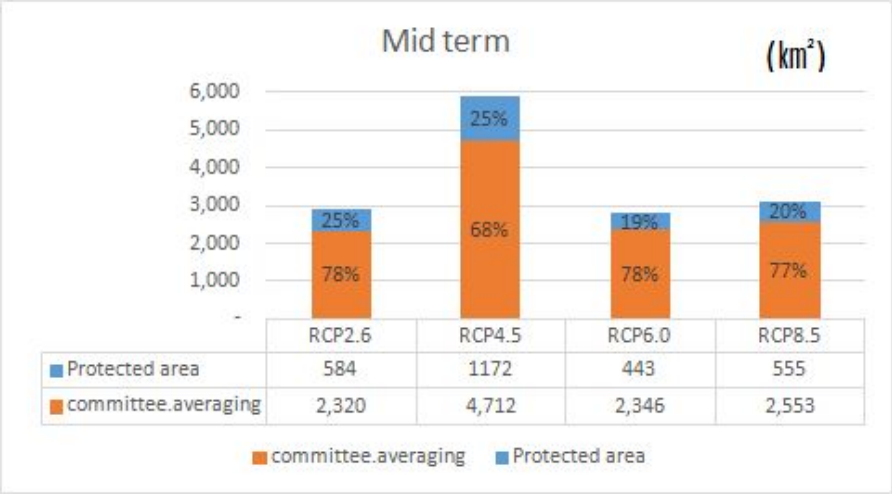
As discussed previously, the Korean pine is an afforestation species, and hence it cannot be planted in protected areas,

including the Ecological Landscape Conservation Area, Wetland Conservation Area, Wildlife Protected Area, National Parks, Development Limited District, Natural Habitat Protected Area, Baekdudaegan Conservation Area, and Forest Genetic Resources Protection Forest. These restrictions will reduce the Korean pine's plantation areas in the future.

Therefore, it was assumed that the current area of protected forest will remain constant in the future, and hence these regions were excluded from the results of the ensemble model.

The ratios of protected area were 25%, 25%, 19%, and 22% with RCPs 2.6, 4.5, 6.0, and 8.5, respectively, in the midterm future. There was no significant difference among the results.

The protected area ratios were 24%, 40%, 31%, and 24% in the long-term future; these results indicate that the area where the Korean pine can be planted will be greatly reduced in the future (Figure 15).



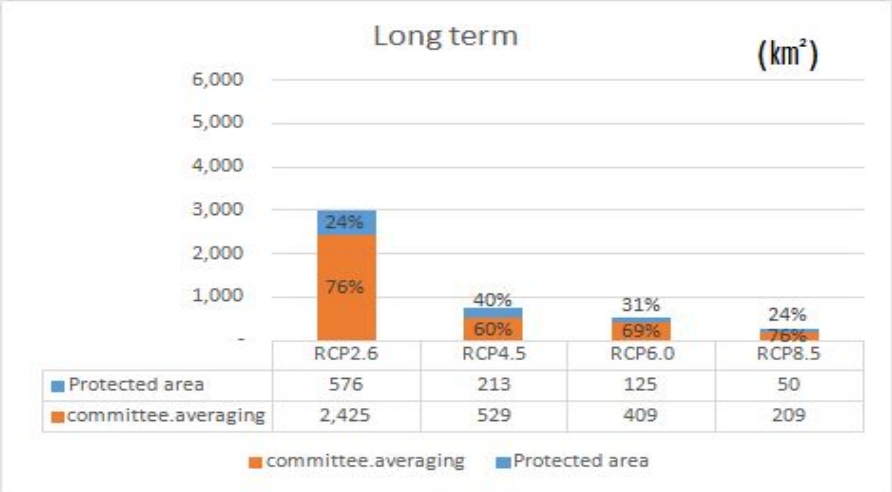


Figure 15. The results of ratio of protected area in mid and long term future(top: midterm future, bottom: long term future)

VI. Conclusion

Species distribution models have been widely used for biodiversity assessment, prioritizing protected area, and environmental planning. However, because of the uncertainties in these models, they have not been used for decision-making, especially with the afforestation sector because any uncertainty can have economic implications. Therefore, this study is significant because these uncertainties are considered when estimating future areas of Korean pine distribution.

To summarize, this study first investigated the range of uncertainties in climate change scenarios. Next, the range of uncertainties among the SDMs was evaluated. Lastly, the ensemble method was implemented. Among the eight models, RF had the highest accuracy, while ANN and Maxent overestimated results. The differences among the models are explained by climate change scenarios in the long-term future, whereas this variation was not seen in the midterm future models.

Consensus results were made by the MCA, uncertainty maps, and overlay maps. The uncertainties produced by the individual models were quantified using uncertainty and overlay maps. As a result, approximately 7% of the uncertainties were reduced by using the ensemble models,

indicating that ensemble models can be used to reduce the uncertainties of individual models. Therefore, multi scenarios and SDMs can be considered for estimation studies.

The suitable area of the Korean pine will decrease in the future, especially in southern parts of Korea where the natural habitat will actually disappear in the long-term future. In addition, suitable habitat areas in some parts of Gangwon province will also decrease. Only the central part of Korea will remain a suitable area for this species.

This study has significance in that it takes the uncertainties of the models into consideration. However, assumptions were made, including that the Korean pine in its natural habitat area was healthy, and that planted trees were in a suitable habitat area and all managed similarly.

One limitation of this study is that it does not take the process of succession into account. Korean pine is a coniferous species, and it could be dominated by deciduous species in the future. In addition, social variables such as land use planning and human disturbance of this afforestation species were not considered.

This study investigated the distribution change of the Korean pine under variables of climate change. The results indicate that suitable habitat area for the Korean pine will decrease in the future. Therefore, a climate change adaptation plan is needed to protect this species in the

future. The results of this study could be useful in the decision-making process, especially in planning the establishment of suitable plantation areas while considering climate change.

VII. Reference

Lee, Jung Hyun and Choi, Byoung Hee, 2010. Evergreen trees, Korean Peninsula, Machilus thunbergii, Neolitsea sericea, northernmost Limit of distribution, Quercus acuta. Korean Journal of Plant Taxonomy, 40(4), pp.267-273.

Choi, Jaeyong·Lee, Peter Sang-Hoon and Lee, Sanghyuk, 2015. Anticipation of the Future Suitable Cultivation Areas for Korean Pines in Korean Peninsula with Climate Change. Journal of the Korea Society of Environmental Restoration Technology, 18(1), pp.103-113.

Chon, Sang-Keun·Shin, Man Yong·Chung, Dong-Jun·Jang,Yong-Seok and Kim,Myung-Soo 1999. Characteristics of the Early Growth for Korean White Pine(*Pinus koraiensis Sieb. et Zucc.*) and Effects of Local Climatic Conditions on the Growth. Journal of Korean Forest Society, 88(1), pp.73-85.

Heo, Inhye·Kwon, Won-tae·Chun, Youngmoon and Lee, Seungho, 2006. Impact of Temperature Rising on the Distribution of Plant. Journal of Environmental Impact Assessment, 15(1), pp.67-78.

Joo,Young-Tuk·Chon,Sang-Kuen and Chung, Dong-Jun, 1999. The Effect of some Meterological Factors on the Seed

Characteristics in Korean White Pine. Korean Journal of Agricultural and Forest Meteorology, 1(1), pp.1-10.

Kim, Hyun Seop·Bae, Sang Won·Jang, Seok Chang and Jeong, Jun Mo, 2011. Stand Structure and Growth Characteristics at Different Elevations of the Korean Pine (*Pinus koraiensis*) Natural Forest on Mt . Seorak. Journal of Korean Forest Society, 27(3), pp.157-167.

Kim, Il Hyun· Shin, Man Yong· Kim, Young Chai· Chon, Sang Keun, 2001. Evaluation of reproductive and vegetative growth in a mature stand of Korean pine under simulated climatic condition. Korean Society Agricultural and Forest Metrology, 3(4), pp.185-198.

Kim, Jum-Su, 2009. A Research on Changes in the Production Environment of Gangwon Province's Forest Products Depending on Climate Change, Research Institute for Gangwon.

Kim, Yong-Kyung· Lee, Woo-Kyung·Park, Young-hwan·Oh, Suhyun and Heo, Jun-Hyeok , 2012. Changes in Potential Distribution of *Pinus rigida* Caused by Climate Changes in Korea. Journal of Korean Forest Society, 101(3). pp.509-516.

Korea Environment Ministry, 2013. Korean Climate Change Assessment Report.

Korea Forest Service, 2008. The study of impact and

response in forestry, Korean Forest institute.

Korea Forest Service, 2014. Statistical Yearbook of Forestry.

Kwon, Hyuk Soo, 2014. Applying Ensemble Model for Identifying Uncertainty in the Species Distribution Models. Journal of Korean Society for Geospatial Information System, 22(4), pp.47-52.

Kwon, Hyuk Soo·Ryu, Ji-eun, Seo, Chang Wan·Kim, Jiyeon·Lim,Dong-ok·Seo, Mi-Hwan, 2012. A Study on Distribution Characteristics of Corylopsis coreana Using SDM. Journal of Environmental Impact Assessment, 21(5), pp.735-74.

Kwon, Hyuk Soo·Seo, Chang Wan and Park, Chong Hwa. 2012. Development of Species Distribution Models and Evaluation of Species Richness in Jirisan region. Journal of the Korean Society for Geospatial Information System, 20(3), pp.11-18.

Lee, Do-Hyung and Hwang, Jae Woo, 2000. A Study for the Site Condition and Vegetation Structure for the Natural Stands of Pinus koraiensis S. et Z. in Mt. Palgong. Journal of Research Development, 19(1), pp.68-76.

Lee, Sangtae·Bae, Sang-Won·Jang, Seok Chang·Hwang, Jae Hong·Chung, Jungmo and Kim, Hyun-Seop, 2009. A Study on the Relationship Between Radial Growth and Climate Factors by Regions in Korean Pine (Pinus

koraiensis). Journal of Korean Forest Society, 98(6), pp.733-739.

Lee, Yong Seok·Sung, Joo Han·Chun, Jung Hwa and Shin, Man Yong, 2012. Development of Site Index Equations and Assessment of Productive Areas Based on Environmental Foctors for Major Coniferous Tree Species. Journal of Korean Forest Society, 101(3), pp.395-404.

Park, Hyun-Chul·Lee, Jung-Hwan·Lee, Gwan-Gyu, 2014. Predicting the suitable habitat of the Pinus pumila under climate change. Journal of Environmental Impact Assessment, 23(5), pp.380-393.

Shin, Hyung Jin · Park, Min Ji and Kim, Seong Joon, 2012. Evaluation of Forest Watershed Hydro-Ecology using Measured Data and RHESSys Model -For the Seolmacheon Catchment. Journal of Korea Water Resources Association , 45(12), pp.1293-1307.

Shin, Hyung-Jin·Park, Geun-Ae·Park, Min-Ji·Kim, Seong-Joon , 2012. Projection of Forest Vegetation Change by Applying Future Climate Change Scenario MIROC3.2 A1B. Journal of the Korean Society for Geospatial Information System, 15(1), pp.64-76.

Shin, Man Yong and Kim, Young Chai, 2002. Effects of Local Climatic Condistions on The Early Growth in Progeny Test Stands of Korean White pine. Korean Journal of

Agricultural and Forest Meteorology, 4(1), pp.1-11.

Shin, Man Yong·Jang, Yong-Seok·Han, Sang-Urk and Kim, Young-Chai 2002. Effects of Local Climatic Conditions on the yearly Cone Production in Progeny Test Stands of Korean White Pine. Korean Society Agricultural and Forest Metrology, 4(3), pp.141-150.

Shin, Man Yong·Jung, Il Bin·Koo, Kyo-Sang and Won, Heong-Gyu, 2006. Development of a Site Index Equation for Pinus Koraiensis Based on Environmental Factors and Estimation of Productive Areas for Reforestation. Korean Journal of Agricultural and Forest Meteorology, 8(2), pp.97-106.

Akos, B.F., Levente, H. and Marton, K., 2014. Impact of climate change on the potential distribution of Mediterranean pines. Quarterl Journal of the Hungarian Meteorological Service, 118(1), pp.41-52.

Arau´jo, M. B., Robert J. W., Richard J. L. and Markus E., 2005a. Reducing uncertainty in projections of extinction risk from climate change. Global Ecol. Biogeogr. 14: pp.529-538.

Arau´jo, M.B. et al., 2005. Reducing uncertainty in projections of extinction risk from climate change. Global Ecology and Biogeography, 14(6), pp.529-538.

Arau´jo, M.B., Pearson, R. G., Thuiller, W. and E.rhard, M., 2005b. Validation of species - climate impact models under climate change. Global Change Biology, 11, pp.1504-1513.

Archer, K.J. and Kimes, R. V., 2008. Empirical characterization of random forest variable importance measures. Computational Statistics and Data Analysis, 52(4), pp.2249-2260.

Austin, M., 2007. species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling, 200(1-2), pp.1-19.

Beale, C. M., and Lennon, J. J. (2012). Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1586), 247-58.

Beale, C.M. and Lennon, J.J., 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 367(1586), pp.247-58.

Boyd, J. and Banzhaf, S., 2007. What are ecosystem services? The need for standardized environmental accounting units. Ecological Economics 63 (2-3), pp.616-626.

Cailleret, M., Heurich, M. and Bugmann, H., 2014. Reduction

in browsing intensity may not compensate climate change effects on tree species composition in the Bavarian Forest National Park. Forest Ecology and Management, 328, pp.179-192.

Calama, R., Gordo, F. J., Mutke, S. and Montero, G., 2008. An empirical ecological-type model for predicting stone pine (Pinus pinea L.) cone production in the Northern Plateau (Spain). Forest Ecology and Management, 255(3-4), pp.660-673.

Calama, R., Mutke, S., Jose. T., Gordo, J. and Montero, G., 2011. Modelling spatial and temporal variability in a zero-inflated variable: The case of stone pine (Pinus pinea L.) cone production. Ecological Modelling, 222(3), pp.606-618.

Chiu, C. A., Chiou, C. R., Lin, J. R., Lin, P. H., & Lin, C. T. 2014. Coldness index does not indicate the upper limit of evergreen broad-leaved forest on a subtropical island. Journal of Forest Research, 19(1), 115-124.

Convertino, M. Munoz-Carpena, R., Chu-Agor, M.L., Kiker, G.a. and Linkov, I., 2014. Untangling drivers of species distributions: Global sensitivity and uncertainty analyses of Maxent. Environmental Modelling and Software, 51, pp.296-309.

De Groot, R., 2006. Function-analysis and valuation as a tool

to assess land use conflicts in planning for sustainable, multi-functional landscapes. Landscape and Urban Planning, 75, pp.175-186.

Edwards, M. and Morse, D.R., 1995. identification in biodiversity research. TREE, 10(4), pp.153-158.

Ehrlich, P.R. and Ehrlich, A.H., 1981. Extinction: the causes and consequences of the disappearance of species. RandomHouse, New York.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. and Zimmermann, N.E., 2006. Novel methods improve pre- diction of species' distributions from occurrence data. Ecography 29, pp.129-151.

Elith, J., Leathwick, J.R. and Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 77(4), pp.802-813.

Elith, J. and Graham, C.H., 2009. Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. Ecography, 32(December 2008), pp.66-77.

Elith, J., Phillips, S. J., Hastie, T., Dudik, M. C., Yung E.Y. and Colin J. 2011. A statistical explanation of Maxent for ecologists. Diversity and Distributions, 17, pp.43-57.

Friedman, J.H. and Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine, 22, 1365-1381.

Green, R.H., 1979. Sampling design and statistical methods for environmental biologists, John Wiley and Sons.

Guisan, A. and Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8(9), pp.993-1009.

Hastie, T., Tibshirani, R. and Buja, A., 1994. Flexible Discriminant Analysis by Optimal Scoring. Journal of the American Statistical Association, 89(428), pp.1255-1270.

Hastie, T., Tibshirani, R. and Andreas, B., 1995. Flexible discriminant and mixture models. Neural networks and statistics, pp.1-23.

Hastie, T., Tibshirani, R. and Friedman, J.H., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

Heibl, C. and Renner, S. S. 2012. Distribution Models and a Dated Phylogeny for Chilean Oxalis species Reveal Occupation of New Habitats by Different.

Higa, M., Nakao, K., Tsuyama, I., Nakazono, E. Y., and Masatsugu M. T., 2013. Indicator plant species selection for monitoring the impact of climate change based on prediction uncertainty. Ecological Indicators, 29, pp.307-315.

Hijmans, R.J. and Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology, 12(12), pp.2272-2281.

Hirzel, A. and Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological Modelling, 157, pp.331-341.

Hu, J. and Jiang, Z., 2011. Climate Change Hastens the Conservation Urgency of an Endangered Ungulate, PLOS ONE, 6(8), e22873.

IPCC, 2014. Working Group Ⅱ Chapter 17. pp.14-15.

Jaeschke, A., Torsten, B., Bjorn, R, and Carl, B., 2013. Can they keep up with climate change? - Integrating specific dispersal abilities of protected Odonata in species distribution modelling. Insect Conservation and Diversity, 6(1), pp.93-103.

Jeon, S. W., Kim, J., Jung, H., Lee, W. K. and Kim, J. S., 2014. species Distribution Modeling of Endangered Mammals for Ecosystem Services Valuation, 17(1), pp.111-122.

Jose´ Alexandre F. Diniz-Filho, Luis Mauricio Bini, Thiago

Fernando Rangel, Rafael D. Loyola, Christian Hof, D.N. and ´jo, ´.-B. and M.B.A., 2009. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography, 32(6), pp.897-906.

Kuncheva, L. and Whitaker, C., 2003. Measures of diversity in classifier ensembles, Machine Learning, 51, pp. 181-207.

Leathwick, J.R. Elith, J., Francis, M. P., Hastie, T. and Taylor, P., 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. Marine Ecology Progress Series, 321, pp.267-281.

Lee W. S., Alan R. Graefe and Doohyun Hwang, 2013. Willingness to Pay for an Ecological Park Experience. Asia Pacific Journal of Tourism Research, 18(3), pp.288-302.

Marmion, M. et al., 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions, 15, pp.59-69.

Meller, L., M. Cabeza, S. Pironon, M. Barbet-Massin, L. Maiorano, D. Georges and W. Thuiller, 2014, Ensemble distribution models in conservation prioritization: from consensus predictions to consensus reserve networks, Diversity and

Distributions. 20 pp.309-321.

Millenium Ecosystem Assessment, 2005. Ecosystems and human well-being.

Miller, J. a., 2014. Virtual species distribution models: Using simulated data to evaluate aspects of model performance. Progress in Physical Geography, 38(1), pp.117-128.

Muñoz, J., &Felicísimo, Á. M. (2004). Comparison of statistical methods commonly used in predictive modelling. Journal of Vegetation Science, 15(2), 285.

Muraoka, H., Saitoh, T.M. and Nagai, S., 2015. Long-term and interdisciplinary research on forest ecosystem functions : challenges at Takayama site since 1993 Takayama chronicle. pp.197-200.

Parmesan, C. and Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature, 421(6918), pp.37-42.

Pearson, R.G. and Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global ecology and biogeography. 12, pp.361-371.

Pearson, R.G. et al., 2006. Model-based uncertainty in species range prediction. Journal of Biogeography, 33, pp.1704-1711.

Pearson, R.G., Thuiller, W., Arau´jo, M. B., Martinez-Meyer, E.

B., Lluls M., Colin M., Lera S., Pedro D., Terence P. and Lees, D. C., 2006. Model-based uncertainty in species range prediction. Journal of Biogeography, 33, pp.1704-1711.

Phillips, S.J., Dudik, M., Elith, J.,Graham, C. H., Lehmann, A.,Leathwick, J.F. and Simon, 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecological Applications, 19(1), pp.181-197.

Pliscoff, P. Luebert, F.H., Hartmut H., Guisan, A., 2014. Effects of alternative sets of climatic predictors on species distribution models and associated estimates of extinction risk: A test with plants in an arid environment. Ecological Modelling, 288, pp.166-177.

Sollich, P. and Krogh, A., 1996. Learning with ensembles: How overfitting can be useful, Advances in Neural Information Processing Systems, volume 8, pp. 190-196,

Spittlehouse, D.L. and Stewart, R.B., 2003. Adaptation to climate change in forest management. J. Ecosyst. Mangment. 4, pp.1-11

Thuiller, W. Miguel B. Arau´jo, Richard G. Pearson, Robert J. W., Lluis B., and Sandra L., 2004. Biodiversity conservation: uncertainty in predictions of extinction risk. Nature 427, pp.145-148

Thuiller, W., 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology, 9(10), pp.1353-1362.

Thuiller, W., Lavorel, S., Arau´ jo, M.B., Sykes, M.T. & Prentice, I.C., 2005 Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences of the United States of America, 102, 8245-8250.

Thuiller, W., Lafourcade, B., Engler, R., Arau´jo, M. B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. Ecography, 32(3), pp.369-373.

Thuiller, A.W. Georges, D., Engler, R.G., Maintainer D., Thuiller, C. W., 2015. Package "biomod2."

Yu, J., Wang, C., Wan, J., Han, S., Wang, Q. and Nie, S., 2014. A model-based method to evaluate the ability of nature reserves to protect endangered tree species in the context of climate change. Forest Ecology and Management, 327, pp.48-54.

Zhang, M.-G. Zhou, Z. Chen, W. Cannon, C. H. Raes, N. Slik, J. W. and Ferryet, S. 2014. Major declines of woody plant species ranges under climate change in Yunnan, China B. Bradley, ed. Diversity and Distributions, 20(4), pp.405-415.

Zukerman, I., Albrecht, D. W., and Albrecht, D. W., 2001.

Predictive Statistical Models for UserModeling. User Modeling and User-Adapted Interaction, 11 (1-2), 5-18.

<Website>

Korea forest institute(http://book.kfri.go.kr)

Korea forest service(http://www.forest.go.kr/)

<Text book>

Lee, Hee Yeon and Noh, Seung Chul, 2013. Statistical analysis. Seoul : Moonwoo

Im, KyungBin, 2001. Illustrated guide to Korean flora. Seoul : Bori

Fortin, M. and Dale, M.R.T., 2014. Spatial Analysis a guide for ecologist, Cambridge University Press.

Franklin, J., 2006. Mapping species distributions : spatial inference and prediction, Cambridge University Press.

Appendix 1. Correlation analysis for selecting variables

	Correlation coefficient																
	existence	log Altitude	logSlope	log yearprep	Junetemp	Apriltemp	Octprep	Sepprep	Janavg temp	Augavg temp	yearavg	yearmax	yearmin	coldindex	Soildepth	Soiltype	Soildrain
existence	1	.223**	.241**	.518**	.073	.001	.105	−.030	−.202**	−.202**	−.448**	−.320**	−.461**	−.054	−.066	−.071	−.129
p value		.002	.001	.000	.303	.989	.140	.677	.004	.004	.000	.000	.000	.447	.358	.322	.070
logAltitude	.223**	1	.545**	.289**	.131	.218**	.216**	.304**	−.615**	−.615**	−.709**	−.676**	−.675**	−.056	.205**	.207**	.028
p value	.002		.000	.000	.066	.002	.002	.000	.000	.000	.000	.000	.000	.433	.004	.004	.697
logSlope	.241**	.545**	1	.253**	.062	.112	.122	.125	−.282**	−.282**	−.421**	−.387**	−.404**	−.170*	.128	.112	−.085
p value	.001	.000		.000	.385	.115	.085	.078	.000	.000	.000	.000	.000	.016	.072	.116	.235
logyearprep	.518**	.289**	.253**	1	.604**	.514**	.346**	.464**	−.104	−.104	−.464**	−.361**	−.453**	−.398**	.049	.041	−.063
p value	.000	.000	.000		.000	.000	.000	.000	.143	.143	.000	.000	.000	.000	.491	.567	.374
Junetemp	.073	.131	.062	.604**	1	.777**	−.112	.246**	.058	.058	−.060	.003	−.076	−.297**	.095	−.002	.046
p value	.303	.066	.385	.000		.000	.114	.000	.413	.413	.399	.969	.285	.000	.181	.977	.516
Augtemp	.633**	.171*	.232**	.785**	.287**	.107	.003	−.035	−.053	−.053	−.430**	−.256**	−.468**	−.264**	−.062	−.058	−.146*
p value	.000	.016	.001	.000	.000	.133	.967	.626	.455	.455	.000	.000	.000	.000	.385	.414	.039
Apriltemp	.001	.218**	.112	.514**	.777**	1	.130	.400**	−.171*	−.171*	−.107	−.137	−.077	−.119	.266**	.190**	.189**
p value	.989	.002	.115	.000	.000		.068	.000	.016	.016	.132	.054	.283	.093	.000	.007	.008
Octprep	.105	.216**	.122	.346**	−.112	.130	1	.790**	−.341**	−.341**	−.321**	−.436**	−.225**	−.043	.096	.143*	.021
p value	.140	.002	.085	.000	.114	.068		.000	.000	.000	.000	.000	.001	.545	.178	.043	.766
Sepprep	−.030	.304**	.125	.464**	.246**	.400**	.790**	1	−.256**	−.256**	−.286**	−.394**	−.203**	−.267**	.160*	.145*	.092
p value	.677	.000	.078	.000	.000	.000	.000		.000	.000	.000	.000	.004	.000	.024	.041	.198
Janavgtemp	−.202**	−.615**	−.282**	−.104	.058	−.171*	−.341**	−.256**	1	1.000**	.815**	.873**	.740**	−.504**	−.301**	−.318**	−.166*
p value	.004	.000	.000	.143	.413	.016	.000	.000		.000	.000	.000	.000	.000	.000	.000	.019
Augavgtemp	−.202**	−.615**	−.282**	−.104	.058	−.171*	−.341**	−.256**	1.000**	1	.815**	.873**	.740**	−.504**	−.301**	−.318**	−.166*
p value	.004	.000	.000	.143	.413	.016	.000	.000	.000		.000	.000	.000	.000	.000	.000	.019
yearavg	−.448**	−.709**	−.421**	−.464**	−.060	−.107	−.321**	−.286**	.815**	.815**	1	.925**	.977**	.075	−.179*	−.191**	−.026
p value	.000	.000	.000	.000	.399	.132	.000	.000	.000	.000		.000	.000	.295	.012	.007	.716
yearmax	−.320**	−.676**	−.387**	−.361**	.003	−.137	−.436**	−.394**	.873**	.873**	.925**	1	.829**	.007	−.192**	−.233**	−.046
p value	.000	.000	.000	.000	.969	.054	.000	.000	.000	.000	.000		.000	.926	.007	.001	.524
yearmin	−.461**	−.675**	−.404**	−.453**	−.076	−.077	−.225**	−.203**	.740**	.740**	.977**	.829**	1	.096	−.177*	−.171*	−.030
p value	.000	.000	.000	.000	.285	.283	.001	.004	.000	.000	.000	.000		.178	.013	.016	.680
coldindex	−.054	−.056	−.170*	−.398**	−.297**	−.119	−.043	−.267**	−.504**	−.504**	.075	.007	.096	1	.173*	.199**	.190**
p value	.447	.433	.016	.000	.000	.093	.545	.000	.000	.000	.295	.926	.178		.015	.005	.007
Soildepth	−.066	.205**	.128	.049	.095	.266**	.096	.160*	−.301**	−.301**	−.179*	−.192**	−.177*	.173*	1	.853**	.868**
p value	.358	.004	.072	.491	.181	.000	.178	.024	.000	.000	.012	.007	.013	.015		.000	.000
Soiltype	−.071	.207**	.112	.041	−.002	.190**	.143*	.145*	−.318**	−.318**	−.191**	−.233**	−.171*	.199**	.853**	1	.687**
p value	.322	.004	.116	.567	.977	.007	.043	.041	.000	.000	.007	.001	.016	.005	.000		.000
Soildrain	−.129	.028	−.085	−.063	.046	.189**	.021	.092	−.166*	−.166*	−.026	−.046	−.030	.190**	.868**	.687**	1
p value	.070	.697	.235	.374	.516	.008	.766	.198	.019	.019	.716	.524	.680	.007	.000	.000	

**. The correlation coefficient is significant at the 0.01 level (both sides) .
*. The correlation coefficient is significant at the 0.05 level (both sides)

Appendix 2. Importance of variables

The importance of variables are based on algorithm of CT(Thuiller, 2015). The importance were given by the Minimum mean square error(MSE) and the frequency of the usage when creating the decision trees. The higher importance means significant contribution on models(Friedman and Meulman, 2003; Elith et al.,2008).

	Maxent	GAM	GLM	GBM	CTA	FDA	RF	ANN
Run1								
Soil drainage	0	0.076	0	0.001	0	0	0.006	0.022
Soil depth	0	0.01	0.034	0	0	0	0.005	0.006
Soiltype	0.001	0.002	0	0	0	0	0.006	0.005
Altitude	0.173	0.02	0	0.067	0.07	0.056	0.074	0.213
Slope	0.14	0.037	0.044	0.044	0	0.029	0.068	0.03
Janavgtemp	0.190	0.079	0.073	0.086	0	0.152	0.111	0.052
Yearprcp	0.554	0.402	0.419	0.437	0.356	0.415	0.396	0.128
Coldindex	0.526	0.566	0.608	0.494	0.59	0.542	0.336	0.774
Run 2								
Soil drainage	0.035	0.058	0	0	0	0	0.004	0.008
Soil depth	0	0.007	0.029	0	0	0	0.004	0.003
Soiltype	0.003	0.002	0	0	0	0	0.005	0.039
Altitude	0.185	0.003	0	0.053	0.071	0	0.09	0.276
Slope	0.124	0.021	0.026	0.024	0	0	0.067	0.128
Janavgtemp	0.184	0.09	0.062	0.109	0.144	0.148	0.114	0.126
Yearprcp	0.628	0.368	0.375	0.431	0.465	0.384	0.388	0.401
Coldindex	0.453	0.566	0.607	0.477	0.469	0.472	0.391	0.558
Run 3								
Soil drainage	0	0.032	0	0.001	0	0	0.004	0.009
Soil depth	0	0	0	0	0	0	0.005	0.007
Soiltype	0	0.002	0	0	0	0	0.007	0.022
Altitude	0.171	0	0	0.053	0.109	0	0.097	0.296
Slope	0.062	0.011	0	0.017	0	0	0.075	0.161
Janavgtemp	0.091	0.089	0.103	0.082	0.19	0.167	0.104	0.149
Yearprcp	0.415	0.287	0.283	0.375	0.371	0.303	0.316	0.399
Coldindex	0.524	0.542	0.584	0.468	0.57	0.528	0.332	0.561
Run 4								

Soil drainage	0.003	0.091	0	0.001	0	0	0.004	0.004
Soil depth	0	0.011	0.036	0	0	0	0.005	0.002
Soiltype	0	0.004	0	0	0	0	0.009	0.023
Altitude	0.014	0.007	0	0.059	0.09	0.07	0.094	0.191
Slope	0.105	0.025	0.03	0.029	0	0	0.057	0.131
Janavgtemp	0.004	0.061	0.046	0.043	0	0.146	0.074	0.14
Yearprcp	0.335	0.326	0.296	0.408	0.414	0.374	0.366	0.455
Coldindex	0.510	0.641	0.696	0.506	0.485	0.638	0.414	0.569
Run 5								
Soil drainage	0	0.039	0	0	0	0	0.002	0.008
Soil depth	0	0.01	0	0	0	0	0.004	0.003
Soiltype	0	0.001	0	0	0	0	0.006	0.044
Altitude	0.058	0.001	0	0.044	0.066	0.021	0.085	0.153
Slope	0.116	0.027	0.034	0.021	0	0.021	0.057	0.085
Janavgtemp	0.020	0.023	0.046	0.037	0	0.052	0.085	0.134
Yearprcp	0.415	0.342	0.362	0.4	0.34	0.451	0.352	0.389
Coldindex	0.508	0.594	0.654	0.548	0.604	0.695	0.426	0.601
Full								
Soil drainage	0.002	0.033	0	0	0	0	0.006	0.003
Soil depth	0	0.006	0	0	0	0	0.01	0.001
Soiltype	0	0	0	0	0	0	0.016	0
Altitude	0.203	0.002	0	0.069	0.112	0.034	0.161	0.129
Slope	0.076	0.021	0.022	0.028	0	0.022	0.136	0.001
Janavgtemp	0.036	0.045	0.087	0.059	0.17	0.083	0.137	0.048
Yearprcp	0.415	0.317	0.319	0.444	0.386	0.34	0.449	0.075
Coldindex	0.647	0.635	0.722	0.497	0.562	0.62	0.375	0.839

Appendix 3. The results of verification of individual models

	RUN1	RUN2	RUN3	RUN4	RUN5	Full
Maxent	0.765	0.74	0.842	0.807	0.753	0.871
GAM	0.802	0.817	0.838	0.835	0.759	0.829
GLM	0.827	0.826	0.826	0.847	0.75	0.824
GBM	0.824	0.853	0.884	0.874	0.807	0.927
CTA	0.8	0.797	0.797	0.779	0.708	0.818
FDA	0.82	0.813	0.818	0.829	0.786	0.83
RF	0.819	0.849	0.868	0.865	0.819	0.998
ANN	0.735	0.802	0.845	0.837	0.705	0.795

Appendix 4. The result of verification of ensemble models

Mean of probability(prob.mean)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.917	577.5	81.325	85.813
Coefficient of variation of probabilities(prob.cv)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	NA	NA	NA	NA
Confidence interval(prob.ci)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.915	517.5	81.627	85.813
Confidence interval(rob.ci.alpha)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.917	637.5	81.024	85.467
Median of probabilities(prob.median)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.887	498.5	84.94	76.817
Model committee averaging(committee.averaging)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.912	457.5	83.434	80.277
Weighted mean of probabilities(prob.mean.weight.decay)				
	Testing.data	Cutoff	Sensitivity	Specificity
ROC	0.919	574.5	81.627	85.813

Appendix 5. The results of response function

Data	Attributes
Drainage of soil (layer1)	1: Very Good drainage, 2: Very Good drainage ~ Good, 3: Good 4: Good ~ Slightly better 5: Bad 6: Very bad
Depth of soil (layer2)	1: Very deep(above 100cm), 2: Very deep ~ Deep 3: deep(50-100cm) 4: Deep ~ Moderate 5: Moderate(20-50cm) 6: Moderate ~ Shallow 7: Shallow(below 20cm) 8:Very Shallow
Soil type (layer3)	0: Rock 1: Gravel 2: Sand 3:Sandy loam 4:Sandy silt 5: Silt clay loam 6:clay loam

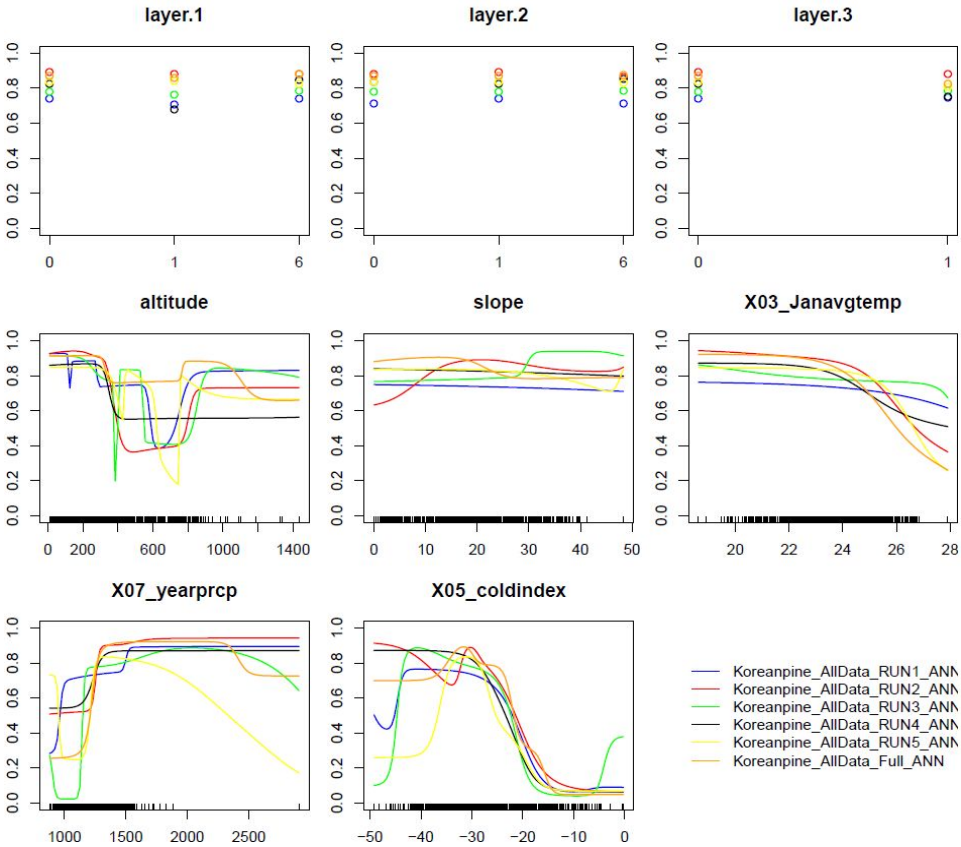


Figure 16. Response function of ANN (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

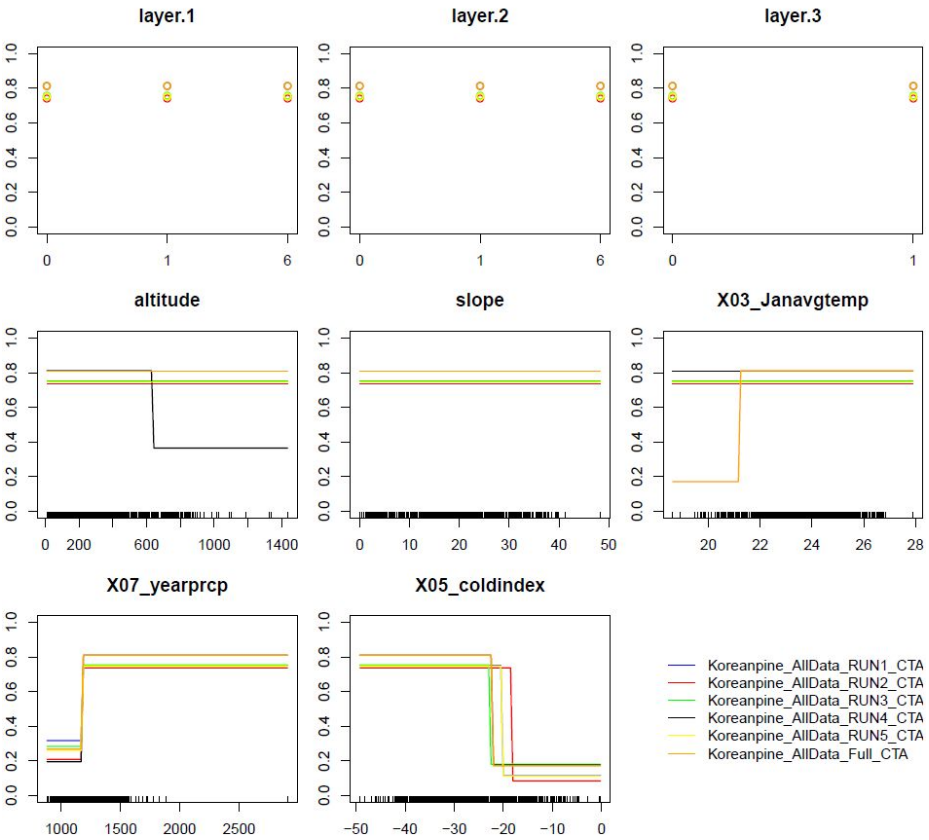


Figure 17. Response function of CTA (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

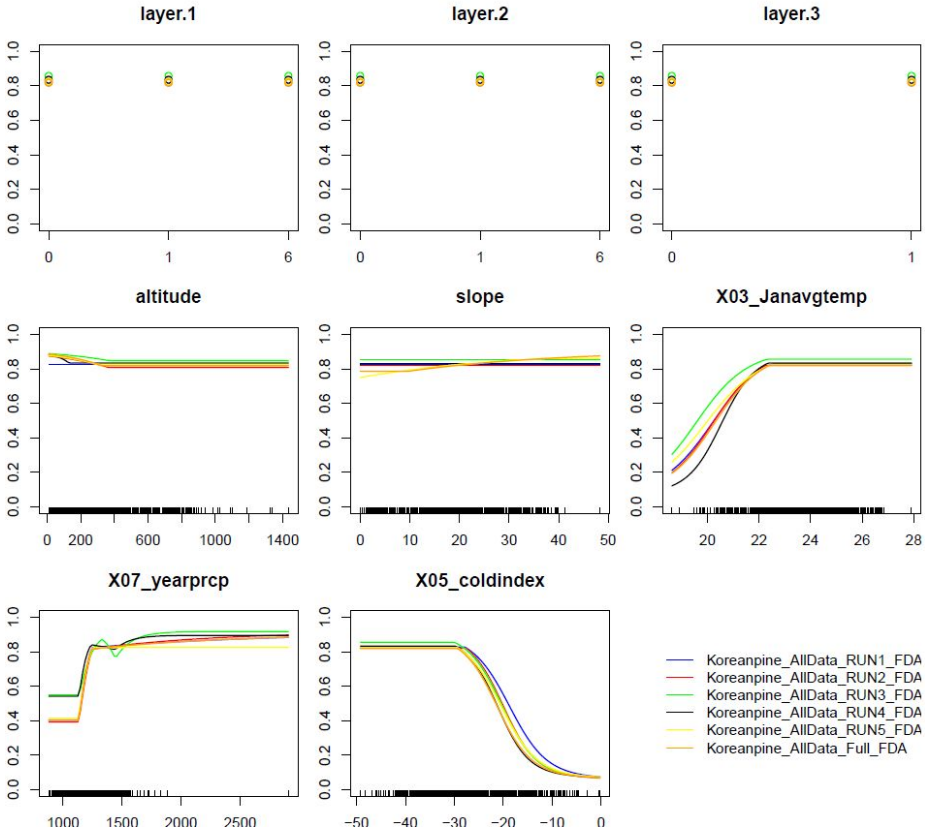


Figure 18. Response function of FDA (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

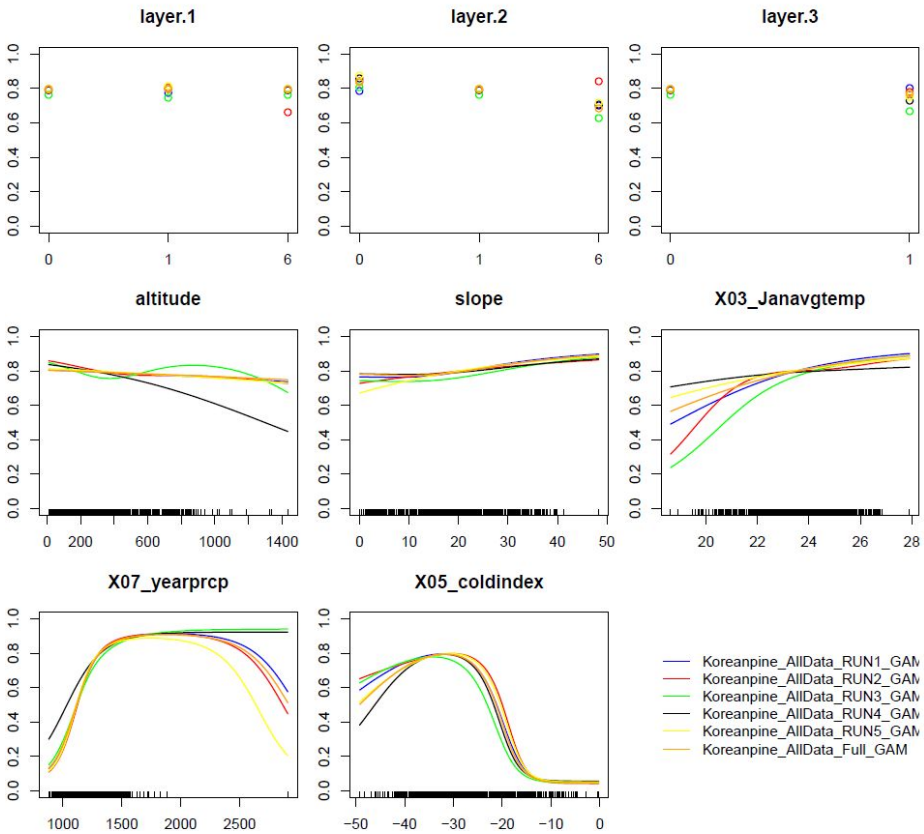


Figure 19. Response function of GAM (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

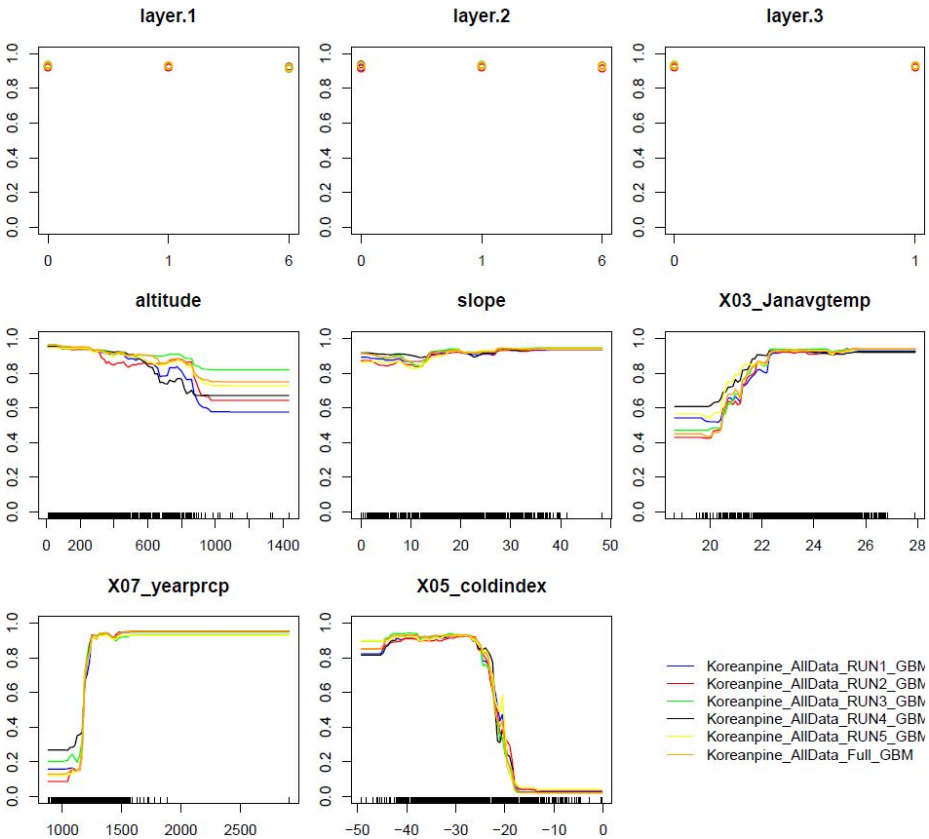


Figure 20. Response function of GBM (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

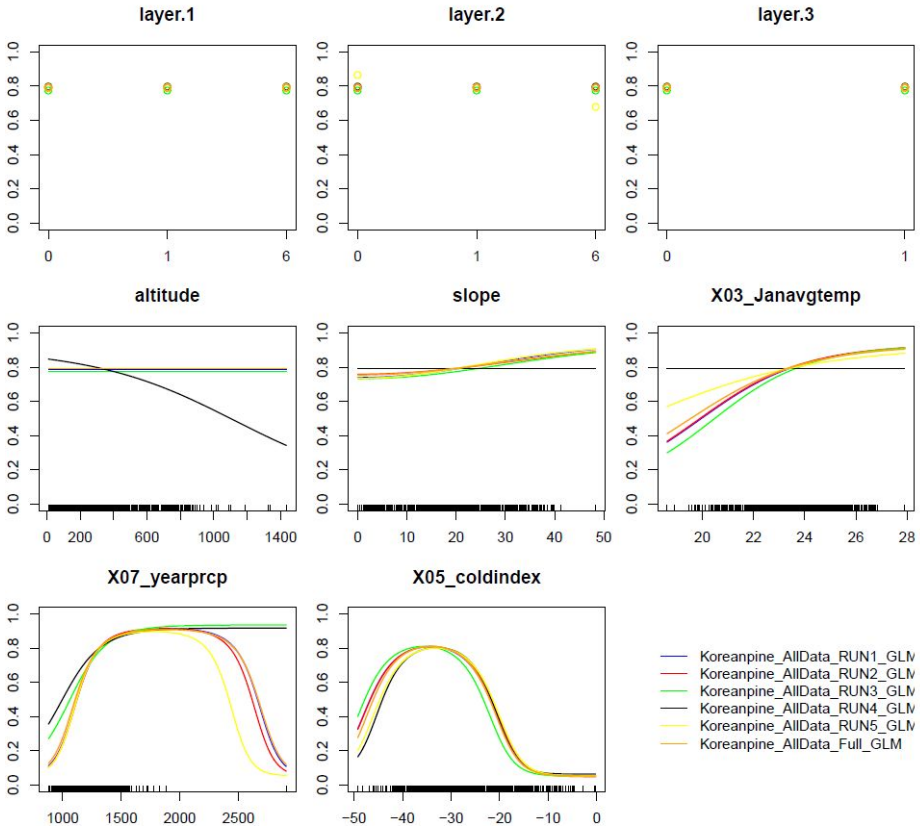


Figure 21. Response function of GLM (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

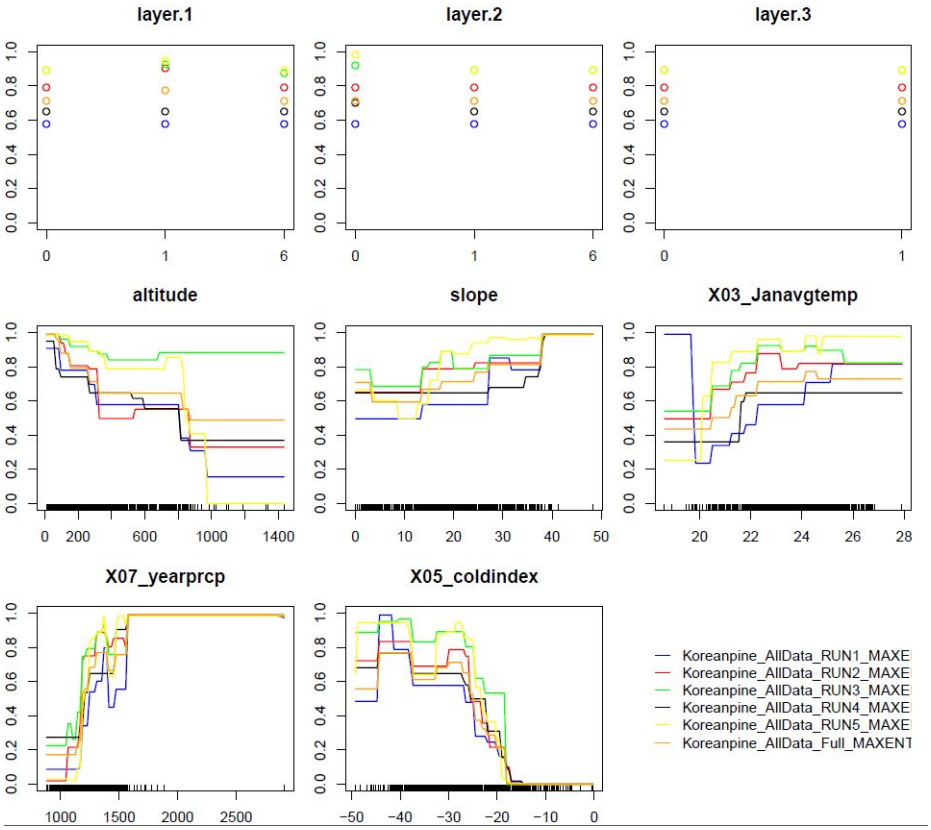


Figure 22. Response function of MAXENT (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

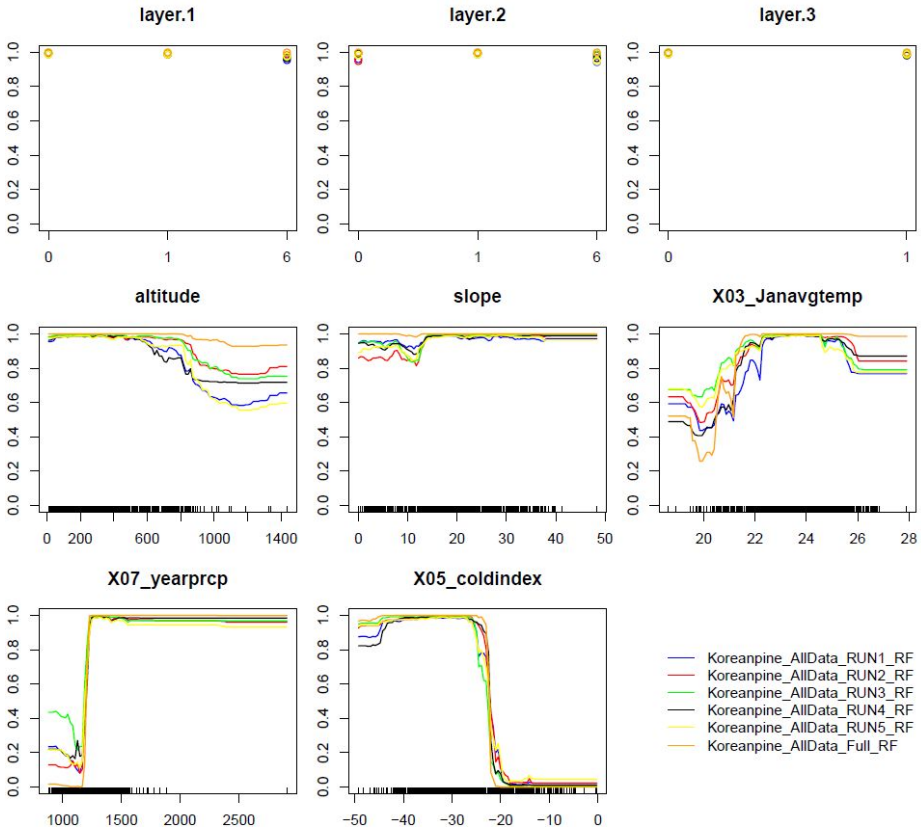


Figure 23. Response function of RF (layer1: Drainage of soil, layer2: Depth of soil, layer3: Soil type, Altitude, Slope, Janavgtemp: Average precipitation in January, Yearprcp: Yearly precipitation, Coldindex: Coldness index)

국문초록

불확실성을 고려한 잣나무의 서식 적지 분포 예측
- 종 분포 모형과 RCP시나리오를 중심으로 -

안 윤 정
조경.지역시스템공학부 생태조경학 전공
서울대학교 대학원
지도교수 이 동 근

기후변화로 인해 산림의 종 분포가 변화될 것이라고 예측되고 있다. 산림의 공익적 이익을 극대화하기 위해 산림의 관리가 중요하다. 조림수종의 경우 경제적 이익을 목적으로 하기 때문에 기후변화에 따른 조림 수종의 잠재적 적지 예측에 대한 연구가 다수 진행되었다. 특히 잣나무(*Pinus koraiensis*)는 한대수종이기 때문에 기후변화에 취약하다. 이에 본 연구는 종분포 모형과 RCP시나리오의 불확실성을 고려하여 기후변화에 따른 잣나무 분포를 예측하였다.

먼저, 모형의 입력 자료로 종의 출현 포인트, 비출현 포인트, 환경변수와 기후변수 자료를 구축하였다. 잣나무의 출현 포인트를 추출하기 위해 현존식생도와 5차임상도를 활용하였다. 현존식생도는 자연림을 조사한 자료이며 임상도는 인공림을 조사한 자료이다. 임상도는 인공림에 대한 정보를 제공하기 때문에 잣나무가 서식하기에 적합한 곳을

예측하기 위하여 임상도에서 출현 포인트 추출 시 지위지수를 고려하였다. 또한, 잣나무가 서식하고 있지 않은 산림에서 비출현 포인트를 추출하였다.

다음으로 문헌고찰을 통하여 잣나무 분포에 영향을 미치는 환경변수를 추출하였다. 추출된 변수를 대상으로 상관분석을 수행하였고 상관분석 결과 공선성이 존재하는 변수를 제외한 나머지 변수를 선택하였다. 그 후 전문가 자문을 거쳐 변수를 확정지었다. 확정된 변수를 모형에 입력하였으며 모형의 검증은 데이터 분리와 ROC검증 방법을 활용하였다. 검증된 모형을 통해 미래를 예측하였다. 마지막으로 미래 예측 결과를 종합하는 앙상블 모형을 구축하였고 최종적으로 앙상블 결과와 불확도 지도 그리고 개별모형의 결과를 중첩한 중첩지도를 도출하였다.

개별모형의 결과는 상이한 결과를 보였다. 8가지 모형 중에서 RF는 가장 뛰어난 예측력을 보였다. 반면, ANN와 Maxent는 다른 모형에 비해 과추정 되는 경향을 보였다. 선행연구를 검토한 결과 이러한 차이는 모형의 알고리즘, 입력 자료의 형태, 검증방법 때문이라고 판단된다. 앙상블모형에서는 개별모형에서 보여 졌던 불확실한 결과는 제외된 최종결과를 도출하였다.

중기미래에서의 결과의 차이는 모형에 의한 차이가 큰 반면 장기미래에서는 모형과 기후변화 시나리오에 따른 차이가 큰 것으로 밝혀졌다. 앙상블 모형의 결과는 중첩지도와 불확실성 지도를 통해 불확실성을 정량화하였다. 개별모형의 결과에서보다 앙상블 모형의 결과에서 모든 모형의 결과의 평균에 가까운 결과를 도출 할 수 있었다.

중기미래 잣나무의 분포는 남부 일부지방, 강원도 일대, 중부지방에 분포하는 것으로 예측되었다. 한편 장기미래에서는 자연림이 분포하

는 남부지방의 분포를 사라졌으며 중부지방과 강원도 일대의 분포지역도 줄어드는 것으로 예측되었다. 더 나아가, 잣나무는 조림수종이라는 것을 고려하였을 때 일부보호지역에서 조림이 제한된다. 이에 따라 앙상블 모형의 결과와 보호지역지도를 중첩하여 보호지역의 분포를 확인하였다. 결과적으로 중기와 장기 미래 보호지역과 잣나무의 분포지역은 약 20~40% 중첩되는 것으로 밝혀졌다.

본 연구에서는 기후변화를 고려한 잣나무의 분포를 예측하였다. 분포 예측 결과 미래 잣나무의 분포는 2090년대 크게 감소할 것으로 예측되었다. 또한 각 모형의 불확실성의 범위를 확인하였으며 앙상블 모형을 통해 종분포 모형의 불확실성을 줄일 수 있음을 확인하였다. 본 연구는 잣나무 분포를 예측함에 있어 종 분포 모형과 기후변화 시나리오의 불확실성을 고려했다는 점에서 의의가 있다. 연구의 결과는 향후 기후변화를 고려한 조림계획 수립 시 활용될 수 있을 것이라 생각된다.

□ *주요어 : Ensemble models, BIOMOD2, Machine learning models, Statistical models, Species distribution, Forest ecosystem*

□ *학 번 : 2013-23255*