Ph.D. Dissertation of Engineering

Assessing Uncertainties in Predicting the Changes in Forest Species Distributions caused by the Climate Change

기후변화를 고려한 산림 수종

분포변화 예측의 불확실성 평가

February 2019

Graduate School of Seoul National University

Interdisciplinary Program in Landscape Architecture

Sunyong Sung

Assessing Uncertainties in Predicting the Changes in Forest Species Distributions caused by the Climate Change

Advisor: Dong Kun Lee

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Interdisciplinary Program in Landscape Architecture in Seoul National University

January 2019

Sunyong Sung

Approved by Thesis Committee

Chair Youngryel Ryu (Seal)

Vice Chair Youngkeun Song (Seal)

Examiner Yoosang Chung (Seal)

Examiner Changwan Seo (Seal)

Examiner Dong Kun Lee (Seal)

Publications

Please note that some part of this dissertation proposal was written as stand-alone papers (see below), and therefore there is some repetition in the methods and results.

1. Sung S. et al., 2018, Assessing Effective Sampling Methods and Sample Size for Species Distribution Modeling of Korean Red Pine (*Pinus densiflora)*, Journal of Faculty Agriculture, Kyushu University, 63(2), pp.211–221

Table of Content

List of Figures

List of Tables

Abstract

Assessing Uncertainties in Predicting the Changes in Forest Species Distributions caused by the Climate Change

Sunyong Sung

Interdisciplinary Doctoral Program in Landscape Architecture

Graduate School, Seoul National University

Supervised by Professor Dong Kun Lee

The adverse impacts of climate change on forest ecosystems are expected to increase, and various measures are being proposed to reduce them. To mitigate the negative impacts of climate change with limited time and resources, and to respond effectively, it is necessary to make an accurate impact assessment based on climate change. To do so, it is necessary to understand and quantify the uncertainties that are inevitable in climate change impact assessment.

The concept of uncertainty, which has been mentioned since the fourth report of the Intergovernmental Panel on Climate Change, is specified in the Fifth Impact Assessment Report and is used as a concept to aid decision making. In Korea, efforts are being made to quantify uncertainties in

assessing the impacts of climate change. However, these studies are still in the early stages, are limited in scope, and do not consider uncertainties in various aspects.

Therefore, in this study, we analyzed the causes of uncertainties that may occur due to climate change by: 1) measuring the effectiveness of sampling methods and sample size, 2) evaluating the uncertainties in model performance and spatial distribution due to Species Distribution Model (SDM) algorithms, and 3) considering the uncertainties if involved in assuming competition among major species, applying four Representative Concentration Pathway (RCP) scenarios to potential distribution ranges.

To measure the effectiveness of the sampling methods, three sampling methods and seven different sample sizes were considered for the one-way t-test. As a result of the one-way t-test, stratified random sampling methods are shown to well represent the population. In addition, if the sample size exceeds a certain number, for this study, 200 samples, the performance of SDMs does not significantly increases.

We applied eight SDMs that were either statistically based or machine learning based algorithms to model the potential distribution of major species in Korea; the performance of the models differed according to the algorithms.

To test the performance of SDMs, the area under the curve (AUC) value and True Skilled Statistics (TSS) value were applied. Machine learning models, especially the random forest (RF) model showed excellent performance while statistical based models (Generalized Linear Model (GLM), Generalized Additive Model (GAM)) showed average performance. When we verified uncertainties in spatial distribution, with thresholds matching the current area of the major species, the uncertainties in the spatial distribution was significant. Ensemble methods need to be applied to minimize uncertainties in the spatial distribution of SDMs.

To consider uncertainties in the competition among major species, the random forest algorithm and Global Agro-Ecological Zones (GAEZ) classification were applied. Modeling results revealed that the multi-species model included higher uncertainties. However, single species models can not include the climate zone changes that we expect in RCP the scenarios. Thus, we need to include the potential introduction of forest species that are suitable in different climate zones.

Through this study, when we establish management strategy for climate change mitigation and adaptation, uncertainties in each step if we predict potential distribution of forest species can be applied to prioritize management target. This can reduce uncertainties in management strategy as

well as find effective monitoring points for counteract adverse changes due to the climate change.

Keywords: Uncertainty, Species Distribution Change, Sampling Methods, BIOMOD2, Random Forest, Forest Management
Student Number: 2013-30714

I. Introduction

The adverse impacts of climate change to human society have increased since the advent of industrialization. Extreme events, which has been caused by climate change damages, significantly damages not only the ecosystem but also human settlement (IPCC, 2014a). To minimize the adverse impact of climate change, various studies have been conducted to establish regional and national polices since the 1990s (IPCC, 2014b). In particular, studies that quantify the future impacts of climate change are being used effectively to mitigate these effects and develop countermeasures to adapt them.

While establishing countermeasures on the adverse impact of climate change, forests are drawing attention since they can play a role as a stepping stone to achieve mitigation and adaptation simultaneously (Lal, 2005). Thus, research on changes in forests for climate change has been actively conducted (Bonan, 2008). It is important to promote the role of forests as carbon sinks because the forest contributes 25% of the carbon dioxide mitigation that is emitted by fossil fuels. In addition, forests contribute to adaptation by mitigating the risks associated with heavy rainfall due to climate change and promoting biodiversity by conservation of habitat. Therefore, discussions are underway to preserve the forest ecosystem and restore the areas expected to be degraded under climate change (IPCC, 2003).

To understand the role of the forest ecosystem in counteracting the adverse impact of climate change, it is important to understand the uncertainties within the impact assessment. The concept of uncertainty originated in the fourth report of Intergovernmental Panel on Climate Change (IPCC). In the fifth impact assessment report, uncertainty was used as a concept to support decision-making processes (IPCC, 2003). In Korea, efforts have been made to quantify the uncertainties in assessing the impact of climate change (Kim et al., 2018). However, these studies are still in the early stages, limited in scope, and do not take into account uncertainties in modeling processes (Kryazhimskiy et al., 2015).

To quantitatively evaluate the impact of climate change on the forest ecosystem, it is necessary to comprehensively understand the various factors affecting these ecosystems as well as the expected changes due to changes in climate. The major impacts of climate change on forest ecosystems are largely divided into changes in productivity and forest growth, changes in forest species, species composition, and increases in disturbances (Lindner et al., 2014). Among these researches on the distribution of forests species are important as their distribution can be considered a baseline for assessing the various effects of forests on future climate change, such as carbon sequestration, productivity of forest species, and biodiversity of ecosystems.

All impacts on forest ecosystems are closely linked to climate change management plan.

To make an effective and reliable management plan for climate change, it is important to consider the uncertainties that may arise in predicting the distribution of forest species. When we are predicting changes in forest species, we need to consider uncertainties in various steps: first, in collecting input data for modeling the potential distribution of forest species. Because we cannot survey all forest species on a national scale. Next, the application of models could cause uncertainties, as differences in model algorithms derive different interpretations of the current distribution of forest species. Finally, the temperature and precipitation changes in different Representative Concentration Pathway (RCP) scenarios could cause different ranges of impacts on species distribution.

However, most of the related studies on the changes in forest species and distribution have been conducted using single models and single sampling methods (Pearson and Dawson, 2003). In this case, using different models and different sampling methodologies makes the interpretation of the variation of in results difficult (Beale and Lennon, 2012). Therefore, it is necessary to consider the uncertainties that are caused from the choice of models and sampling methods (Ananda and Herath, 2009; Hannemann et al., 2016). At the same time, it is necessary to recognize the uncertainties from different temperatures and amount of precipitations in RCP scenarios that could causes different interpretations of climate change impact.

Therefore, this study aims to quantitatively assess the uncertainties in predicting the distribution of major forest species under climate change. To consider the uncertainties in modeling the distribution of future forest species, this study will consider 1) effective sampling methods 2) various algorithms of models and spatial distribution of species, and 3) multiple RCP scenarios used to comprehensively evaluate the impact of forests under climate change on a national scale. In addition, the level of uncertainties will be different in different factors of modeling changes of major forest species under climate change. This will the quantify impact of climate change and utilize our result as a reference data for decision making in planning future forest adaptation and management to encounter the adverse impact of climate change.

II. Literature Review

1. Definition of Uncertainty

Uncertainties are always involved in our understanding of the current status and modeling of future changes. Uncertainty is defined as a "state without adequate information", rather than as knowing or not-knowing (Walker et al., 2003). To systematically identify the source of uncertainty, it can be approached in terms of 1) the “location” where uncertainty occurs, 2) the “level” of uncertainty, and 3) the “source” of uncertainty (Walker et al., 2003). First, the location of uncertainty includes the uncertainty in the context, model, input, and parameter.

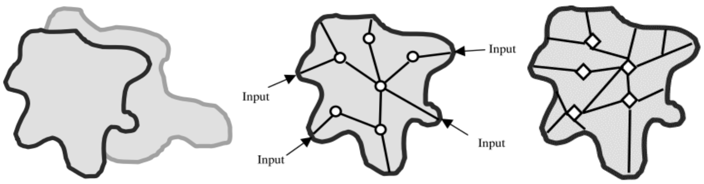


Figure 1 Uncertainties from context (left), and model structure and inputs (right) (Walker et al., 2003)

The uncertainty in the context will depend on how the model's scope is set. In the model, it is difficult to analyze all phenomena; therefore, only a part of reality is reflected. The uncertainty in the model is caused by the diversity of the model structure, which explains the reality and uncertainty in the input data.

Next, the difference in the level of uncertainty starts from statistical uncertainty to scenario uncertainty, recognized ignorance, and total ignorance (Walker et al. 2008). From a deterministic point of view, uncertainties can be explained statistically, and scenario uncertainties can be solved through scenario design. Perceived uncertainties can also be reduced through research and experimentation. However, uncertainties beyond human perception are indeterminacies.

Finally, the causes of uncertainty can be classified into two categories: variability and lack of knowledge (Van Asselt and Rotmans, 2002). Uncertainty due to variability implies uncertainty in human and natural systems inherent in society, economy, and technology. This includes uncensored human behavior, unpredictable social uncertainty and unpredictable effects of new technology. At the same time, uncertainties arising from lack of knowledge can be reduced by research or empirical observations.

2. Modeling Potential Impact of Forest Species Distribution under the Climate Change

Climate change includes the increase in mean temperature, changes in precipitation patterns and changes in CO_2 concentration. Climate change is linked with phenology, mortality, disturbances from invasive species, extreme events. Changes linked with productivity and the potential distribution of species are connected to the composition of species in the forest and the optimal ranges of species under future climate change. Productivity is related to carbon sequestration and the potential distribution of species is linked with biodiversity. Both are connected to the climate change adaptation plan.

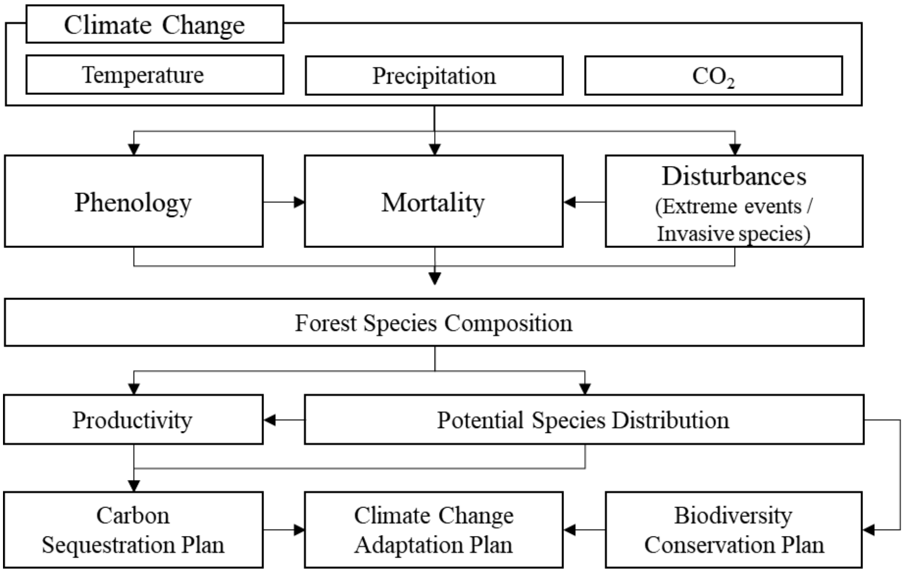


Figure 2 Impact of climate change to forest ecosystem and link with national plans

The distribution of climatic zones was first published in 1884, taking into account the distribution of forests using temperature and precipitation patterns (Belda et al., 2014). Choi et al. (2011) estimated the potential distribution of forests using the Warmth Index (WI), Minimum Temperature of the Coldest Month Index (MTCI) and Precipitation Effective Index (PEI). In addition, studies have been conducted to predict the impact of climate change on future distribution by using the Growing Degree Days (GDD) to separate the climate zones (Gang et al., 2015). In addition, related studies have been conducted to predict future distribution changes using BIOCLIM data and species distribution models (Beaumont et al., 2005; Kriticos et al., 2012; Shin et al., 2012).

A few studies applied the site index to model the future distribution of forest. The site index can be defined as the average height of a dominant tree or co-dominant tree in a given age (base age). Applying the site index has the advantage of providing a comprehensive approach to the effects of local environmental factors such as topography, soil characteristics, and climate, which are important environmental factors for forest growth (Carmean et al., 1989). However, since the distribution of the site index differs for each species and region, it is necessary to derive an appropriate index function for the region and species. The potential distribution of future forests under climate change is modeled by analyzing the site index of the area where the species are distributed and estimating the site index based on related climate and environmental variables. Thereafter, the ability of species to adapt to changes in status index can be analyzed (Korea Forest Research Institute, 2014).

In Korea, a study was conducted to predict the potential forest distribution through a multinomial logit model using topography, maximum monthly temperature, summer average precipitation, soil base saturation, and soil organic matter content (Shin et al., 2012). In these studies, the potential forest distribution was predicted based on the appropriate range of species, but uncertainties due to climate scenarios are inherent because of the consideration of only a single scenario. In addition, since the future forest distribution is predicted by using the present optimum distribution of the indices, succession of forest cannot be considered.

In modeling the potential distribution of species due to climate change, some studies have used multiple indices to estimate the appropriate ranges (Véga and St-Onge, 2009), while other studies have modeled the potential distribution of species using environmental variables at occurrence point in the Species Distribution Model (SDM) (Guisan and Thuiller, 2005; Gutiérrez et al., 2016).

The most widely used methods for modeling the potential distribution of species are the ecological niche model and the habitat suitability model (Kwon, 2014). The ecological niche model predicts the future distribution of species, based on the idea that the current distribution is the most appropriate ecological niche for the species and predicts the potential distribution of the species through the assumption that this ecological niche will not change (Elith and Leathwick, 2009; Guisan and Zimmermann, 2000; Pearson and Dawson, 2003; Wang et al., 2015). Therefore, it is essential to accurately identify the niche species for modeling the potential distribution of species by SDMs.

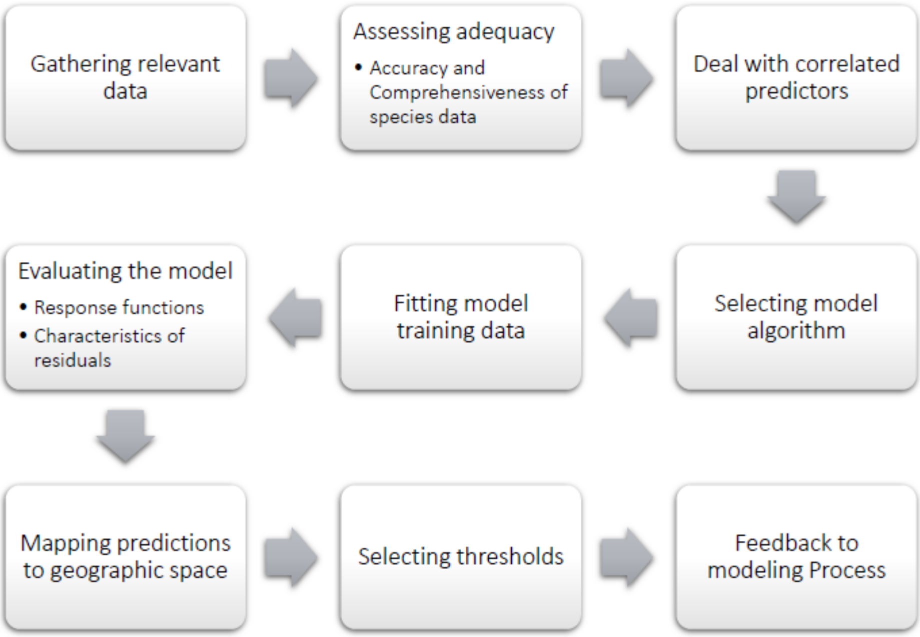


Figure 3 Basic flowchart of SDM (modified from Elith and Leathwick, 2009)

3. Cause of Uncertainties in Species Distribution Modeling

The degree of uncertainty in estimating the impact of climate change increases with the steps of impact assessment (Jones 2000; Wilby and Dessai 2010). Uncertainties arising from climate change impact assessment are subject to uncertainties in: 1) future societies, 2) greenhouse gas emissions, 3) climate models, both regional and global, and 4) impact assessment models. It is necessary to identify and quantify the main sources of uncertainty that can occur at each step to provide information at a reliable interval. In the quantitative assessment of the impact of climate change, uncertainties in climate change can be reduced if uncertainties are considered at each assessment step.

Table 1 Different uncertainties from climate change impact assessment and description

Source of uncertainties	Description
Emission Scenarios	Climate change are based on the future carbon dioxide emission from human activity. Policy implication could cause uncertainties
Global Climate Change	Climate change scenario can exceed expected range of uncertainties
Regional Climate Model	From downscaling global climate to regional scenario, downscaling method and parameters can cause uncertainties
Modeling	Modeling framework can cause uncertainties

3.1. Uncertainties in Climate Change

There are cause of uncertainties within climate scenarios. First, the climate scenario itself has uncertainties as their projections are based on socio-economic scenarios. The global community tries to reduce greenhouse gases, which can affect future climate scenario projections. These changes can alter future climate projections (Knutti and Sedláček, 2012). Next, the climate scenario has assumptions in how the greenhouse will change with different model algorithms and scenarios. Even if we apply all climate scenarios for assessing the potential impact of climate change, there are some changes that exceed the range of model projections (Jones, 2000). Therefore, we need to consider the potential climate change within quantifiable ranges of current climate scenarios.

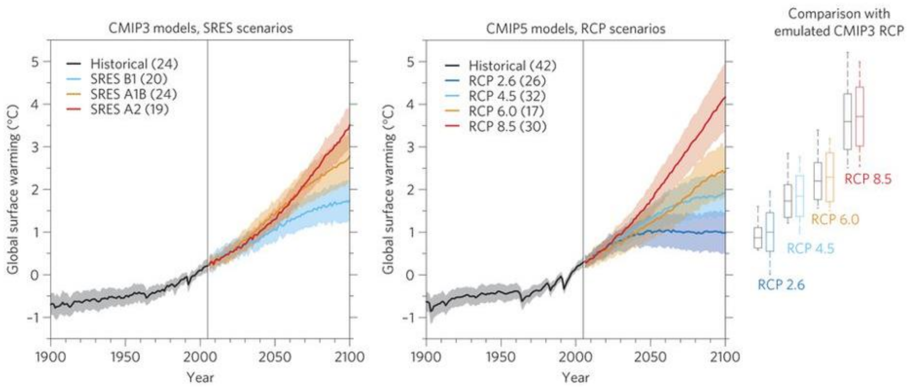


Figure 4 Global temperature change and uncertainty (Knutti and Sedláček, 2012)

3.2. Uncertainties in Sampling Methods

There is a limit to all the information that can be acquired for predicting a phenomenon, and it is important to extract an appropriate sample that can represent the phenomenon by effectively utilizing limited information and resources. There are four major problems to be considered in extracting the samples: size of the sample, design method of the sample, representative value that can characterize the sample and variation and the confidence interval of the sample that should be set (Hengl, 2009).

There are two methods for extracting samples to be used in social science research: probability sampling and non-probability sampling. The probability sampling method uses a random method to extract sample components when the probability, that all the research subjects are extracted as a sample, is known; otherwise, a non-probabilistic method is used. The probability sampling method includes simple random sampling, systematic sampling, stratified sampling, and cluster sampling. Non-probability sampling is a non-random sampling method, e.g. judgment sampling, and quota sampling.

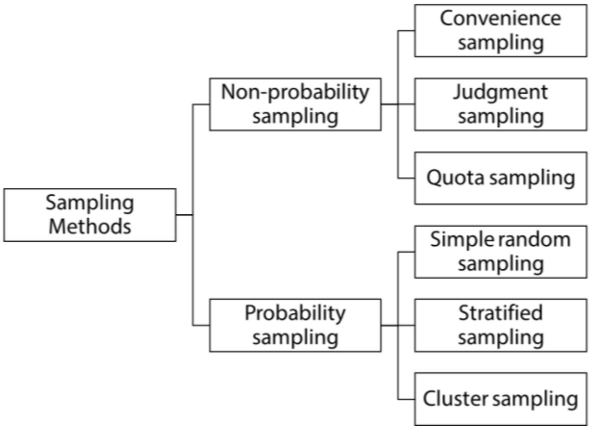


Figure 5 Sampling methods in forest

However, dealing with spatial data requires a different approach than the one dealt with in social sciences. Studies related to geography, such as vegetation and landscape, assume that everything is related but 'closer is more relevant than far'[1], so we can explain the phenomenon by extracting the appropriate sample. Basically, a method of extracting a spatial sample utilizes methodologies such as extracting a sample at intervals or randomly extracting a sample. For spatial analysis, it is necessary to use different sampling methods depending on the type and purpose of the data to be used. These sampling characteristics could cause uncertainties in modeling the potential distribution of forest (Sun–Yong Sung et al., 2018).

[1] "Everything is related to everything else, but near things are more related than distant things (Tobler, 1970)"

In forest modeling, generally four different sampling methods are applied: simple random sampling, systematic sampling, cluster sampling and stratified sampling (Food and Agriculture Organization, 2004; Figure 6). For effective understanding of forests, sampling design should consider spatial balance, uncertainties and the cost of survey in the sampled area. In addition, the size of the sample is important as a sample size is too small will increase uncertainty while a sample size that is too large will increase the cost of survey unnecessarily high.

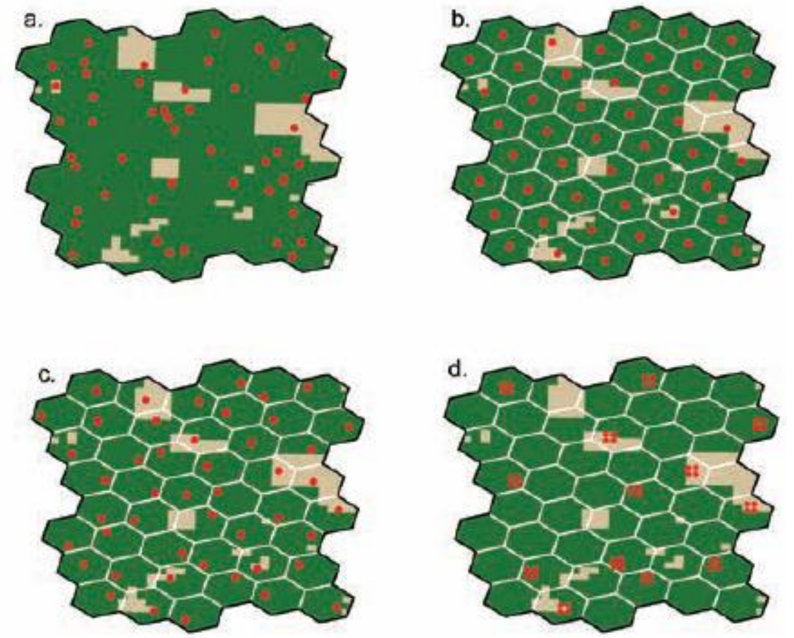


Figure 6 Basic sampling patterns in forest inventory (Food and Agriculture Organization, 2004)
(a) simple random sampling design, (b)aligned systematic sampling design
(c) unaligned systematic sampling design (d) unaligned, clustered, systematic design

3.3. Uncertainty in Species Distribution Model Algorithm

Recently, a number of studies have been conducted to predict the distribution of forests using the SDMs, and these studies identify the strengths and weaknesses of the models according to their respective characteristics (Franklin, 2010a). The SDMs can be divided into a statistical based model and a machine learning model. The statistical models can be classified as Generalized Linear Models (GLM) and Bayesian model. Machine learning models include decision tree (DT), Artificial Neural Networks (ANN), Genetic Algorithms (GA), and MAXENT models.

Table 2 Types of SDM and their performance (modified from Franklin and Miller, 2010)

Classification	Model	Performances
Statistical	Generalized linear models (GLM)	Effective global modeling methods; performs well with adequate data
	Generalized additive models (GAM)	Performs well when not over-fit to data
	Multivariate adaptive regression splines (MARS)	Performs similarly to and slightly better than GLM
	Bayesian Modeling	Not widely compared with other SDM
	Spatial autoregressive models (SAR)	Tend to perform better than non-spatial models, but limited spatial density of data
Machine learning	Decision Trees (DT)	Single DT perform poorly compared to other methods
	Random forest (RF)	Ensemble DTs tend to have good predictive performances. And provides importance of predictors and response functions
	Artificial neural networks (ANN)	Good performances when used by skilled performers.
	Genetic algorithm (GA)	Poor performance in comparison with other methods
	Maxent	Performs well in data-poor situations.
	Support vector machines	Only a few SDM applications to data

Previous applications of SDMs used single model or multi model approaches to quantify uncertainties from different model algorithms (Jarnevich et al., 2018; Sunyong Sung et al., 2018). Researchers have suggested several methods to minimize uncertainties (Kim et al., 2018; Wiens et al., 2009). However, it is very difficult to propose one optimum algorithms or method for modeling the potential distribution of species. Therefore, it is necessary to understand the uncertainties in the SDMs.

4. Summary

In considering the uncertainties involved in modeling the potential distribution of forest species according to future climate change, we reviewed the definition of uncertainty, the modeling of the potential distribution forests using different algorithms, causes of uncertainties within SDMs. However, the quantification of uncertainties in modeling potential distribution of forests species is limited (Beale and Lennon, 2012).

There are uncertainties in modeling the potential distribution of forest with SDMs (Figure 7). To apply different SDMs for predicting the potential distribution of forest species, we need to collect the presence/absence point. In collecting the presence point, we cannot select all sampling points or presence points as data is not available in some countries. In these cases, uncertainties can arise in selecting samples for SDMs. In addition, the climate change has uncertainties in projecting future temperature and precipitation changes, since there are significant changes in greenhouse gases, based on human activities, which have uncertainties.

Similarly, there are uncertainties in selecting different SDMs, as each has different algorithms. When we interpret the result of SDMs, model performance and spatial distribution should be considered carefully. Modeled results do not include species composition in terms of competition among the major forest species. Thus, the uncertainties in modeling potential distribution should be carefully examined in managing the forest ecosystem effectively. Then, a management strategy should be proposed for effective response to the potential adverse impact of climate change.

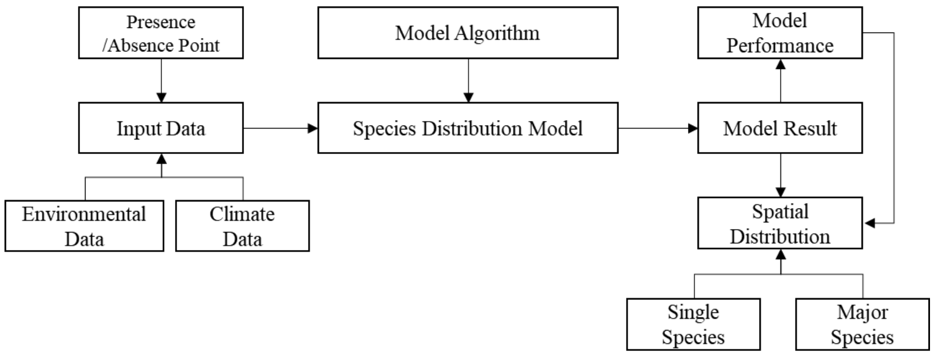


Figure 7 General framework of species distribution modeling and uncertainties

III. Scope and Methods

1. Study Scope

1.1. Research Flow

The research flow of this study is shown in Figure 8. First, the purpose of this study is contained in the introduction. In the literature review, the definition of uncertainty, related studies on modeling the potential distribution of major forest species and the causes of uncertainties are reviewed. Then, we set the study scope and describe materials and methods for quantifying uncertainties in the modeling of the potential distribution of major forest species.

In the result and discussion section, the potential distribution of major forest species is presented for quantifying the different consequences in selecting RCP scenarios. Then, the uncertainties in selecting sampling methods and different SDMs are described. Based on the results for quantifying uncertainties, I suggested a forest management strategy to effectively counteract the forest ecosystem to minimize the negative effects that could be derived by climate change.

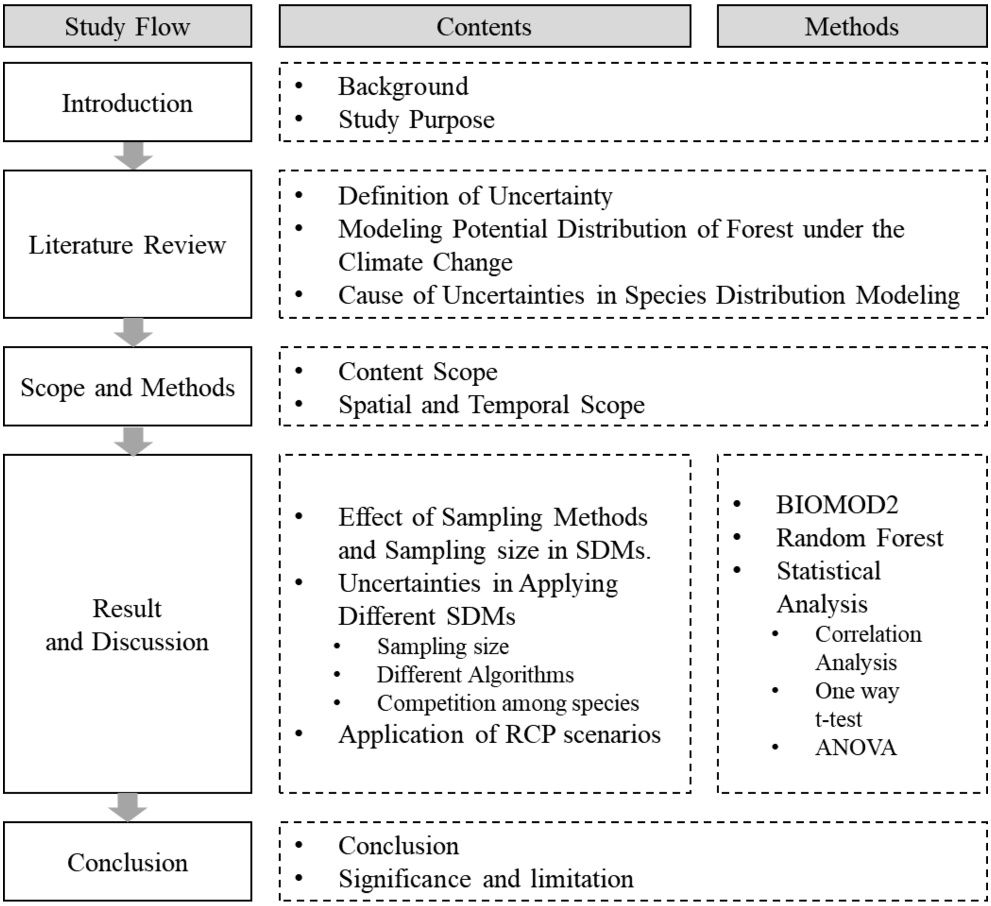


Figure 8 Research flow

1.2. Spatial and Temporal Scope

To predict the potential distribution of forest species under climate change, the scope of study was limited to the inland areas of South Korea (Figure 9). Because South Korea includes diverse vegetation zones as its terrain is complex, and the forests species are in various climatic zones, it is expected that the effects on forests species due to climate change will be significant.

At the same time, it is possible to obtain more accurate and meaningful impact assessment results because it is possible to obtain high-resolution data essential for predicting future climate change impact.

We set the current climate conditions from 2001 to 2010. We set the target period for analyzing long-term changes in forests to the 2090 's from 2091 to 2100 as the forest ecosystem changes in a relatively longer period. In this study, the effects on forests species due to future climate change should be predicted spatially. Thus, this study was conducted based on the resolution of 1 km × 1 km, which is the highest resolution of climate data provided by the Korea Meteorological Administration.

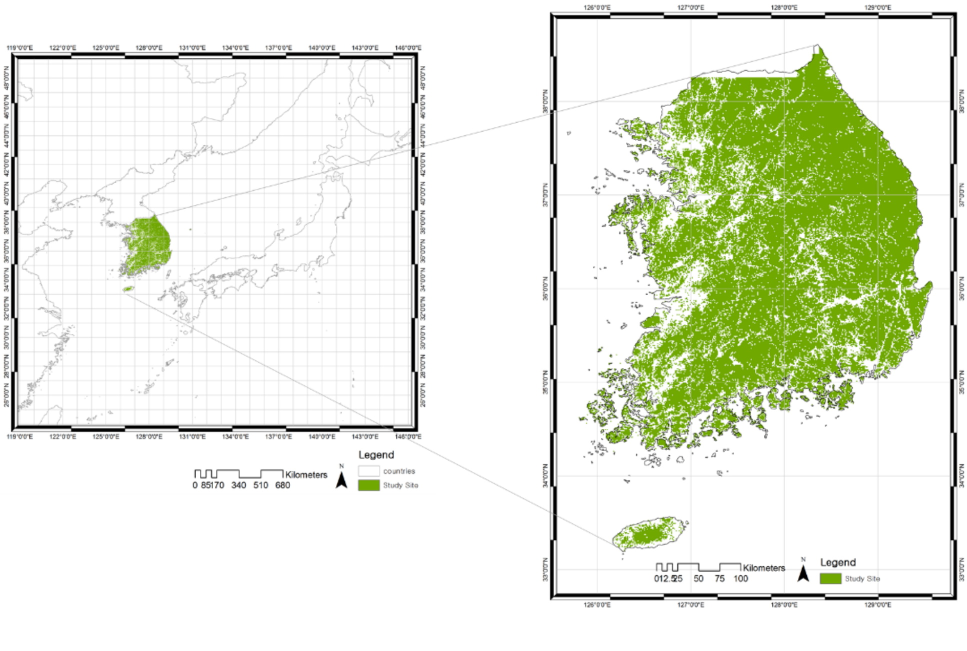


Figure 9 Study site

2. Materials and Methods

2.1. Materials

We used four categories of environmental variables—vegetation, climate, topography, and soil (Table 3)—in the SDMs. Vegetation data were based on a 1:5000 detailed forest map from the Korea Forest Service (KFS). Forest age, type, and density were collected for each forest stand. Climate data were derived from the Korea Metrological Administration (KMA), and all climate datasets were statistically downscaled to 1 km × 1 km resolution.

Table 3 List of environmental variables

Category	Variables	Data type	Resolution	Source
Vegetation	Forest Type	Feature (SHP)	1:5000	Korea Forest Service
Topography	Altitude	Raster (TIFF)	30m	National Geography Information Institute
	Slope			
	Radiation			Ministry of Environment
	Distance from water			
	Distance from sea			
Soil	Soil Depth	Feature (SHP)	1:25000	Korea Forest Service
	Soil Organic Matter Content in Layer A			
Climate	Warmth index	Raster (TIFF)	1 km × 1 km	Korea Metrological Administration
	Isothermality			
	Min temperature of coldest month			
	Precipitation of wettest month			
	Precipitation of driest month			
	Climate Zone		30m	Korea Forest Service

From the monthly temperature and precipitation data, we generated 19 bioclimatic variables averaged over 10 years, from 2001 to 2010, to match the WorldClim database (Hijmans et al., 2005). The BioCLIM variable is used by Hijmans et al., (2005) to express ecologically meaningful variables (e.g., annual climate and seasonality, extreme climatic and limiting requirements) utilizing monthly precipitation and monthly temperatures, which are widely used in the analysis of biological species distribution (Ahn et al., 2015; Beale and Lennon, 2012; Beaumont et al., 2005). We calculated future (2091-2100) BIOCLIM data by using the Dismo Package of R.

Table 4 List of BioClim variables (Kriticos et al., 2012)

Variable Number	Variable
Bio01	Annual mean temperature (°C)
Bio02	Mean diurnal temperature range (mean (period max-min)) (°C)
Bio03	Isothermality (Bio02 ÷ Bio07)
Bio04	Temperature seasonality (C of V)
Bio05	Max temperature of warmest week (°C)
Bio06	Min temperature of coldest week (°C)
Bio07	Temperature annual range (Bio05-Bio06) (°C)
Bio08	Mean temperature of wettest quarter (°C)
Bio09	Mean temperature of driest quarter (°C)
Bio10	Mean temperature of warmest quarter (°C)
Bio11	Mean temperature of coldest quarter (°C)
Bio12	Annual precipitation (mm)
Bio13	Precipitation of wettest week (mm)
Bio14	Precipitation of driest week (mm)
Bio15	Precipitation seasonality (C of V)
Bio16	Precipitation of wettest quarter (mm)
Bio17	Precipitation of driest quarter (mm)
Bio18	Precipitation of warmest quarter (mm)
Bio19	Precipitation of coldest quarter (mm)

We also constructed a WI and a coldness index (CI), which are considered efficient indicators for monitoring interactions between climate and species distribution (Kira, 1945; Yim, 1977) (Equations 1, 2). The WI was calculated for months in which the temperature (t) was greater than 5℃, and the CI was calculated for months in which the temperature was less than 5℃. We used climate zone data derived from the KFS for selecting strata for the stratified sampling method.

$$\text{Warmth index (WI)} = \sum (t - 5) \tag{1}$$

For months in which t > 5℃

$$\text{Coldness index (CI)} = -\sum (5 - t) \tag{2}$$

For months in which t < 5℃

Topographical layers included altitude, slope, and aspect. These datasets were derived from digital elevation models (DEMs) from the National Geography Information Institute (NGII). A land cover map from the Ministry of Environment (ME) was used to extract land cover data. Distance from water and distance from the sea (Schulze, 2005) were calculated using Euclidian distance. Soil depth and soil organic matter content in the A-horizon were extracted from a Korean soil forest map (Brady, 2008). All environmental data were resampled with 1km by 1km resolution for modeling with the ArcGIS resampling tool. Environmental variables that have discrete characteristics are resampled with the nearest algorithm. On the other hand, environmental variables are resampled with bilinear resampling methods.

We conducted a correlation analysis in R to identify auto-correlation among the environmental variables, and environmental variables were selected with respect to multi-collinearity. If Pearson correlation coefficients were larger than 0.7, we removed relevant variables from the list[2] (Dormann et al., 2013). We also conducted a literature review to select variables potentially important for the species distribution (Nakao et al., 2014; Park et al., 2016; Takahashi and Okuhara, 2012).

2.2. Measuring Uncertainties in Modeling Potential Species Distribution

In modelling the potential distribution of major forest species, there are several factors that could cause uncertainty. First, when input data are prepared for modeling, three sources of uncertainty as can be categorized as environmental data, collecting presence/absence point data and climate data for projecting the future distribution of major forest species. Second, while

[2] Please refer appendix for correlation analysis result

applying the species distribution model, the different model algorithm could cause uncertainty. Finally, modeled results should be carefully considered as the spatial distribution can be different even though they have similar performance. In this study, the uncertainties in environmental data (e.g., topographic variables, soil parameters) were not considered because they included large uncertainties linked with socio-economic changes such as land use change and planning.

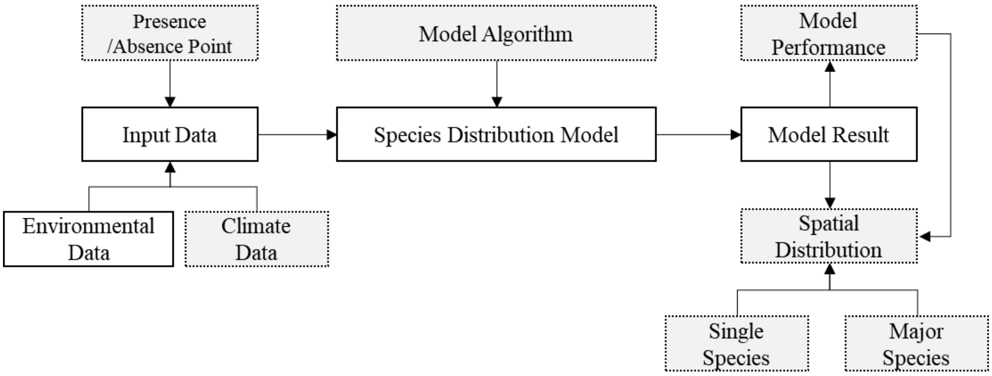


Figure 10 General framework of species distribution modeling and uncertainties (Considered uncertainties in this study were shaded)

2.2.1. Sampling Size and Methods

We applied 1:5000 forest inventory map to collect sampling point for species distribution modeling. In 1:5000 forest inventory map, the natural forest and artificial forest were divided with combination of land registration map, afforestation and reforestation map, aerial photos and field survey data by KFS (Korea Forest Research Institute, 2012). In this study, a sample was extracted, limited to the natural forest of 0.5 ha or more, which is designated as the minimum size of the forest in 1:5000 scale forest map. In addition, a natural niche of forests species can be selected by selecting samples in natural forest only. We randomly created one sample point in each forest stand from the forest inventory map.

To test the effects of sample size in species distribution modeling, we used the following sample sizes: 30, 50, 100, 200, 500, 1000, and 3000. Then, we compared three sampling methods for testing the performance of SDMs: simple random sampling, stratified random sampling, and area-weighted sampling. In simple random sampling we randomly selected samples among the created points in each forest stand. In stratified random sampling. we set strata by the climate zone in the forest soil inventory map. In area-weighted sampling, we selected points from the largest area stands in decreasing order up to the number of points selected. For example, if we collected 30 samples, we took one sample from each of the 30 largest forest stands. The sampling design tool and SQL query of ArcGIS were utilized. We conducted a one-way t-test for each environmental variable used in species distribution modeling to validate which sampling methods well represented the population.

2.2.2. SDMs Algorithm

The package Biomod2 (version 3.1.64) in R (version 3.1.2) was used to model the distribution of Korean red pine (R Core Team, 2014; Thuiller et al., 2009), which enabled us to run 10 cutting-edge species distribution modeling techniques to describe and model the relationships between Korean red pine and its environment. Biomod2 uses the ecological niche of a particular species, based on environmental variables, such as temperature, precipitation, and altitude, to project a potential habitat based on current or future environmental variables (Thuiller et al., 2015).

There are two categories of SDMs in BIOMOD2: statistically based models and machine learning based models (Table 5). Generalized linear models (GLMs), generalized additive models (GAMs), and multivariate adaptive regression splines (MARS) are all statistically based models. Machine learning based models include the generalized boosted regression model (GBM), classification tree analysis (CTA), artificial neural network (ANN), rectilinear envelope like BIOCLIM (SRE), flexible discriminant analysis (FDA), random forest (RF), and maximum entropy (MAXENT). Of these, eight models were used for analysis (the SRE and MARS models were excluded as they cannot handle categorical variables).

Table 5 Characteristics of eight species distribution models and their relative performance (revised from tables in (Franklin, 2010b; H. G. Kim et al., 2015; Thuiller et al., 2010)).

Category	Model	Characteristics	Performance
Statistical	Generalized linear model (GLM)	Flexible modern regression models	Effective global modeling methods
	Generalized additive model (GAM)	Multiple regression but with curve fitting splines or other methods	Performs slightly better than GLM
Machine learning based	Artifi1cial neural network (ANN)	Nonlinear model Using concept of artificial neural network	Performance sometimes worse than statistical model
	Maximum entropy algorithm (MAXENT)	Nonlinear model Using concept of maximum entropy Validated by ROC curve	Performs well in data-poor situation
	Random forest (RF)	Estimate many tree models based on subset of data and averaging result	Ensemble of decision tree have good performance
	Generalized boosted regression model (GBM)		
	Flexible Discriminant Analysis (FDA)	Classification based on mixture models	-
	Classification tree analysis (CTA)	Divisive model	Single decision trees perform poorly

To analyze the performance of SDMs, we used the area under the receiver operating characteristic (ROC) curve. The ROC curve is an effective method to determine the relationship between the false positive fraction (1-specificity) and the sensitivity for a range of thresholds. A good model has a curve that maximizes sensitivity for low values of 1- specificity (Neovius et al., 2004). The area between the 1:1 line and the curve represents the model performance, and this value is called the area under the curve (AUC). Additionally, the AUC is an effective model evaluation index and is independent of prevalence(Franklin, 2010b). We considered AUC values 0.9–1.0, excellent;

0.8–0.9, very good; 0.7–0.8, good; 0.6–0.7, average; and 0.5–0.6, poor (Hansson et al., 2005). An analysis of variation (ANOVA) was conducted using SPSS 18.0 to test the differences in AUC among the sampling methods and SDMs(SPSS Inc, 2009). We also applied Ture Skill Statistics (TSS) to measure performance of species distribution model. TSS corrects the dependency on prevalence while maintaining the advantages of kappa. If TSS is lower than 0.4, it is considered poor accuracy, 0.4-0.6 is moderate, 0.6-0.8 is good and above 0.8 is excellent (Kwon, 2014).

2.2.3. Representative Concentration Pathway Scenarios

To quantify the potential distribution of major species from different RCP scenarios, we applied the same sampling size and the same sets of SDMs. The only difference is the four RCP scenarios for modeling the potential distribution of major forest species. In this study, the RCPs provided by the Korea Meteorological Administration (KMA) were used to predict future climate change. The KMA provides climate data for the Korean Peninsula at 12.5km × 12.5km resolution using the HadGEM3-RA model and then provides temperature and precipitation data at 1km × 1km resolution for the four RCP scenarios through statistical downscaling. By using the four RCP scenarios, we can quantify the different patterns of impact from RCP.

Table 6 Detailed information on Regional Climate Model (RCM)

Classification	Contents
Base Model	HadGEM3-RA
RCP pathway	RCP2.6/4.5/6.0/8.5
Spatial scope	South Korea
Temporal scope	2001-2100
Spatial resolution	1km
Temporal resolution	Daily, Monthly
Climate variables	Minimum temperature, Maximum temperature, Average temperature, Precipitation

We applied ensemble methods to minimize uncertainties from selecting the species distribution algorithm (Thuiller et al., 2015). Each ensemble method uses different methods to integrate probabilities or binary values from SDM results. We utilized the ensemble methods including 1) mean of probabilities 2) confidence interval 3) median of probabilities 4) models committee averaging 5) weighted mean of probabilities with all five pseudo-absence points and five repetitions in each model. To compare uncertainties from different SDM algorithms, we applied a threshold that matches the total area of the current forest area. Then we compared the differences in spatial distribution among the SDMs int the binary map.

2.2.4. Competition among Major Forest Species

To model competition among the major forest species, we applied the random forest algorithm. The random forest algorithm is one of the machine learning models that ensembles decision tree (Breiman, 2001). Random forest selects a random subset from the input data for classification or prediction and then selects a random variable to select a decision tree. In this process, multiple decision trees are made, and then a collection of trees are called as forest (Mi et al., 2017).

Random forest has the advantage of being able to handle large-scale data in the model and distinguish them by using various input variables (Wang et al., 2015). At the same time, many variables can be extracted and utilized without the user having to delete the variables. In addition, it is possible to classify other data through the constructed tree.

Therefore, in this study, we constructed a prediction model that can model the competition among the major forest species using random forest algorithms. The random forest package in R was applied to construct the model, and the number of trees was set to 1000, with reference to the previous study (Jin et al., 2016). We limit the number of nodes, using 5 out of 12 variables. This is to limit the variables with low impact and to shorten the computation time of the random forest module. For the prediction of multi-species under climate change, we applied Global Argo-Ecological Zones (GAEZ) model (Food and Agriculture Organization, 2012) to limit prediction of the random forest model within current climate zones.

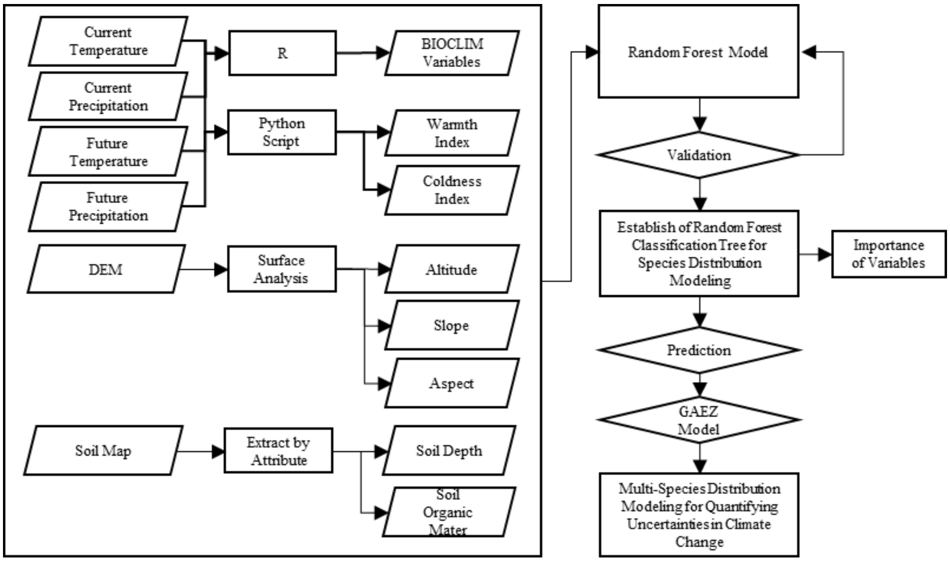


Figure 11 Random forest model framework for quantifying uncertainties

1) Defining Forest Type

In 1: 5000 scale forest maps, 43 different forest species are classified in Korea. To increase the accuracy of the model and minimize the uncertainty, we selected 11 species for modeling, consisting 99% of the total forest in Korea. To evaluate the uncertainty according to the sampling method, we selected the representing one forest species by forest type with the largest area and evaluated the uncertainty according to the model difference and sampling methods.

Table 7 Forest classification of 1:5000 scale forest map

Forest Type	Species Name (Korean)	Scientific Name	Selected Species	CODE
Needle-leaved Forest	소나무	*Pinus densiflora*	○	PD
	잣나무	*Pinus koraiensis*		
	낙엽송	*Larix kaempferi* (Lamb.)		
	리기다소나무	*Pinus rigida* Mill.		
	곰솔	Pinus thunbergii Parl.	○	PT
	전나무	*Abies holophylla*		
	편백나무	*Chamaecyparis obtusa*		
	삼나무	*Cryptomeria japonica*		
	가문비나무	*Picea jezoensis*		
	비자나무	*Torreya nucifera* (L.)		
	은행나무	*Ginkgo biloba* L.		
	기타침엽수	-		
DecidiousBroad-leaved Forest	상수리나무	*Quercus acutissima* Carruth	○	QQ
	신갈나무	*Quercus mongolica* Fisch.	○	QQ
	굴참나무	*Quercus variabilis* Blume	○	QQ
	기타 참나무류	-	○	QQ
	오리나무	*Alnus japonica* (Thunb.)		
	고로쇠나무	*Acer pictum* subsp. *mono*		
	자작나무	*Betula platyphylla* var. *japonica*		
	박달나무	*Betula schmidtii* Regel		
	밤나무	*Castanea crenata*		
	물푸레나무	*Fraxinus rhynchophylla* Hance		
	서어나무	*Carpinus laxiflora* Blume		
	때죽나무	*Styrax japonicus*		
	호두나무	*Juglans regia* L.		
	백합나무	*Liriodendron tulipifera* L.		
	포플러	*Populus lasiocarpa* Oliv.		
	벚나무	*Prunus serrulata* var.		
	느티나무	*Zelkova serrata*		
	층층나무	*Cornus controversa* Hemsl.		
	아까시나무	*Robinia pseudoacacia* L.		
	기타활엽수	-	○	EB
Evergreen-broad leaved forest	가시나무	*Quercus myrsinifolia* Blume		
	구실잣밤나무	*Castanopsis sieboldii* (Makino) Hatus.		
	녹나무	*Cinnamomum camphora* (L.) J. Presl		
	굴거리나무	*Daphniphyllum macropodum* Miq.		

	황칠나무	*Dendropanax morbiferus* H.Lév.		
	사스레피나무	*Eurya japonica* Thunb.		
	후박나무	*Machilus thunbergii*		
	새덕이	*Neolitsea aciculata* (Blume) Koidz.		
	기타상록활엽수	-	○	EG
Mixed Forest	침활혼효림	-	○	MM
Bamboo Forest	죽림	-		

2) Sampling Methods for random forest model

Selecting the presence point for SDMs is essential to predicting the potential distribution of major forest species under climate change. In selecting the presence point for modeling the potential distribution of species, it is difficult to specify the presence point of unlike animals. This is because vegetation lives in a community, so it is necessary to extract representative points in one forest stand which, however, could increase uncertainty.

The number of samples was selected according to the ratio of each forest type to total natural forest. To select the optimum number of samples, the sample was selected by setting the sample at a 95% confidence level and the confidence interval at ± 5%.

$$\hat{\mu} = \bar{y} = \sum_{i=1}^{n} y_i / n \tag{1}$$

$$\mathrm{Var}(\bar{y}) = \frac{\sigma^2}{n} \cdot \left(\frac{N-n}{N-1}\right) \tag{2}$$

$$1.96\sqrt{Var(\bar{y})} \cong 2\sqrt{\frac{\sigma^2}{n} \cdot \left(\frac{N-n}{N-1}\right)} = B \quad (3)$$

$$\text{n} = \frac{N\sigma^2}{(N-1)D + \sigma^2}, D = B^2/4 \quad (4)$$

N: number of parent population n: number of samples

Oak tree species, such as Sawtooth oak, Mongolian oak and Oriental cork oak were 29.4% of the total forest area in Korea, followed by Korean red pine (26.4%), mixed deciduous forests (24.3%), and mixed forests (14.2%). As for other species, Black pine sin was 4.4% and evergreen broad-leaved tree was 0.2%. As a result of selecting the number of sampling points according to the area ratio, 400 samples of oak species were estimated, followed by 312 samples in pine trees and 295 samples in other mixed forests.

Table 8 Classification result and sample size

Classification	Scientific name	Area(㎢)	Percent (%)	Sample size
Evergreen needle leaved forest	*Pinus densiflora*	12966.6	26.4	312
	Pinus thunbergii	2134.4	4.4	67
Deciduous broad-leaved forest	*Quercus acutissima* *Quercus mongolica* *Quercus variabilis*	14423.3	29.4	400
	Mixed species	11903.3	24.3	295
Evergreen broad-leaved forest	Mixed species	87.2	0.2	3
Mixed forest	-	6943.73	14.2	195

As a result of examining the method of selecting the presence point for modeling, it was shown that the most accurate method is to run the species distribution model using the area-weighted stratified sampling method. As a result, large forests were extracted mainly in Gangwon and southern provinces, and evergreen broad- leaved forests were distributed in the Jeju-do area.

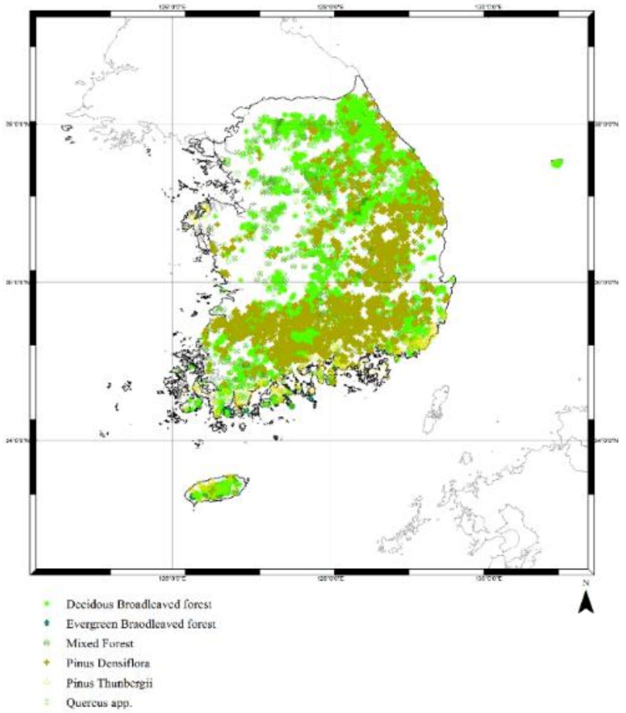


Figure 12 Sampling points for random forest modeling

IV. Result and Discussion

1. Effect of Sampling Method and Sampling Size in SDMs

1.1. Verify Effectiveness of Sampling Methods

We selected Korean red pine as representative species for verifying effective sampling methods and sample sizes. In all sampling sizes, stratified random sampling well represents the population compared to simple random sampling and area-weighted sampling methods. Simple random sampling methods did not represent the population among the three sampling methods in min temperature of coldest month and precipitation of wettest month. Area-weighted sampling methods showed average effectiveness among three sampling methods.

Table 9 Environmental characteristics of population and each sampling method

Environmental Variables*	Population	Simple Random Sampling	Stratified Random Sampling	Area-weighted Sampling
Altitude	323.61	294.40	326.18	292.77
Slope	20.34	16.58	20.25	21.03
Radiation	4589.75	4569	4600.85	4697.57
Distance from water	1358.78	1334.81	1346.92	1213.71
Distance from sea	36156.01	36253.59	34910.71	39725.71
Soil Depth	66.77	86.78	64.66	58.34
Warmth Index	82.63	84.61	83.58	89.58
Isothermality	29.75	37.76	29.69	30.67
Min temperature of Coldest month	-8.13	-2.56	-8.13	-7.24
Precipitation of Wettest month	458.20	387.53	457.54	479.15
Precipitation of driest month	7.74	7.58	7.65	8.53

In simple random sampling methods, when the sample size increases by more than 500, some environmental variables (e.g., altitude, slope, distance from water and sea, isothermality) showed different characteristics when compared with population. The reliability of sampling can be flexible if apply simple random sampling methods are applied.

Table 10 The p-value of one-way t-test result for environmental variables (simple random sampling)

Environmental Variables*	Sample Size						
	30	50	100	200	500	1000	3000
Altitude	0.634	0.228	0.202	0.379	0.001	0.001	0.001
Slope	0.039	0.012	0.089	0.013	0.001	0.001	0.001
Radiation	0.050	0.873	0.198	0.763	0.198	0.694	0.313
Distance from water	0.453	0.303	0.693	0.309	0.044	0.070	0.001
Distance from sea	0.264	0.534	0.751	0.978	0.009	0.863	0.990
Soil Depth	0.038	0.095	0.213	0.08	0.006	0.001	0.001
Warmth Index	0.987	0.215	0.028	0.39	0.068	0.004	0.001
Isothermality	0.825	0.605	0.548	0.888	0.046	0.136	0.057
Min temperature of Coldest month	0.739	0.702	0.405	0.482	0.004	0.023	0.001
Precipitation of Wettest month	0.191	0.781	0.459	0.573	0.006	0.001	0.155
Precipitation of driest month	0.474	0.301	0.520	0.005	0.786	0.112	0.442

Stratified random sampling considered as well constructed sampling methods for collecting samples in population. Only a few sample size (30, 200 samples) and environmental variables (radiation, isothermality, precipitation of direst month) have different average values when compared with population. Stratified random sampling methods can effectively represent the environmental characteristics of Korean red pine forest.

Table 11 The p-value of one-way t-test result for environmental variables (stratified random sampling)

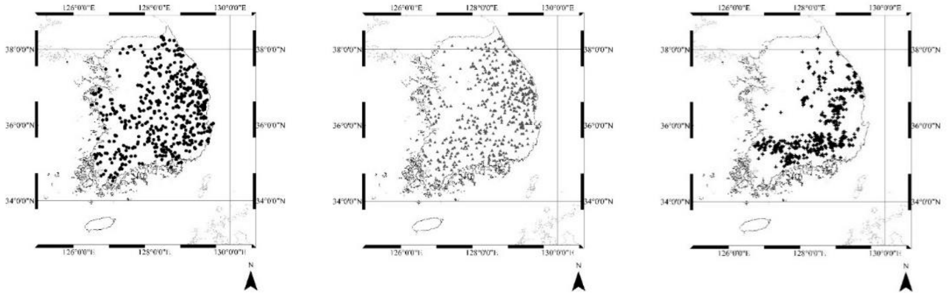
Environmental Variables*	Sample Size						
	30	50	100	200	500	1000	3000
Altitude	0.565	0.481	0.977	0.615	0.23	0.825	0.913
Slope	0.362	0.75	0.677	0.554	0.791	0.633	0.195
Radiation	0.027	0.695	0.622	0.168	0.181	0.528	0.749
Distance from water	0.267	0.794	0.902	0.445	0.349	0.474	0.782
Distance from sea	0.557	0.328	0.925	0.741	0.283	0.354	0.762
Soil Depth	0.099	0.851	0.123	0.867	0.42	0.95	0.333
Warmth Index	0.992	0.177	0.523	0.335	0.5	0.898	0.787
Isothermality	0.929	0.768	0.244	0.022	0.585	0.937	0.818
Min temperature of Coldest month	0.728	0.599	0.549	0.936	0.323	0.974	0.758
Precipitation of Wettest month	0.596	0.275	0.141	0.679	0.203	0.991	0.105
Precipitation of driest month	0.029	0.733	0.368	0.298	0.169	0.088	0.655

On the other hand, area-weighted sampling provided limited representation of population. Radiation, warmth index and isothermality showed different average in all sampling sizes. In addition, precipitation of the wettest month and precipitation of the driest month differed in sampling areas. Forest patches that were selected by area-weighted sampling method have different characteristics in some environmental variables. Selected samples of Korean red pine by area-weighted samples appeared to have higher radiation which can explain higher warmth index and minimum temperature of the coldest month. In addition, the precipitation was larger than the entire population of Korean red pine.

Table 12 the p-value of one-way t-test result for environmental variables (area-weighted sampling)

Environmental Variables*	Sample Size						
	30	50	100	200	500	1000	3000
Altitude	0.913	0.136	0.075	0.005	0.001	0.001	0.001
Slope	0.532	0.363	0.249	0.31	0.127	0.351	0.662
Radiation	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Distance from water	0.817	0.406	0.087	0.001	0.001	0.001	0.001
Distance from sea	0.127	0.221	0.037	0.011	0.022	0.001	0.001
Soil Depth	0.365	0.549	0.216	0.039	0.035	0.089	0.128
Warmth Index	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Isothermality	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Min temperature of Coldest month	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Precipitation of Wettest month	0.016	0.001	0.001	0.001	0.001	0.001	0.001
Precipitation of driest month	0.041	0.076	0.018	0.002	0.001	0.002	0.011

The distribution of samples exhibited different spatial patterns based on the sampling method. In the simple random sampling method, the samples were dispersed in an area distant from the primarily mountainous areas of South Korea. However, samples that were selected using area-weighted sampling were in the southern part of the Korean peninsula. Stratified random sampling was used to collect samples based on the climate zone proportion of the population. These results can reduce the bias when collecting samples for forest modeling. These differences showed the effectiveness of the sampling method.



(a) Random sampling (b) Stratified sampling (c) Area-weighted sampling

Figure 13 Distribution of samples used in species distribution models by sampling method. Each method was used to select 500 samples.

1.2. Verify Effectiveness of Sampling Methods in Species Distribution Modeling

The sampling method caused differences in model performance. The area-weighted sampling method performed better than the stratified random sampling and simple random sampling methods. The average AUC value for models based on area-weighted sampling was 0.777, which was considered good, while stratified random sampling and simple random sampling had AUC values of 0.663 and 0.622, respectively, and were thus considered average. Additionally, area-weighted sampling demonstrated stable performance, even across different sample sizes, as shown by the standard deviation of model performance.

Table 13 Average performance (AUC) of SDMs by sampling method for all sample sizes and SDM types as determined by the analysis of variance.

Sampling methods	Average AUC	Standard deviation	F-Value[1]	Post-Hoc Analysis[2]
Simple random sampling	0.622	0.050	1436.904	Simple random sampling < Stratified sampling, Area-weighted sampling
Stratified random sampling	0.663	0.070		Stratified random sampling > Simple random sampling Stratified random sampling < Area-weighted sampling
Area-weighted Sampling	0.777	0.087		Area-weighted sampling > Simple random sampling, Stratified random sampling
1 * P<0.0001, 2 Games-Howell tests. P<0.05				

We found that sample size is a significant factor in deciding the reliability of species distribution modeling. Generally, a larger sample size increases model performance (Moudrý and Šímová, 2012; Wisz et al., 2008). However, our results showed that as the sampling size increased, the AUC of a given model did not respond linearly; instead the correlation coefficient decreased logarithmically with increase in sample size. Thus, increases in sampling size after a certain number, 200 samples in this study, do not increase model performance significantly. Different ecological niches decrease model performance.

Table 14 Average performance of SDMs by sample size (AUC) and ANOVA results for all SDMs.

Sampling methods	Sample size	Average	Standard deviation	F-Value[1]	Post-Hoc Analysis[2]
Random sampling	30s	0.610	0.121	40.991	30s < 200s, 500s, 1000s, 3000s ; 30s > 100s
	50s	0.585	0.089		50s < 200s, 500s, 1000s, 3000s
	100s	0.570	0.066		100s < 30s, 100s, 200s, 500s, 1000s, 3000s
	200s	0.642	0.066		200s > 30s, 50s, 100s
	500s	0.643	0.052		500s> 30s, 50s, 100s
	1000s	0.648	0.046		1000s > 30s, 50s, 100s
	3000s	0.653	0.043		3000s > 30s, 50s, 100s
Stratified sampling	30s	0.624	0.104	24.149	30s < 100s, 200s, 500s, 1000s, 3000s
	50s	0.639	0.075		50s < 200s, 500s, 1000s, 3000s
	100s	0.660	0.071		100s > 30s; 100s <200s, 3000s
	200s	0.681	0.058		200s > 30s, 50s, 100s
	500s	0.669	0.051		500s > 30s, 50s; 500s < 3000s
	1000s	0.675	0.045		1000s> 30s, 50s; 1000s < 3000s
	3000s	0.693	0.043		3000s > 30s, 50s, 100s, 500s, 1000s
Area-weighted sampling	30s	0.762	0.129	4.633	30s < 200s
	50s	0.779	0.116		-
	100s	0.786	0.083		100s > 3000s
	200s	0.798	0.077		200s > 30s, 1000s, 3000s
	500s	0.782	0.058		500s > 3000s
	1000s	0.775	0.060		1000s < 200s
	3000s	0.760	0.045		3000s < 200s, 500s

1 * P<0.0001, 2 Games-Howell test. P<0.05

When we analyzed the altitude of each collected sample by sampling method, we found that the area-weighted sampling method showed a different altitude distribution than the simple random sampling and stratified random sampling methods did. Anything from 0-500m altitude was a suitable habitat for Korean red pine (Lee and Jo 2003). The area-weighted sampling method showed 88% of its samples within this 0-500m range, while the simple random sampling and stratified random sampling methods had 80% and 79% of their samples within this range, respectively.

In addition, samples by area-weighted sampling method were collected in mostly age classes 4 and 5 (40-50 years). Area-weighted sampling selected more aged (90.80%) forests compared to the selection of other two sampling methods (simple random sampling: 85.23% and stratified random sampling 85.32%). These differences led to the inclusion of the ecological preferences of Korean red pine, which can affect model performance.

Sampling methods can change performance of SDMs. When we compared the three different sample selection methods, the area-weighted sampling methods showed better performance compared to the stratified sampling and random sampling methods. However, the characteristics of sample matched well with the population if we applied stratified sampling methods.

Sample size partially influenced the performance of SDMs partially. If enough samples were acquired, the performance of the model did not change significantly. As a result of this study, if the sample size exceeds 200, the performance of SDMs did not increase in direct proportion to the sampling size. Thus, for modeling changes of forest under future climate change, we can utilize sampling methods and sample size for effective monitoring in developed countries that have limited budget and human resources.

2. Uncertainties in Applying Different SDMs

2.1. Performance Changes by SDMs and Sampling Methods

As a result of modeling the differences for the Korean red pine, which is a representative species of needle-leaved forest, according to the sampling methods of species, the AUC value was 0.727 the highest value for the results of extracting the occurrence point using the area-weighted method. When the simple random sampling method was applied, the AUC value was the lowest (0.615). Since the simple random sampling method extracts the sample without considering the ecological characteristics of the forest species, it is consistent with the existing results, which are known to have the lowest accuracy. The TSS values also showed a similar pattern as the AUC values. Higher accuracy was found machine learning models such as RF and GBM.

Among the SDMs, the RF model showed the highest accuracy with an AUC value of 0.806. The performance of the RF model is higher than other models as the RF model applied bootstrapping for selecting the potential distribution (Wang et al., 2015). The ANN model is consistent with the results of previous studies which tend to be overestimated potential distribution of forest and show poor results (AUC value 0.581) compared to GAM or GLM (Franklin, 2010b; Thuiller, 2003)

Table 15 AUC values according to sampling methods (Korean red pine forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.583	0.574	0.669
CTA	0.571	0.597	0.675
FDA	0.637	0.623	0.773
RF	0.651	0.648	0.806
GLM	0.644	0.638	0.761
GBM	0.649	0.645	0.794
GAM	0.610	0.629	0.759
ANN	0.577	0.585	0.581
Average	0.615	0.617	0.727

Table 16 TSS values according to sampling methods (Korean red pine forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.230	0.200	0.347
CTA	0.189	0.202	0.355
FDA	0.298	0.316	0.522
RF	0.326	0.319	0.543
GLM	0.325	0.322	0.476
GBM	0.316	0.325	0.526
GAM	0.275	0.292	0.519
ANN	0.175	0.216	0.209
Average	0.267	0.274	0.437

When we compared the performance of SDMs which predicted the potential distribution of oak with different sampling methods, the AUC value was the lowest for stratified random sampling, while it was value the highest for area-weighted sampling (0.643). This is similar to the result of the previous studies used to select habitat sites when the forest area is large (J. Kim et al., 2015). Among the different SDMs, the performance of machine learning models such as the GBM and RF model, was the highest.

Table 17 AUC values according to sampling methods (Oak forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.659	0.619	0.783
CTA	0.610	0.620	0.763
FDA	0.674	0.660	0.838
RF	0.682	0.667	0.876
GLM	0.665	0.658	0.862
GBM	0.694	0.672	0.877
GAM	0.661	0.659	0.843
ANN	0.595	0.586	0.779
Average	0.655	0.643	0.828

Table 18 TSS values according to sampling methods (Oak forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.328	0.270	0.557
CTA	0.242	0.246	0.529
FDA	0.349	0.329	0.606
RF	0.372	0.342	0.651
GLM	0.362	0.334	0.626
GBM	0.382	0.358	0.657
GAM	0.352	0.334	0.616
ANN	0.212	0.186	0.522
Average	0.325	0.300	0.595

The AUC values of the mixed forest model showed the highest AUC values in area-weighted sampling (0.739). As with mixed forests and coniferous forests, the accuracy of SDMs that applied the simple random sampling method is lowest (0.597). The machine learning models such as the

RF model and GBM model showed the highest AUC values. The TSS value was also similar to the AUC value, because the accuracy of the machine learning model was higher than that of statistical based models.

Table 19 AUC values according to sampling methods (Mixed forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.582	0.615	0.693
CTA	0.558	0.595	0.690
FDA	0.616	0.641	0.767
RF	0.619	0.647	0.795
GLM	0.607	0.647	0.758
GBM	0.625	0.650	0.785
GAM	0.616	0.650	0.771
ANN	0.551	0.559	0.650
Average	0.597	0.626	0.739

Table 20 TSS values according to sampling methods (Mixed forest)

	Simple random sampling	Stratified random sampling	Area-weighted sampling
MAXENT	0.203	0.265	0.393
CTA	0.159	0.201	0.375
FDA	0.274	0.293	0.494
RF	0.292	0.312	0.520
GLM	0.294	0.331	0.473
GBM	0.300	0.312	0.511
GAM	0.286	0.318	0.500
ANN	0.142	0.158	0.299
Average	0.244	0.274	0.446

2.1.1. Importance of Environmental Variables in each SDMs[3]

Altitude is the most important variable in modeling the Korean red pine forest, and precipitation of wettest month and precipitation of driest month

[3] Relative importance of all environmental variables is described in Appendix

are considered the second most important variables. Radiation, soil depth and organic matter in the A layer were important variables in the simple random sampling method.

Table 21 Three relative important variables for each model (Korean red pine)

	Area weighted sampling	Random sampling	Stratified sampling
MAXENT	ALT, BIO13, BIO14	RAD, Soil_depth, WI	ALT, SLO, Soil_Depth
CTA	ALT, Dis_water, BIO13	RAD, Soil_OM, BIO13	ALT, SLO, BIO13
FDA	ALT, WI, BIO6	SLO, Soil_depth, WI	ALT, WI, BIO14
RF	ALT, Dis_waterm BIO13	ALT, Soil_depth, BIO13	ALT, SLO, BIO14
GLM	ALT, Soil_OM, BIO13	Soil_depth, Soil_OM, WI	Soil_depth, Soil_OM, WI
GBM	ALT, BIO13, BIO14	RAD, Soil_depth, BIO13	ALT, SLO, BIO13
GAM	ALT, Soil_OM, WI	Soil_depth, Soil_OM, WI	Soil_depth, Soil_OM, WI
ANN	ALT, Dis_water, Dist_sea	Dis_water, Dist_sea , Soil_depth	ALT, Dis_water, Dist_sea,

ALT: Altitude, SLO: Slope, RAD: Radiation, Dis_water: Distance from water, Dis_sea: Distance from sea, Soil_depth: Soil depth, Soil_OM: Organic matters in soil layer A, WI: Warmth Index, BIO3: Isothemarlity, BIO6: Min temperature of coldest month, Bio13: Precipitation of wettest month, Bio14: Precipitation of driest month

Altitude is the most important variable in modeling deciduous as well as coniferous forest. Among the climatic variables, precipitation of the wettest month plays an important role in predicting the distribution of oak forest. This is consistent with previous studies that analyzed the existing oak distribution area. Those studies, indicated that oak forests prefer dry places, and their cold tolerance is high (Kim and Kim, 2017). Maximum temperature in the random sampling method has the highest influence, while slope is an important variable in the stratified sampling methodology.

Table 22 Relative important variables in each model (Oak forest)

	Area-weighted sampling	Random sampling	Stratified sampling
MAXENT	ALT, BIO13, BIO14	ALT, SLO, BIO6	Slope, Dis_water, WI
CTA	ALT, Dis_water, BIO13	ALT, SLO, BIO6	ALT, SLO, Dis_sea
FDA	ALT, WI, BIO6	ALT, SLO, BIO6	SLO, WI, BIO6
RF	ALT, Dis_water, BIO13	ALT, SLO, BIO6	ALT, SLO, BIO6
GLM	ALT, Soil_OM, BIO13	ALT, Soil_depth, BIO6	SLO, Soil_OM, BIO6
GBM	ALT, BIO13, BIO14	ALT, SLO, BIO6	ALT, SLO, Dis_sea
GAM	ALT, Soil_OM, WI	Soil_depth, Soil_OM, BIO6	Soil_depth, Soil_OM, Bio6
ANN	ALT, Dis_water, Dis_sea	ALT, Dis_water, Dis_sea	ALT, Dis_water, Dis_sea

ALT: Altitude, SLO: Slope, RAD: Radiation, Dis_water: Distance from water, Dis_sea: Distance from sea, Soil_depth: Soil depth, Soil_OM: Organic matters in soil layer A, WI: Warmth Index, BIO3: Isothemarlity, BIO6: Min temperature of coldest month, Bio13: Precipitation of wettest month, Bio14: Precipitation of driest month

As a result of evaluating the importance of the variables in mixed forest projection, in all sampling methods, topographic variables such as altitude and slope area considered important variables for modeling the potential distribution of forest. Climatic variables are relatively less important for modeling the current distribution of mixed forest.

Table 23 Relative important variables in each model (mixed forest)

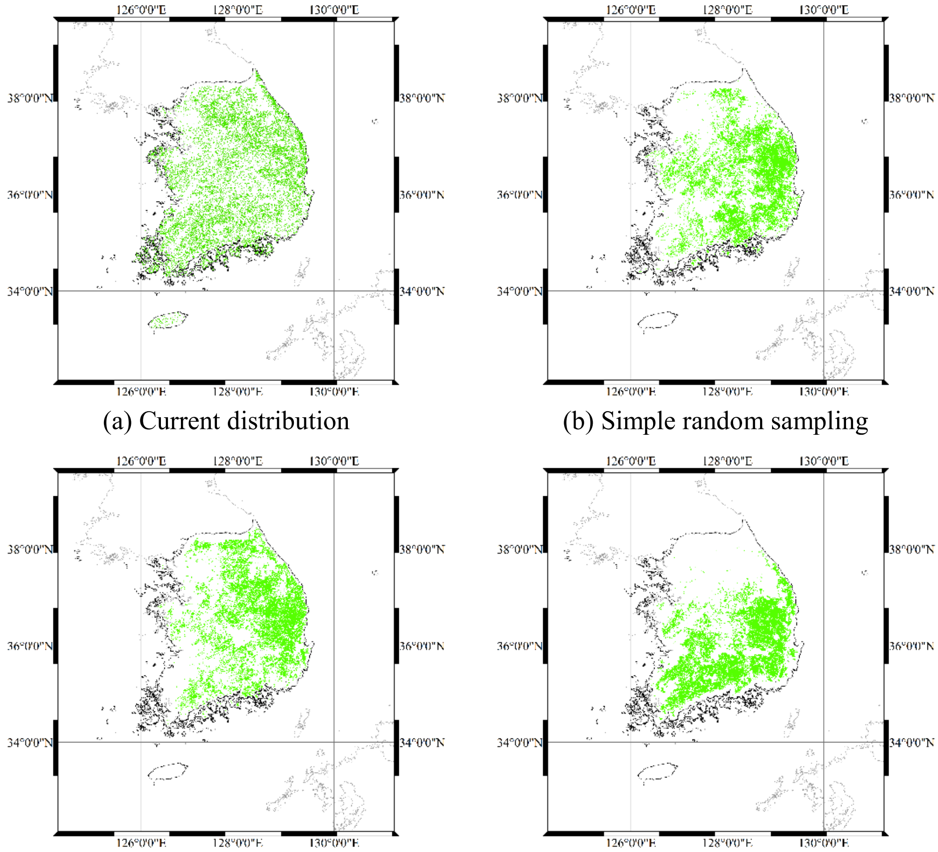
	Area-weighted sampling	Random sampling	Stratified sampling
MAXENT	ALT, SLO, Dis_sea	SLO, Soil_depth, BIO6	SOL, RAD, Soil_depth
CTA	ALT, SLO, Dis_water	SLO, Soil_depth, BIO13	ALT, SLO, Soil_depth
FDA	ALT, BIO6, BIO13	ALT, SLO, WI	SLO, RAD, Soil_depth
RF	ALT, SLO, BIO14	ALT, SLO, BIO13	SLO, RAD, Soil_depth
GLM	ALT, Soil_depth, Soil_OM	SLO, Soil_depth, BIO13	SLO, Dis_sea, Soil_depth
GBM	ALT, SLO, BIO13	ALT, SLO, Soil_depth	SLO, RAD, Soil_depth
GAM	ALT, Soil_depth, Soil_OM	Soil_depth, Soil_OM, WI	Soil_depth, Soil_OM, WI
ANN	ALT, Dis_water, Dis_sea	ALT, Dis_water, Dis_sea	ALT, Dis_sea, Soil_depth

ALT: Altitude, SLO: Slope, RAD: Radiation, Dis_water: Distance from water, Dis_sea: Distance from sea, Soil_depth: Soil depth, Soil_OM: Organic matters in soil layer A, WI: Warmth Index, BIO3: Isothemarlity, BIO6: Min temperature of coldest month, Bio13: Precipitation of wettest month, Bio14: Precipitation of driest month

2.1.2. Spatial Distribution Patterns from Different SDMs

For Korea red pine forest, which represents needle leaved trees, the largest AUC change was the RF model, while the ANN model performed poorly. The difference in distribution exhibited by the model is influenced by the variables used in the model. For the RF model, the important variables varied according to the sampling method. However, for the ANN model, the selected important variables are the distance from the freshwater and the distance from the ocean, regardless of the sampling method.

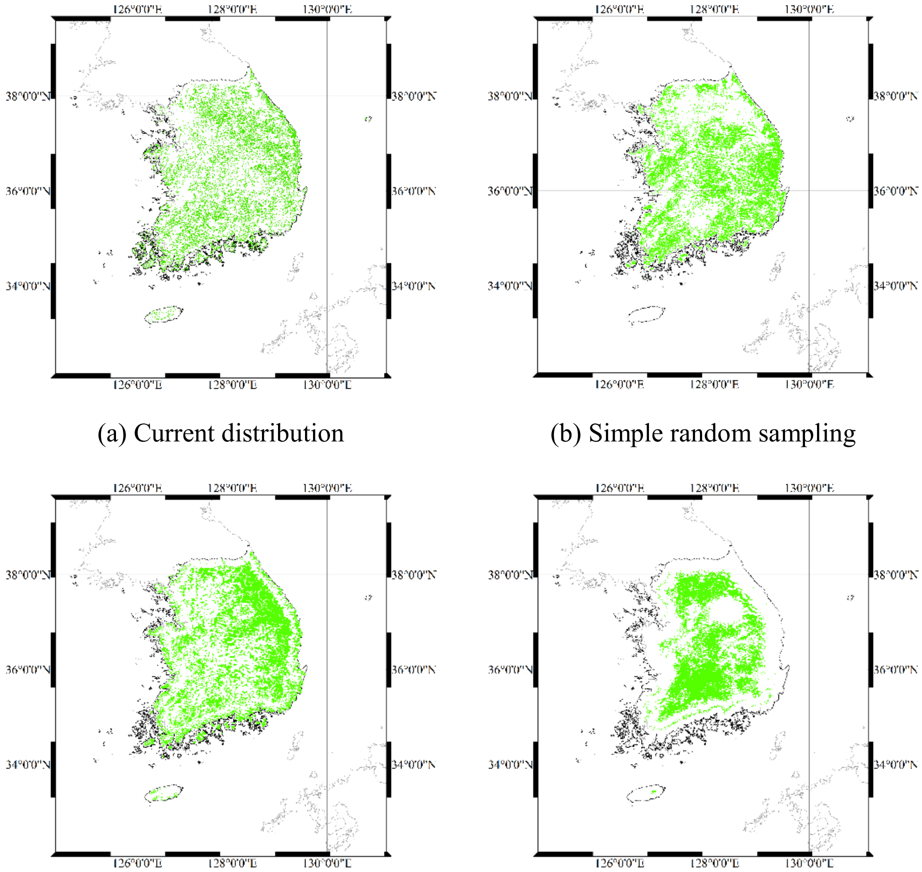
As shown in Figure 14, the spatial distribution of Korea red pine per the RF model was found to be clustered in the south-eastern region. On the contrary, the ANN model showed a tendency to overestimate the distribution of forests in the entire country which is largely different from the current distribution as the distance from sea was selected as an important variable for modeling the potential distribution of Korean red pine. According to the spatial distribution of SDMs, Korean red pine will be distributed mainly in the Gangwon province and mountainous areas. However, the western part of the Korean peninsula showed higher uncertainty regarding the potential distribution of Korean red pine.



(a) Current distribution (b) Simple random sampling

(c) Stratified random sampling (d) Area based sampling

Figure 14 Spatial distribution of RF model by sampling methods (Korean red pine)

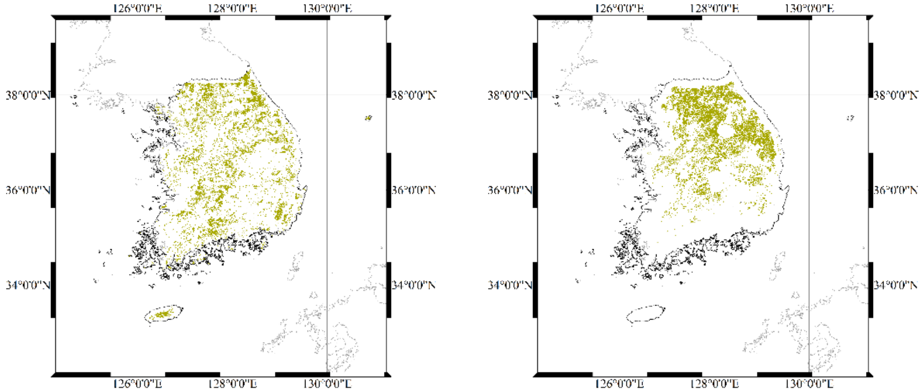


(a) Current distribution (b) Simple random sampling

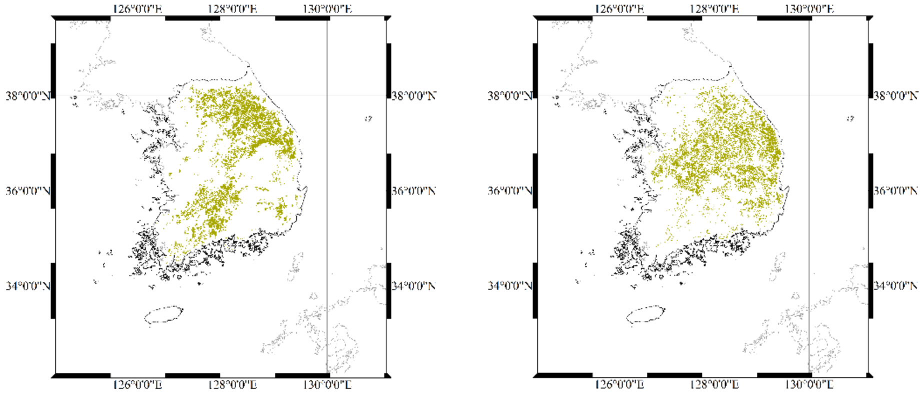
(c) Stratified random sampling (d) Area based sampling

Figure 15 Spatial distribution of ANN model by sampling methods (Korean red pine)

In the case of oak forest, the model with the largest AUC changes was the RF model while the CTA model showed smallest changes. The RF model projected the potential distribution of the entire forest region in Korea. However, when the area-weighted sampling method was applied, it was analyzed that the broad-leaved forest was mainly distributed in the high-altitude mountainous area. For the CTA model, the projected result from the random sampling was overestimated in the forest areas. However, as a result of using the area-weighted sampling, the distribution range of the forest is relatively decreased, and the accuracy of the model is improved. Even though the performance of the RF model was higher than that of the CTA model, the spatial distribution patterns were close to the current distribution of oak forest.

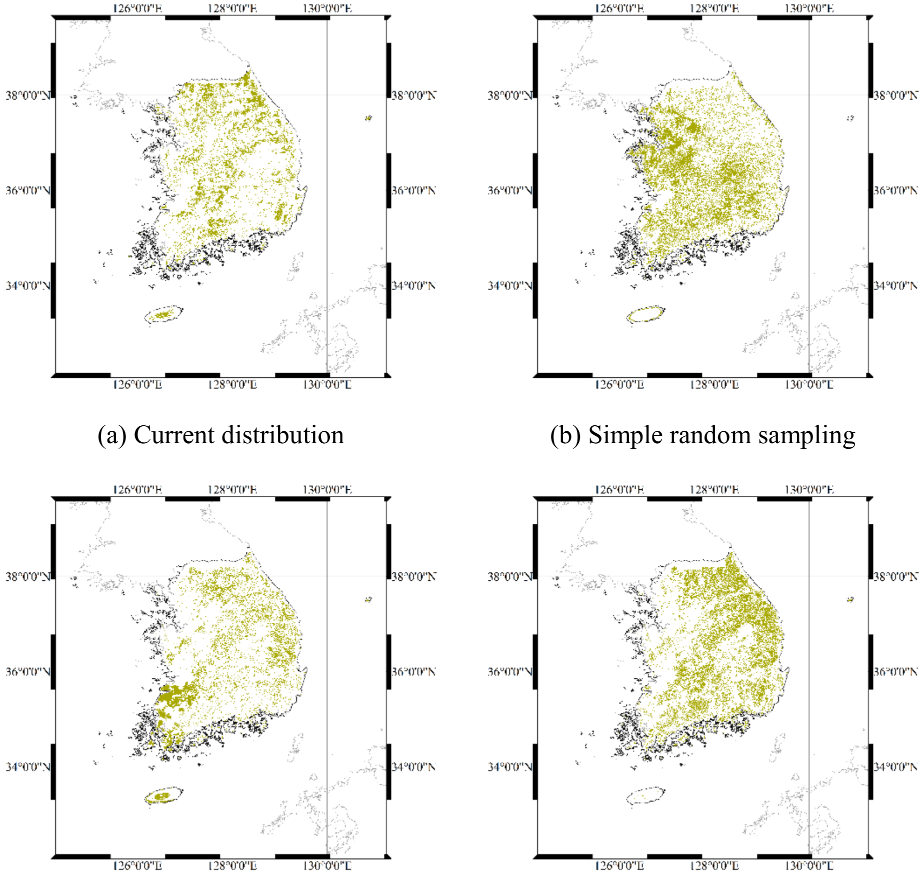


(a) Current distribution (b) Simple random sampling



(c) Stratified random sampling (d) Area based sampling

Figure 16 Spatial distribution of RF model by sampling methods (Oak forest)

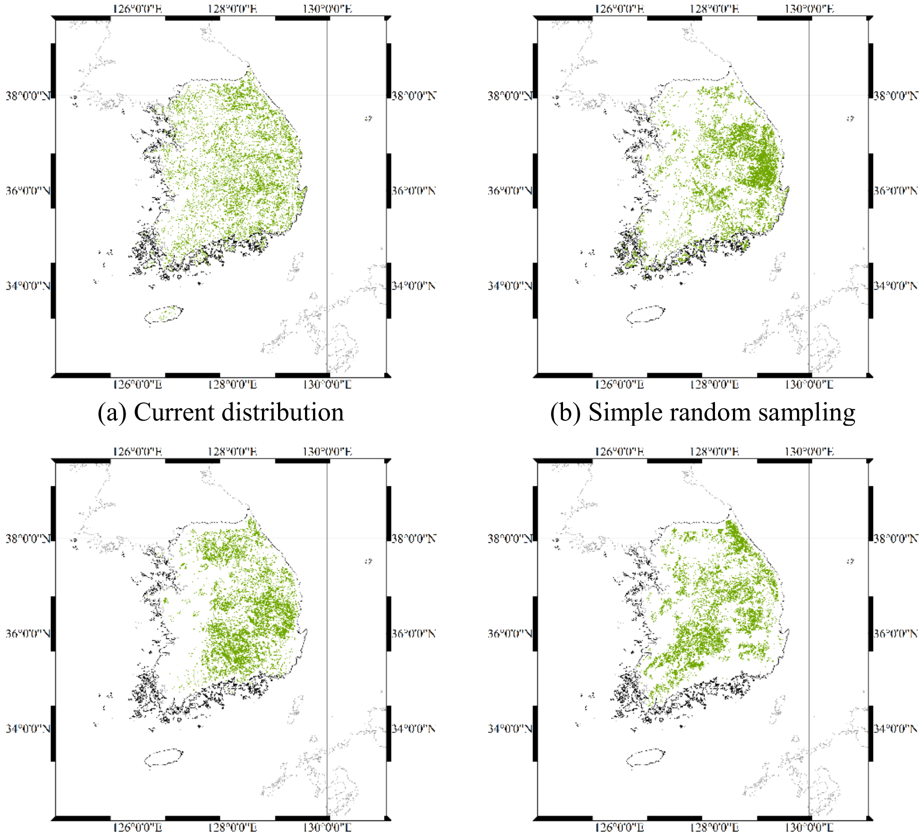


(a) Current distribution (b) Simple random sampling

(c) Stratified random sampling (d) Area based sampling

Figure 17 Spatial distribution of CTA model by sampling methods (Oak forest)

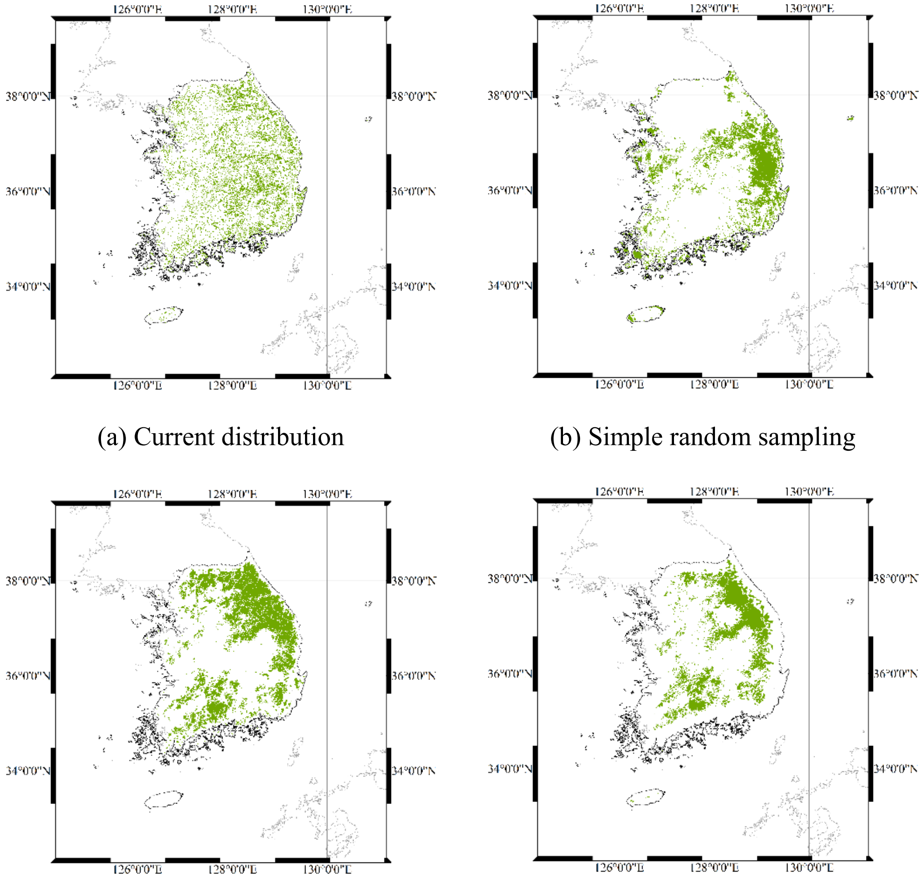
In case of mixed forests, the increased in AUC in the RF model was the highest as in the case of the deciduous forest, and the increase in AUC in the CTA model was the lowest. When we look at the present distribution of the mixed forest, it is considered difficult to categorize the distribution characteristics because the mixed forest is distributed all over the country, rather than being clustered characteristically in a certain area. When we investigate the spatial distribution, the ANN model does not accurately predict the current forest distribution, while the RF model overestimates the distribution of the actual mixed forest. The spatial distribution of the mixed forest was different despite of the similarity of the model's performance in the stratified sampling of the RF model and area-weighted sampling. The RF model estimated the potential distribution in the eastern part of Korea while ANN model modeled the potential distribution in high mountainous areas.



(a) Current distribution (b) Simple random sampling

(c) Stratified random sampling (d) Area based sampling

Figure 18 Spatial distribution of RF model by sampling methods (Mixed forest)



(a) Current distribution (b) Simple random sampling

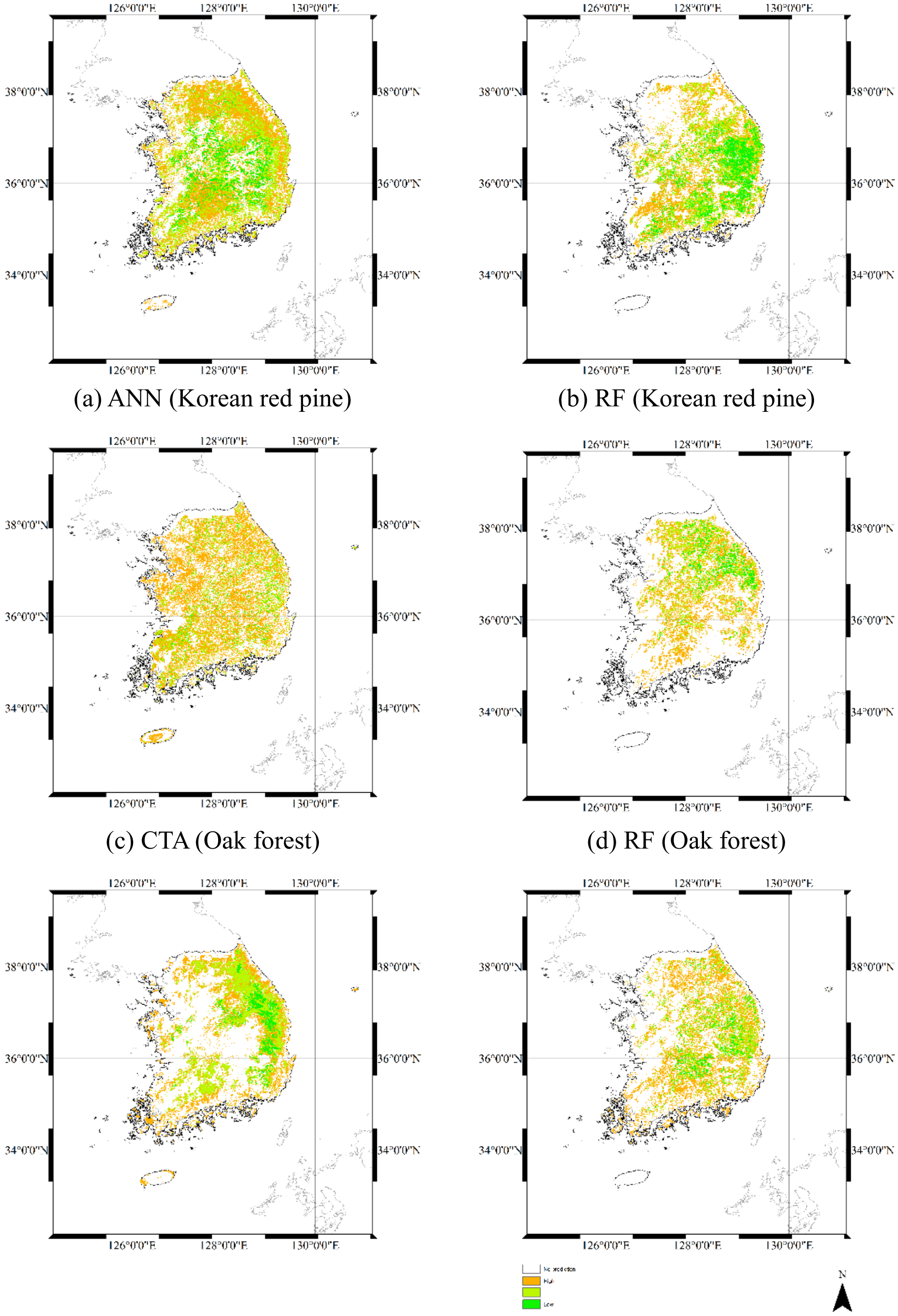
(c) Stratified random sampling (d) Area based sampling

Figure 19 Spatial distribution of ANN model by sampling methods (Mixed forest)

As a result of modeling the potential distribution of forest using the multi species distribution model, machine learning algorithms such as RF and GBM show higher accuracy than statistical based algorithms do. This suggests that a machine learning model is appropriate when applying the species distribution model in South Korea. However, as the spatial resolution changes, the algorithm should be considered carefully.

The modeled accuracy of each SDM varies according to the sampling methods. The RF model showed the largest AUC change according to the sampling method. In the case of the RF model, a decision tree with a small correlation is created through a combination of various variables and nodes, and the RF, based on the majority rule, is presented through the voting of the decision tree, so that the accuracy of the model is higher than that of the CTA model. Conversely, the CTA model showed the least increase in accuracy. This can be attributed to the disadvantages of the algorithm of the CTA model. If the correlation between variables shows a linear or continuous response in predicting the distribution of forests, then it is difficult to extract the thresholds, which reduces the stability of the model (Hastie, 2009).

When we measure the uncertainties of sampling methods about spatial distribution, the model with a higher accuracy (RF) prediction of the potential distribution of major forest species has lower uncertainties compared to lower accuracy (ANN and CTA) (Figure 20). In the case of Korean red pine, the RF modeled distribution area with lower uncertainty was 23.7%, while for the ANN model it was 13.6%. Meanwhile, for RF model projection the lower uncertainty area 11.4% while for the CTA model it was 4.3% for oak forest. On the other hand, for mixed forest, the ANN modeled lower uncertainty area was 10.4% while for the RF 9.4% of total area has lower uncertainty.



(a) ANN (Korean red pine) (b) RF (Korean red pine)

(c) CTA (Oak forest) (d) RF (Oak forest)

(e) ANN (Mixed forest) (f) RF (Mixed forest)

Figure 20 Uncertainties of species distribution change by model

Even if the performance is higher in statistical measures, the differences in spatial distributions should be considered while we apply different SDMs. In the application of different SDMs, ensemble modeling can be a good option for minimizing uncertainties in modeling the potential distribution of forests. As ensemble methods were applied, the performance of SDMs increased regardless of ensemble method. Thus, the ensemble methods have been widely applied for modeling species distribution to reduce the uncertainties from the modeling (Ahn et al., 2015; H. G. Kim et al., 2015; Kim et al., 2018).

3. Uncertainty Considering Competition among Major Forest Species

3.1. Ensemble Modeled Distribution of Single Forest species in Current Climate Condition

As a result of application of species distribution models, there are several differences in the AUC and TSS value by modeling algorithms. Among the machine learning based models, the GBM and RF models which apply an ensemble of models within the model algorithm performs better than the CTA model, which is a single decision-tree-based model. Also, the statistical based models (GAM, GLM) performs well in modeling the potential distribution of Korean red pine.

Table 24 Modeled performance (AUC and TSS) for single species (Korean red pine)

	AUC		TSS	
	Average	STD	Average	STD
MAXENT	0.805	0.035	0.518	0.062
CTA	0.744	0.059	0.454	0.094
FDA	0.823	0.029	0.538	0.058
RF	0.855	0.030	0.606	0.058
GLM	0.818	0.032	0.534	0.062
GBM	0.847	0.028	0.596	0.064
GAM	0.843	0.026	0.573	0.055
ANN	0.654	0.077	0.277	0.125

Among the five ensemble methods, weighted mean of probabilities was the most accurate. The other ensemble methods had high AUC values. Thus, ensemble method achieved high reliability compared to single SDMs.

Table 25 Evaluation result of ensemble models

Ensemble Methods	AUC	Cutoff	Sensitivity	Specificity
Mean of Probabilities	0.924	437.5	90	78.831
Confidence Interval	0.923	403.5	90	78.567
Median of Probabilities	0.908	512.5	83.5	83.030
Model committee average	0.914	632.5	85.0	80.329
Weighted mean of probabilities	0.925	479.5	86.5	82.824

When we utilized the weighted mean of probabilities ensemble method, the spatial patterns between the modeled distribution of Korean red pine and the current distribution of Korean red pine was different in the mid-part of the Korean peninsula and northern part of South Korea. On the other hand, the Korean red pine located in the southern part of Korea was well simulated, as the sampled Korean red pine forest stands are in mountainous and southern part of Korea.

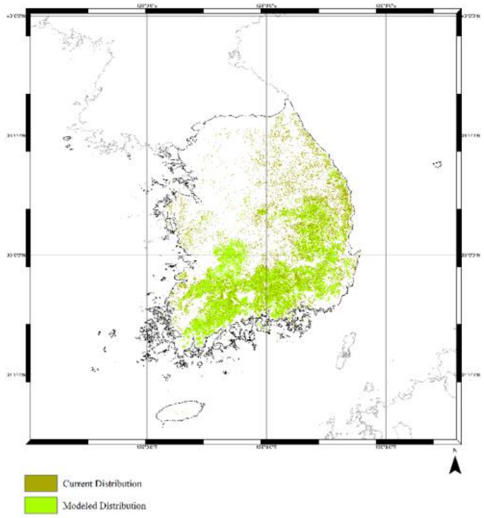


Figure 21 Comparison of Korean red pine between current and modeled distribution

3.2. Modeled Distribution of Considering Competition among Major Forest Species

Evaluation of the accuracy of the constructed RF model revealed an accuracy of the model was 61%. In the case of Korean red pine forest, black pine forest, and oak forest, the accuracy was more than 60% higher than the average. However, deciduous broad forest, evergreen broadleaf and mixed forest exhibited accuracies lower than average. For Korean red pine forest, and black pine forest, the species classification is relatively accurate in the species distribution model, and the classification accuracy is relatively high, even in the case of the deciduous forest because the ecological characteristics are similar.

However, other species such as broad-leaved trees, evergreen broad-leaved trees, and mixed forests did not have representatives for each plant type, and various ecological characteristics within the forest type are present, which is why the model has a low accuracy in classifying forest type.

Table 26 Confusion matrix for forest type classification of South Korea

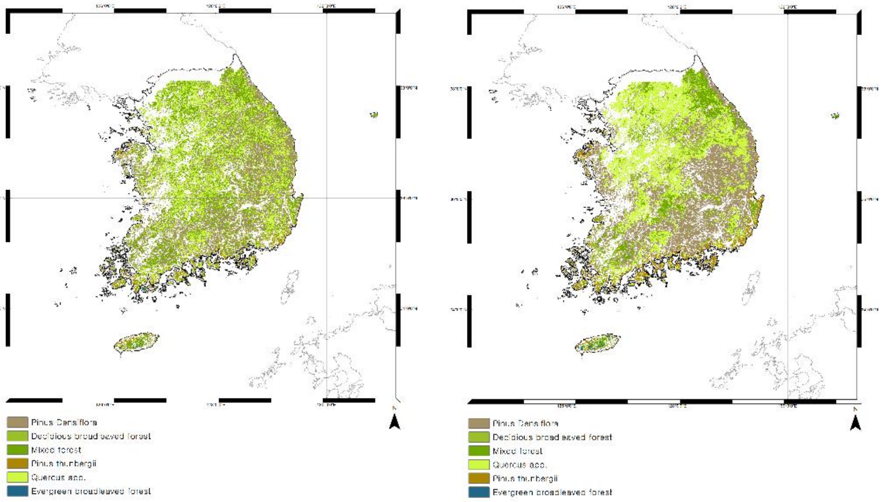
	Deciduous broad forest	Ever broad forest	Mixed forest	Pinus densiflora	Pinus thumbergii	Quercus app.	Total	User's Accuracy (%)
Deciduous broad forest	751	25	109	189	32	717	1823	41.2%
Ever broad forest	30	35	14	0	4	10	93	37.6%
Mixed forest	165	17	364	273	37	342	1198	30.4%
Pinus densiflora	101	1	96	1422	26	307	1953	72.8%
Pinus thumbergii	19	6	21	50	223	17	336	66.4%
Quercus app.	543	10	125	243	34	1537	2492	61.7%
Total	1609	94	729	2177	356	2930		
Producer's Accuracy (%)	46.7%	37.2%	49.9%	65.3%	62.6%	52.5%		
OOB estimate of error rate: 45.13%								

When we compared distribution of forest between the forest map and predicted results, the differences of the other deciduous broad-leaved forests are the largest, while the differences of the Korean red pine forests followed. It is considered that general deciduous broad-leaved trees have a higher difference than the other forest types because the ecological characteristics of various deciduous broad-leaved trees are considered as one forest type.

Table 27 Total area and ratio of Korean forest by forest type (unit: km^2)

	Current (forest map)	Ratio	Project (RF model)	Ratio	Difference
Pinus Densiflora	19,025	28.7%	24,604	37.1%	8.4%
Pinus Thumbergii	2,696	4.1%	4,333	6.5%	2.5%
Quercus app	20,108	30.3%	23,979	36.2%	5.8%
Deciduous broadleaved forest	16,139	24.4%	7,039	10.6%	-13.7%
Evergreen broadleaved forest	102	0.2%	201	0.3%	0.1%
Mixed forest	8,191	12.4%	6,106	9.2%	-3.1%
Total	66,262		66,262		

As a result of modeling, Korean red pine tree forests were distributed in the southern part of Korea, and the modeled distribution of black pine forest were located around the coast. However, there were differences in predicting the distribution of mixed forests. This is because the mixed forest consisted of various species. In addition, the distribution of actual forests in Korea is heterogeneously distributed in mixed forests and deciduous forests.



(a) Current forest distribution (b) Simulated forest distribution

Figure 22 Comparison of forest distribution between forest map and simulation

3.2.1. Importance of Environmental Variables in Constructing a Multi-Species Model

The results of the RF model showed that the altitude value was the most important factor in classifying the forest species. Isothermality (BIO3), minimum temperature of coldest month (BIO6) and radiation (RAD) were also important variables. The altitude plays a most important role in classifying the forest type, which is consistent with the altitude variables accounting for a large part of the vegetation growth conditions presented in the previous studies. In addition, isothermality (BIO3) is closely related to the appropriate temperature at which plants can survive in terms of temperature fluctuation at maximum and maximum monthly temperatures. The amount of radiation is closely related to the amount of energy that can be utilized for the growth of plants, so it is used to distinguish between a sun tree and tolerant tree. The distance to the sea is used to distinguish species with low salt tolerance.

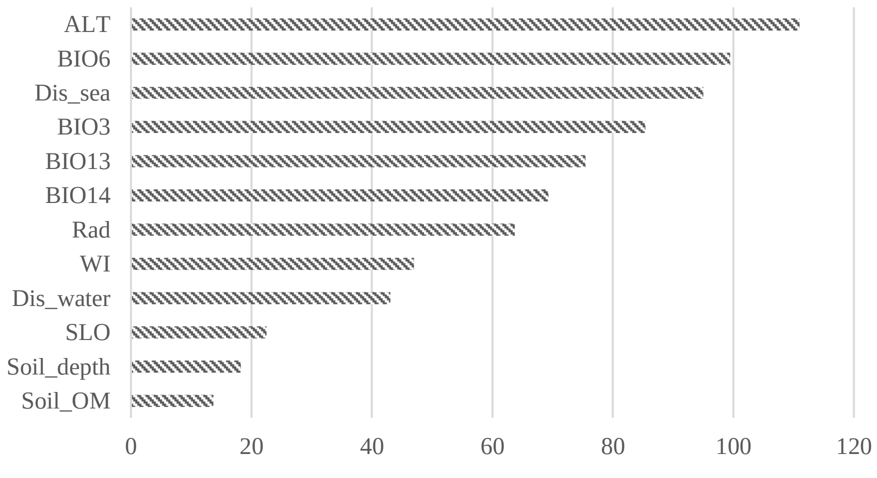


Figure 23 Mean decrease accuracy of over all classes from the RF classification[4]

[4] Please refer to appendix for mean decrease accuracy of variables by major species

In the classification tree in the RF model, altitude is selected as the first node for distinguishing forest type. In addition, minimum temperature of coldest month (BIO6), radiation, isothermality (Bio6) and WI, which are related to the temperature selected around the second node for forest type classification, and precipitation and location parameters such as distance from water and sea, and precipitation in wettest or driest month in considered as the third node for forest type classification. This is consistent with previous studies that show that temperature, precipitation, and topographic variables act as significant environmental variables at medium resolution with a spatial scale of about 1 km (Pearson and Dawson, 2003)

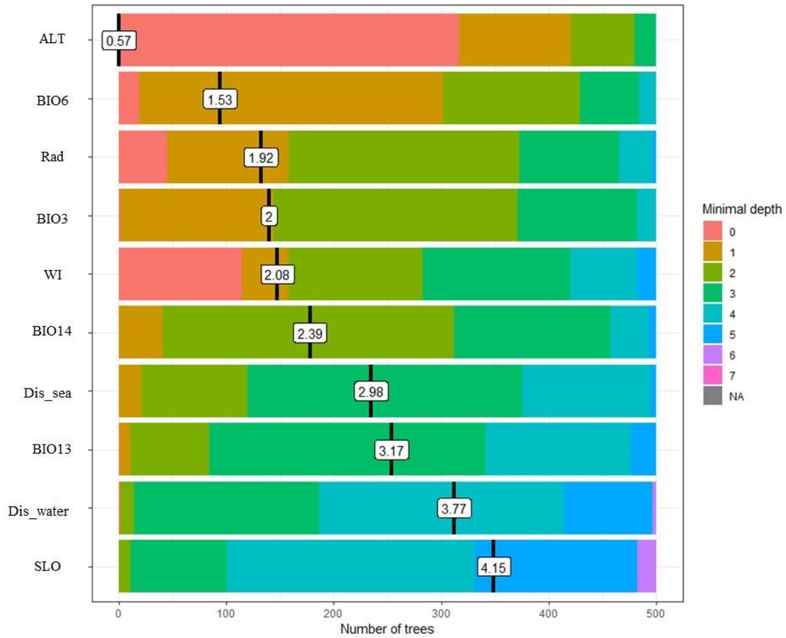


Figure 24 Distribution of minimal depth of environmental variables and their means in RF

4. Potential Distribution of Forest Species in Different RCPs

4.1. Potential Distribution of Korean Red Pine in SDMs

When the potential distribution was modeled with different RCP scenarios, the potential distribution varied among the climate scenarios. As the temperature increased, the distribution of Korean red pine moved to the northern part of South Korea. However, the modeled distributed area of Korean red pine increased as this model only included climatic conditions for analyzing the potential distribution of Korean red pine. This result is consistent with previous studies.

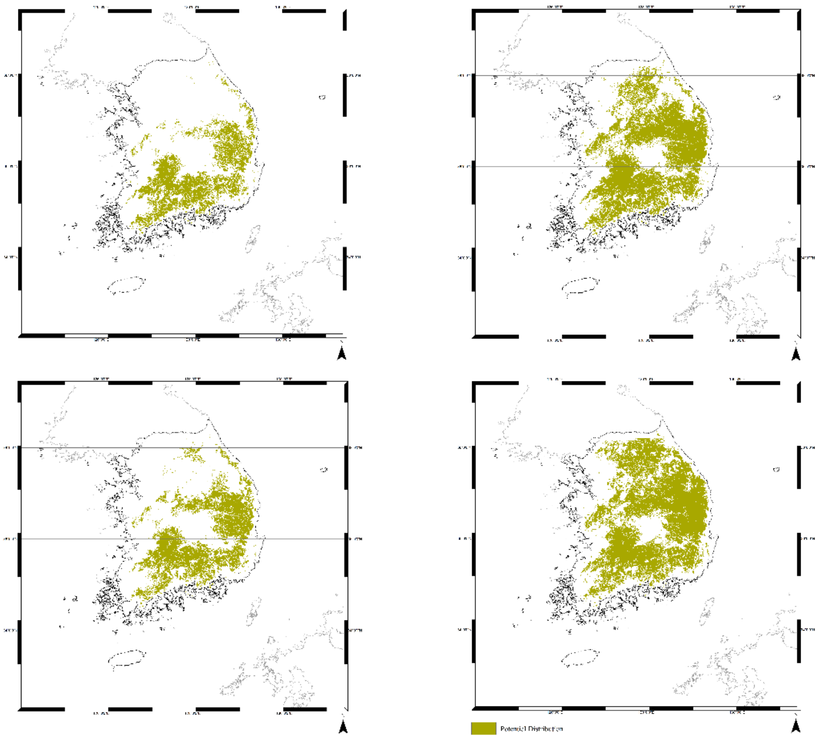


Figure 25 Potential distribution of Korean red pine under RCP scenarios in 2090s (Top left: RCP 2.6, Top right: RCP 4.5; Bottom left: RCP 6.0, Bottom right: RCP 8.5)

However, the total area of Korean red pine increased in all RCP scenarios. This should be carefully interpreted, as climate change can introduce different forest species in different climate zones (Sung et al., 2016). As temperature and precipitation changes, climate zones in South Korea will change, especially in the southern part of Korea. On the other hand, mountainous areas can be considered, and refuges for forest species are vulnerable to climate change. Therefore, vulnerable areas with different types of biome that are expected should be considered in a forest management plan or climate change adaptation plan.

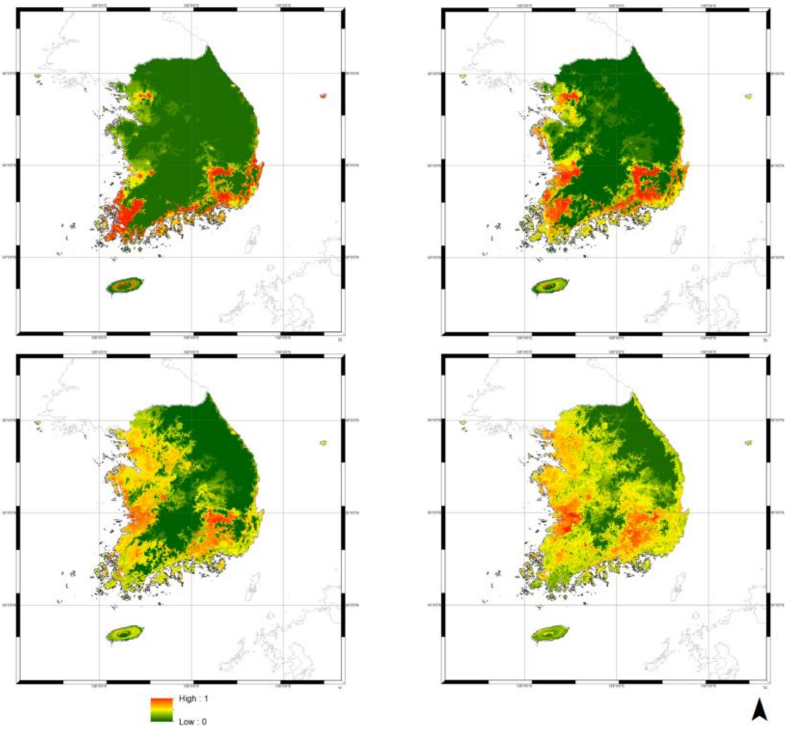


Figure 26 The possibilities of species change from 2010s to 2090s (Sung et al., 2016) (top left: RCP 2.6, top right: RCP 4.5 bottom left: RCP 6.0, bottom right: RCP 8.5)

4.2. Potential Distribution of Korean Red Pine in RF model

As a result of the distribution of major forest species and the proportion of temperate forests, which constitute the majority of Korea, decreased, and subtropical forests increased. In temperate forests, coniferous forests showed a gradual northward appearance. Deciduous forests showed an increase in total area in all scenarios except the RCP 8.5 scenario. Tropical forests showed the greatest increase in the RCP 8.5 scenario.

Table 28 Forest areas by each forest type under different RCP scenarios (Unit: km^2)

	RCP26	RCP45	RCP60	RCP85
Pinus Densiflora	21,076	22,385	24,961	14,004
Pinus thumbergii	744	506	1,685	191
Quercus app	14,729	8,262	9,364	6,448
Deciduous broadleaved forest	4,467	5,941	4,623	3,734
Evergreen broadleaved forest	36	6	33	0
Mixed forest	10,688	2,752	8,788	466
Sub-tropical forest	12,693	18,200	12,721	3,1209
Tropical forest	1,325	7,704	3,581	9,705
Total	65,756	65,756	65,756	65,756

In the RCP 2.6 scenario, where Korean red pine forests are expected to be affected least due to climate change, these are mainly distributed in the southern regions, but in the RCP 8.5 scenarios they are distributed to Gangwon and central regions. This is consistent with previous studies that predicted that Korean red pine trees would be distributed northward to adapt to climate change (Chun and Lee, 2013).

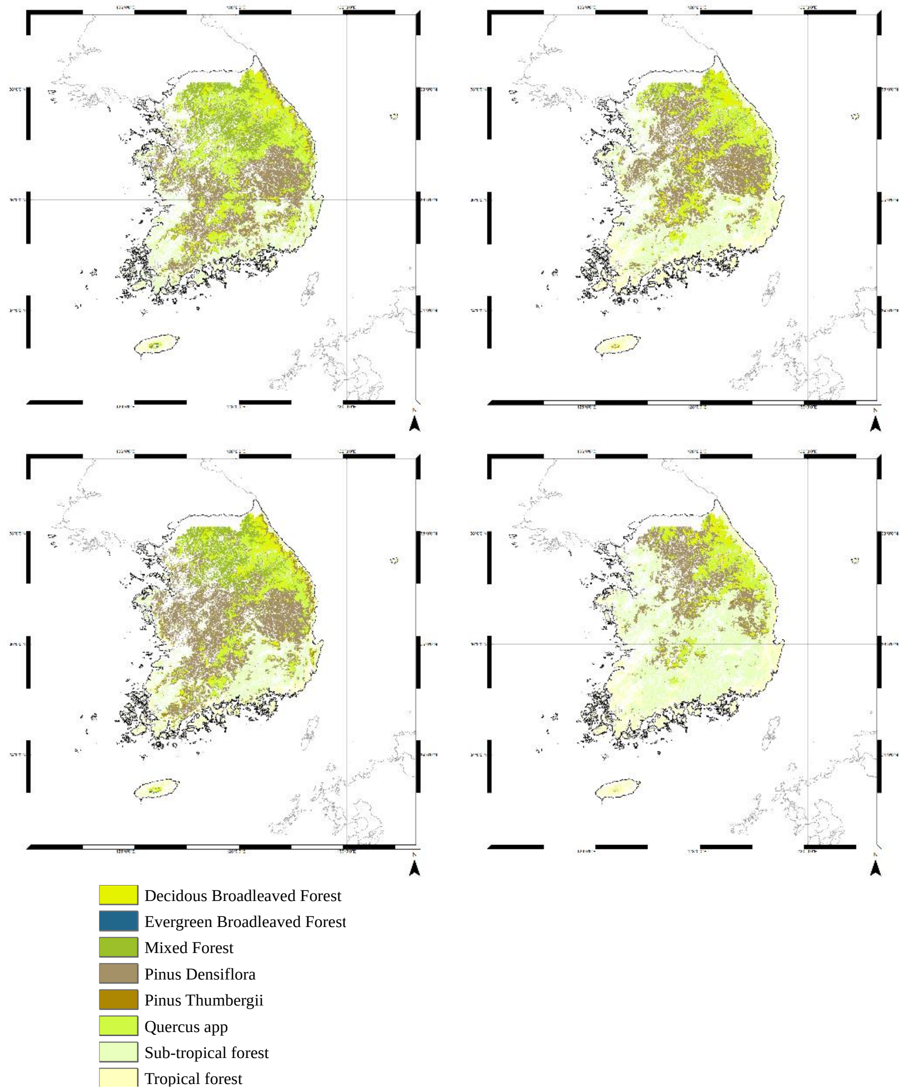


Figure 27 Forest type distribution change under RCP scenarios in 2090s
(Top left: RCP 2.6, Top right: RCP 4.5; Bottom left: RCP 6.0, Bottom right: RCP 8.5)

4.3. Comparing Potential Distribution Range by RCP Scenarios

In this study, response of forest species by RCP scenarios has been examined with four different RCP scenarios considering the competition between major forest species. In the single-species model, the uncertainties are high in the northern part of South Korea, as the single-species model calculated the potential distribution of Korean red pine without considering the suitability of other species. However, in the multi-species distribution model, changes in the species distribution are easier to find. However, the uncertainties increased as multi-species are modeled in one model. On the other hand, the core area for monitoring the trend of potential changes of species can be selected, considering the different ranges of temperature and precipitation by RCP scenarios.

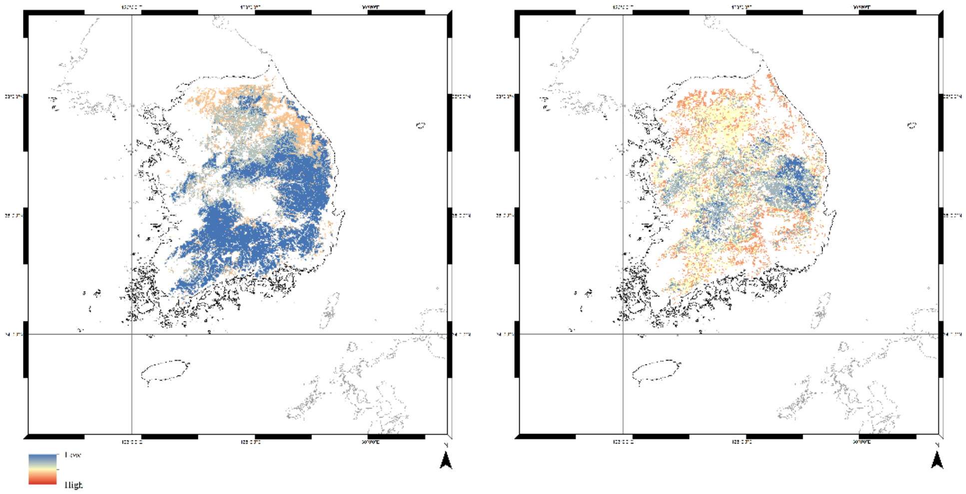


Figure 28 Distribution change ranges from different RCP scenarios
(left: single species model. Right: multi species model)

As the future greenhouse gases concentrations in the atmosphere are not in a fixed state, a carbon sequestration strategy and climate adaptation plan should be based on state-of-art projection of greenhouse gases. The trajectory of greenhouse gases changes every year, and the global community tries to reduce greenhouse gases from industries (Le Quéré et al., 2018). Thus, the most persuadable climate change scenarios can be changed as greenhouse gas emissions change by global agreements.

To reduce uncertainties from temperature and precipitation ranges due to the RCP scenarios, diverse climate scenarios and RCP should be considered in modeling the potential distribution of forest species. To apply a high-resolution climate model, this study was applied on the climate model with four RCP scenarios. Despite of the limitations of this study, potential climate pattern changes under the four different RCPs are estimated to provide a basis for establishing a forest management plan or climate change mitigation and adaptation plan.

V. Conclusion

The adverse impact of climate change on the forest ecosystem is expected to increase. Thus, various countermeasures are proposed to mitigate and adapt to them. To minimize the negative impact of climate change effectively within limited time and resources, it is necessary to make an accurate impact assessment with the consideration of uncertainties, as uncertainties in assessing the potential impact of climate change are inevitable. Thus, step-by-step quantification of uncertainties in modeling the potential impact of climate change to forest species is necessary.

In this study, different source of uncertainties, 1) sampling methods and sample size 2) application of different SDM algorithms and, 3) competition among the major forest species has been examined for modeling the potential distribution of major forest species.

To understand uncertainties in sampling methods, different sampling methods and sizes were selected, and a one-way t-test was conducted to verify effectiveness of sampling methods. Also, the performance of SDMs was tested under the different sampling methods and sample sizes and ANOVA was applied to test the statistical significance of differences in model performance. The uncertainties in different SDM algorithms were analyzed with model performance and spatial distribution of each model. Then, we compared spatial distribution of each model to test uncertainties. The random forest algorithm was applied to consider the statistical modeling of the competition among major forest species. Then, we compared the modeled distribution of Korean red pine in the single-species model. Finally, we applied RCP scenarios to measure different ranges of temperature and precipitation changes in establishing a forest management plan.

As a result of this study, the uncertainties in sampling methods and sampling size affect model performance. The stratified random sampling method was effective as it well represents the population of forest species. In addition, this study found that selecting suitable sample sizes for SDMs can save time and resources in gathering presence data. In developing countries, a surveying presence dataset throughout the country requires an enormous amount of time and effort. If we can apply effective sampling methods and determine effective sample sizes as demonstrated by this study, we can estimate the potential species distribution under recent climate change in time to make adaptation plans to protect the ecosystem.

In this study, even though the performance of SDMs are similar, the spatial interpretation of SDMs should be carefully conducted as the forest management plan. The performance of the model derived from statistical approaches should be based on spatial regime. Also, as a result of modeling the potential distribution of major forest species by RCP scenarios, we found that the changes in temperature and precipitation range drives significant changes in the potential distribution of forest species regardless of applied SDMs. Thus, the ranges of different RCP scenarios should be carefully examined spatially for planning forest management strategies and national adaptation plans.

Many kinds of SDMs are used for modeling species distributions. Due to the complexity of these models, it is important to understand the uncertainties inherent in each model. In this study, due to the characteristics of SDMs, feedback for modeling in the temporal scale was not included. Recent studies (Case and Lawler, 2016; Hill et al., 2017) have used two-stage modeling or hybrid modeling techniques to overcome these uncertainties. Species niches may affect model performance because the variations in climatic and environmental variables interact differently (Buisson et al., 2010).

Additionally, most models, except MAXENT, use pseudo-absence data; this means true absence data should also be carefully examined. Using environmental factors in SDMs requires further study, as we do not comprehensively understand the interactions among these factors in the context of models, at present. These uncertainties can then translate into uncertainties in policies and decision-making processes during planning and conservation.

Despite these limitations, various aspects of uncertainty in predicting changes distribution of major forest species have been discussed in this study. The impact assessment on forest species under the climate change included different kinds of uncertainties in to spatial distribution due to different modeling techniques. Understanding these uncertainties will help to establish effective forest management plan and climate change adaptation strategies on a national scale.

VI. Bibliography

Ahn, Y. jung, Lee, D.-K., Kim, H. gul, Park, C., Jiyeon, K., Kim, J. wook, 2015. Estimating Korean Pine(Pinus koraiensis) Habitat Distribution Considering Climate Change Uncertainty - Using Species Distribution Models and RCP Scenarios -. J. Korea Soc. Environ. Restor. Technol. 18, 51–64.

Ananda, J., Herath, G., 2009. A critical review of multi-criteria decision making methods with special reference to forest management and planning. Ecol. Econ. 68, 2535–2548. https://doi.org/10.1016/j.ecolecon.2009.05.010

Beale, C.M., Lennon, J.J., 2012. Incorporating uncertainty in predictive species distribution modelling. Philos. Trans. R. Soc. B Biol. Sci. 367, 247–258. https://doi.org/10.1098/rstb.2011.0178

Beaumont, L.J., Hughes, L., Poulsen, M., 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecol. Modell. 186, 251–270. https://doi.org/10.1016/j.ecolmodel.2005.01.030

Belda, M., Holtanová, E., Halenka, T., Kalvová, J., 2014. Climate classification revisited: From Köppen to Trewartha. Clim. Res. 59, 1–13. https://doi.org/10.3354/cr01204

Bonan, G.B., 2008. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. Sci. 320, 1444–1449. https://doi.org/10.1126/science.1155121

Brady, N.C., 2008. The nature and properties of soils / Nyle C. Brady, Ray

R. Weil., 14th ed. . ed. Upper Saddle River, N.J. : Pearson/Prentice Hall, c2008., Upper Saddle River, N.J.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Chang. Biol. 16, 1145–1157. https://doi.org/10.1111/j.1365-2486.2009.02000.x

Carmean, W.H., Hahn, J.T., Jacobs, R.D., 1989. Site Index Curves for Forest Tree Species in the Eastern United States.

Case, M.J., Lawler, J.J., 2016. Integrating mechanistic and empirical model projections to assess climate impacts on tree species distributions in northwestern North America. Glob. Chang. Biol. https://doi.org/10.1111/gcb.13570

Choi, S., Lee, W., Kwak, D., Lee, S., Son, Y., Lim, J., Saborowski, J., 2011. Predicting forest cover changes in future climate using hydrological and thermal indices in South Korea. Clim. Res. 49, 229–245. https://doi.org/10.3354/cr01026

Chun, J.H., Lee, C.-B., 2013. Assessing the Effects of Climate Change on the Geographic Distribution of Pinus densiflora in Korea using Ecological Niche Model. Korean J. Agric. For. Meteorol. 15, 219–233. https://doi.org/10.5532/KJAFM.2013.15.4.219

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography (Cop.). 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Food and Agriculture Organization, 2012. Global ecological zones for FAO forest reporting: 2010 Update. For. Resour. Assess. Work. Pap. 179 42. https://doi.org/citeulike-article-id:12358238

Food and Agriculture Organization, 2004. Sampling Designs for National Forest Assessments [WWW Document]. FAO. URL http://www.fao.org/forestry/44859-02cf95ef26dfdcb86c6be2720f8b938a8.pdf

Franklin, J., 2010a. Mapping species distributions. University Press, Cambridge. https://doi.org/dx.doi.org/10.1017/CBO9780511810602

Franklin, J., 2010b. Mapping species distributions. Spatial inference and prediction, Ecology, biodiversity and conservation. Cambridge, Cambridge. https://doi.org/10.1017/CBO9781107415324.004

Gang, C., Wang, Z., Zhou, W., Chen, Y., Li, J., Cheng, J., Guo, L., Odeh, I., Chen, C., 2015. Projecting the dynamics of terrestrial net primary productivity in response to future climate change under the RCP2.6 scenario. Environ. Earth Sci. 74, 5949–5959.

https://doi.org/10.1007/s12665-015-4618-x

Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Gutiérrez, A.G., Snell, R.S., Bugmann, H., 2016. Using a dynamic forest model to predict tree species distributions. Glob. Ecol. Biogeogr. 25, 347–358. https://doi.org/10.1111/geb.12421

Hannemann, H., Willis, K.J., Macias-Fauria, M., 2016. The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. Glob. Ecol. Biogeogr. 25, 26–35. https://doi.org/10.1111/geb.12381

Hansson, S.L., Röjvall, A.S., Rastam, M., Gillberg, C., Gillberg, C., Anckarsäter, H., 2005. Psychiatric telephone interview with parents for screening of childhood autism - Tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): Preliminary reliability and validity. Br. J. Psychiatry 187, 262–267. https://doi.org/10.1192/bjp.187.3.262

Hastie, T., 2009. The Elements of Statistical Learning : Data Mining, Inference, and Prediction. Springer Ser. Stat.

Hengl, T., 2009. A Practical guide to Geostatistical Mapping, Scientific and Technical Research series. https://doi.org/10.1016/0277-

9390(86)90082-8

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. https://doi.org/10.1002/joc.1276

Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., Brown, N., 2017. Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. Ecol. Evol. 7, 1043–1056. https://doi.org/10.1002/ece3.2661

IPCC, 2014a. Summary for Policy Makers. Clim. Chang. 2014 Impacts, Adapt. Vulnerability - Contrib. Work. Gr. II to Fifth Assess. Rep. 1–32. https://doi.org/10.1016/j.renene.2009.11.012

IPCC, 2014b. Climate Change 2013 - The Physical Science Basis. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781107415324

IPCC, 2003. Good Practice Guidance for Land Use, Land-Use Change and Forestry. Practice 177, 576. https://doi.org/citeulike-article-id:1260638

Jarnevich, C.S., Hayes, M.A., Fitzgerald, L.A., Yackel Adams, A.A., Falk, B.G., Collier, M.A.M., Bonewell, L.R., Klug, P.E., Naretto, S., Reed, R.N., 2018. Modeling the distributions of tegu lizards in native and potential invasive ranges. Sci. Rep. 8, 10193. https://doi.org/10.1038/s41598-018-28468-w

Jin, Y., Sung, S., Lee, D.K., Biging, G.S., Jeong, S., 2016. Mapping deforestation in north korea using phenology-based multi-index and random forest. Remote Sens. 8, 1–15. https://doi.org/10.3390/rs8120997

Jones, R.N., 2000. Managing uncertainty in climate change projections–issues for impact assessment. Clim. Change 45, 403–419. https://doi.org/10.1016/j.urology.2013.09.061

Kim, H.G., Lee, D.K., Park, C., Ahn, Y., Kil, S.H., Sung, S., Biging, G.S., 2018. Estimating landslide susceptibility areas considering the uncertainty inherent in modeling methods, Stochastic Environmental Research and Risk Assessment. Springer Berlin Heidelberg. https://doi.org/10.1007/s00477-018-1609-y

Kim, H.G., Lee, D.K., Park, C., Kil, S., Son, Y., Park, J.H., 2015. Evaluating landslide hazards using RCP 4.5 and 8.5 scenarios. Environ. Earth Sci. 73, 1385–1400. https://doi.org/10.1007/s12665-014-3775-7

Kim, J., Jung, H., Jeon, S.W., Lee, D.-K., 2015. Predicting the Potential Distribution of Korean Pine (Pinus koraiensis) Using an Ensemble of Climate Scenarios. J. Korea Soc. Environ. Restor. Technol. 18, 79–88.

Kim, Y., Kim, J.-W., 2017. Distributional Uniqueness of Deciduous Oaks (Quercus L .) in the Korean Peninsula. J. korean Soc. Environ. Restor. Technol. 20, 37–59.

Kira, T., 1945. A new classification of climate in eastern Asia as the basis for agricultural geography. Horticulturla Institute, Kyoto University, Kyoto.

Knutti, R., Sedláček, J., 2012. Robustness and uncertainties in the new CMIP5 climate model projections. Nat. Clim. Chang. 3, 369.

Korea Forest Research Institute, 2014. Prediting the changes of productive areas for major tree species under climate change in Korea.

Korea Forest Research Institute, 2012. Climate change, forest and human.

Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J., Scott, J.K., 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. Methods Ecol. Evol. 3, 53–64. https://doi.org/10.1111/j.2041-210X.2011.00134.x

Kryazhimskiy, A., Rovenskaya, E., Shvidenko, A., Gusti, M., Shchepashchenko, D., Veshchinskaya, V., 2015. Towards harmonizing competing models: Russian forests' net primary production case study. Technol. Forecast. Soc. Change 98, 245–254. https://doi.org/10.1016/j.techfore.2015.06.003

Kwon, S.H., 2014. Applying Ensemble Model for Identifying Uncertainty in the Species Distribution Models. J. Korean Soc. geo-spatial Inf. Syst. 22, 47–52.

Lal, R., 2005. Forest soils and carbon sequestration. For. Ecol. Manage. 220, 242–258. https://doi.org/10.1016/j.foreco.2005.08.015

Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A.C., Korsbakken, J.I., Peters, G.P., Canadell, J.G., Jackson, R.B., Boden, T.A., Tans, P.P., Andrews, O.D., Arora, V.K., Bakker, D.C.E., Van Der Laan-Luijkx, I.T., Van Der Werf, G.R., Van Heuven, S., Viovy, N., Vuichard, N., Walker, A.P., Watson, A.J., Wiltshire, A.J., Zaehle, S., Zhu, D., 2018. Carbon cycle-Global Carbon Budget 2017. Earth Sci. Syst. Data 10, 405–448. https://doi.org/10.5194/essd-10-405-2018

Lindner, M., Fitzgerald, J.B., Zimmermann, N.E., Reyer, C., Delzon, S., van der Maaten, E., Schelhaas, M.-J., Lasch, P., Eggers, J., van der Maaten-Theunissen, M., Suckow, F., Psomas, A., Poulter, B., Hanewinkel, M., 2014. Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management? J. Environ. Manage. 146, 69–83. https://doi.org/10.1016/j.jenvman.2014.07.030

Mi, C., Huettmann, F., Guo, Y., Han, X., Wen, L., 2017. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ 5, e2849. https://doi.org/10.7717/peerj.2849

Moudrý, V., Šímová, P., 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. Int. J. Geogr. Inf. Sci. 26, 2083–2095. https://doi.org/10.1080/13658816.2012.721553

Nakao, K., Higa, M., Tsuyama, I., Lin, C.-T., Sun, S.-T., Lin, J.-R., Chiou, C.-R., Chen, T.-Y., Matsui, T., Tanaka, N., 2014. Changes in the potential habitats of 10 dominant evergreen broad-leaved tree species in the Taiwan-Japan archipelago. Plant Ecol. 215, 639–650. https://doi.org/10.1007/s11258-014-0329-8

Neovius, M.G., Linné, Y.M., Barkeling, B.S., Rossner, S.O., 2004. Sensitivity and specificity of classification systems for fatness in adolescents. Am. J. Clin. Nutr. 80, 597–603. https://doi.org/80/3/597 [pii]

Park, S.U., Koo, K.A., Kong, W.-S., 2016. Potential Impact of Climate

Change on Distribution of Warm Temperate Evergreen Broad-leaved Trees in the Korean Peninsula. J. Korean Geogr. Soc. 51, 201–217.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob. Ecol. Biogeogr. 12, 361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x

R Core Team, 2014. R: A language and environment for statistical computing.

Schulze, E.-D. (Ernst-D., 2005. Plant ecology / Ernst-Detlef Schulze, Erwin Beck, Klaus Müller-Hohenstein . Berlin : Springer, c2005., Berlin.

Shin, H.-J., Park, G.-A., Park, M., Kim, S.-J., 2012. Projection of Forest Vegetation Change by Applying Future Climate Change Scenario MIROC3.2 A1B. J. Korean Assoc. Geogr. Inf. Stud. 15, 64–75.

SPSS Inc, 2009. PASW Statistics for Windows, Version 18.0.

Sung, S., Kwon, Y.-S., Lee, D.K., Cho, Y., 2018. Predicting the Potential Distribution of an Invasive Species, Solenopsis invicta Buren (Hymenoptera: Formicidae), under Climate Change using Species Distribution Models. Entomol. Res. 48, 505–513. https://doi.org/10.1111/1748-5967.12325

Sung, S., Lee, D., Park, C., Kim, H., Kil, S., Chae, H., Park, G., Ohga, S., 2018. Assessing Effective Sampling Method and Sample Size for Species Distribution Modeling of Korean Red Pine (Pinus densiflora). Fac. Agric. Kyushu Univ. 63, 211–221.

Sung, S., Nicklas, F., Georg, K., Lee, D.K., 2016. Estimating Net Primary

Productivity under Climate Change by Application of Global Forest Model (G4M). J. Korean Soc. People Plants Environ. 19, 549–558. https://doi.org/10.11628/ksppe.2016.19.6.549

Takahashi, K., Okuhara, I., 2012. Comparison of climatic effects on radial growth of evergreen broad-leaved trees at their northern distribution limit and co-dominating deciduous broad-leaved trees and evergreen conifers. Ecol. Res. 27, 125–132. https://doi.org/10.1007/s11284-011-0879-3

Thuiller, W., 2003. BIOMOD: Optimising predictions of species distributions and projecting potential future shift under global change. Glob. Chang. Biol. 9, 1353–1362.

Thuiller, W., Georges, D., Engler, R., 2015. biomod2 package manual.

Thuiller, W., Lafourcade, B., Araujo, M., 2010. The Presentation Manual for BIOMOD.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD - A platform for ensemble forecasting of species distributions. Ecography (Cop.). 32, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Van Asselt, M.B. a, Rotmans, J., 2002. Uncertainty in Integrated Assessment modelling. From positivism to pluralism. Clim. Change 54, 75–105. https://doi.org/10.1023/A:1015783803445

Véga, C., St-Onge, B., 2009. Mapping site index and age by linking a time series of canopy height models with growth curves. For. Ecol. Manage. 257, 951–959. https://doi.org/10.1016/j.foreco.2008.10.029

Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B. a., Janssen, P., Krayer von Krauss, M.P., 2003. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integr. Assess. 4, 5–17. https://doi.org/10.1076/iaij.4.1.5.16466

Walker, W.E., Lempert, R.J., Kwakkel, J.H., 2008. Deep Uncertainty. Encycl. Oper. Res. Manag. Sci. 1, 395–402.

Wang, T., Wang, G., Innes, J., Nitschke, C., Kang, H., 2015. Climatic niche models and their consensus projections for future climates for four major forest tree species in the Asia–Pacific region. For. Ecol. Manage. 360, 357–366. https://doi.org/10.1016/j.foreco.2015.08.004

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 19729–19736. https://doi.org/10.1073/pnas.0901639106

Wilby, R.L., Dessai, S., 2010. Robust adaptation to climate change. Weather 65, 180–185. https://doi.org/10.1002/wea.543

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Zimmermann, N.E., 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Yim, Y.-J., 1977. Distribution of Forest Vegetation and Climate in the

Korean Penninsula: III. Distribution of Tree Species Along the Thermal Gradient. Japanese J. Ecol. 27, 177–189.

Appendix 1. Correlation Analysis of All Environmental variables

	Aspect	bio1	bio10	bio11	bio12	bio13	bio14	bio15	bio16	bio17	bio18	bio19	bio2	bio3
Aspect	1	0.05	0.05	0.05	-0.02	-0.05	0.03	-0.06	-0.04	0.03	-0.04	0.04	0.03	0.06
bio1	0.05	1	0.93	0.95	-0.22	-0.45	0.12	-0.39	-0.41	0.17	-0.39	0.29	-0.36	-0.08
bio10	0.05	0.93	1	0.77	-0.38	-0.44	0.01	-0.11	-0.44	-0.05	-0.41	0.05	-0.12	0.03
bio11	0.05	0.95	0.77	1	-0.08	-0.43	0.24	-0.61	-0.36	0.35	-0.35	0.48	-0.49	-0.12
bio12	-0.02	-0.22	-0.38	-0.08	1	0.78	0.56	-0.22	0.89	0.71	0.87	0.64	-0.17	-0.07
bio13	-0.05	-0.45	-0.44	-0.43	0.78	1	0.18	0.4	0.96	0.23	0.95	0.12	0.16	0.05
bio14	0.03	0.12	0.01	0.24	0.56	0.18	1	-0.54	0.32	0.89	0.34	0.85	-0.28	-0.12
bio15	-0.06	-0.39	-0.11	-0.61	-0.22	0.4	-0.54	1	0.23	-0.69	0.24	-0.77	0.49	0.13
bio16	-0.04	-0.41	-0.44	-0.36	0.89	0.96	0.32	0.23	1	0.38	0.99	0.27	0.06	0
bio17	0.03	0.17	-0.05	0.35	0.71	0.23	0.89	-0.69	0.38	1	0.38	0.97	-0.37	-0.12
bio18	-0.04	-0.39	-0.41	-0.35	0.87	0.95	0.34	0.24	0.99	0.38	1	0.28	0.1	0.05
bio19	0.04	0.29	0.05	0.48	0.64	0.12	0.85	-0.77	0.27	0.97	0.28	1	-0.39	-0.07
bio2	0.03	-0.36	-0.12	-0.49	-0.17	0.16	-0.28	0.49	0.06	-0.37	0.1	-0.39	1	0.88
bio3	0.06	-0.08	0.03	-0.12	-0.07	0.05	-0.12	0.13	0	-0.12	0.05	-0.07	0.88	1
bio4	-0.04	-0.58	-0.23	-0.79	-0.24	0.25	-0.35	0.82	0.14	-0.58	0.16	-0.68	0.63	0.21
bio5	0.06	0.82	0.95	0.64	-0.44	-0.42	-0.08	-0.01	-0.46	-0.15	-0.41	-0.05	0.16	0.29
bio6	0.04	0.9	0.68	0.97	-0.02	-0.41	0.3	-0.65	-0.33	0.41	-0.33	0.52	-0.66	-0.3
bio7	-0.02	-0.56	-0.23	-0.75	-0.23	0.23	-0.39	0.75	0.11	-0.56	0.14	-0.64	0.86	0.52
bio8	0.05	0.87	0.98	0.7	-0.42	-0.43	-0.01	-0.04	-0.45	-0.1	-0.4	-0.01	-0.04	0.07
bio9	0.06	0.91	0.75	0.95	-0.11	-0.46	0.25	-0.61	-0.39	0.34	-0.38	0.49	-0.4	-0.03
CI	-0.05	-0.97	-0.81	-0.99	0.13	0.44	-0.2	0.54	0.39	-0.29	0.37	-0.42	0.46	0.11
dis_river	-0.03	-0.01	-0.06	0.04	0	-0.08	0.11	-0.15	-0.06	0.13	-0.07	0.13	-0.23	-0.24
dis_sea	0.07	-0.2	-0.07	-0.28	-0.15	0.08	-0.13	0.34	-0.01	-0.2	0.03	-0.2	0.56	0.51
dem	0.07	-0.75	-0.83	-0.6	0.42	0.42	0.12	0.01	0.44	0.18	0.42	0.11	0.16	0.1
orm	-0.03	-0.51	-0.56	-0.41	0.23	0.3	-0.04	0.05	0.28	0.01	0.27	-0.02	0.2	0.18
radiation	0.08	0.47	0.44	0.46	-0.31	-0.34	-0.28	-0.15	-0.37	-0.2	-0.34	-0.07	0.15	0.37
slope	0.1	-0.4	-0.45	-0.32	0.23	0.26	0.01	0	0.25	0.06	0.24	0.04	0.19	0.19
soil_depth	0.02	0.51	0.55	0.43	-0.24	-0.33	0.05	-0.08	-0.31	0.01	-0.31	0.03	-0.27	-0.25
WI	0.04	1	0.93	0.94	-0.24	-0.45	0.11	-0.37	-0.42	0.14	-0.4	0.26	-0.38	-0.11

	bio4	bio5	bio6	bio7	bio8	bio9	CI	dis_river	dis_sea	dem	orm	radiation	slope	soil_depth	WI
Aspect	-0.04	0.06	0.04	-0.02	0.05	0.06	-0.05	-0.03	0.07	0.07	-0.03	0.08	0.1	0.02	0.04
bio1	-0.58	0.82	0.9	-0.56	0.87	0.91	-0.97	-0.01	-0.2	-0.75	-0.51	0.47	-0.4	0.51	1
bio10	-0.23	0.95	0.68	-0.23	0.98	0.75	-0.81	-0.06	-0.07	-0.83	-0.56	0.44	-0.45	0.55	0.93
bio11	-0.79	0.64	0.97	-0.75	0.7	0.95	-0.99	0.04	-0.28	-0.6	-0.41	0.46	-0.32	0.43	0.94
bio12	-0.24	-0.44	-0.02	-0.23	-0.42	-0.11	0.13	0	-0.15	0.42	0.23	-0.31	0.23	-0.24	-0.24
bio13	0.25	-0.42	-0.41	0.23	-0.43	-0.46	0.44	-0.08	0.08	0.42	0.3	-0.34	0.26	-0.33	-0.45
bio14	-0.35	-0.08	0.3	-0.39	-0.01	0.25	-0.2	0.11	-0.13	0.12	-0.04	-0.28	0.01	0.05	0.11
bio15	0.82	-0.01	-0.65	0.75	-0.04	-0.61	0.54	-0.15	0.34	0.01	0.05	-0.15	0	-0.08	-0.37
bio16	0.14	-0.46	-0.33	0.11	-0.45	-0.39	0.39	-0.06	-0.01	0.44	0.28	-0.37	0.25	-0.31	-0.42
bio17	-0.58	-0.15	0.41	-0.56	-0.1	0.34	-0.29	0.13	-0.2	0.18	0.01	-0.2	0.06	0.01	0.14
bio18	0.16	-0.41	-0.33	0.14	-0.4	-0.38	0.37	-0.07	0.03	0.42	0.27	-0.34	0.24	-0.31	-0.4
bio19	-0.68	-0.05	0.52	-0.64	-0.01	0.49	-0.42	0.13	-0.2	0.11	-0.02	-0.07	0.04	0.03	0.26
bio2	0.63	0.16	-0.66	0.86	-0.04	-0.4	0.46	-0.23	0.56	0.16	0.2	0.15	0.19	-0.27	-0.38
bio3	0.21	0.29	-0.3	0.52	0.07	-0.03	0.11	-0.24	0.51	0.1	0.18	0.37	0.19	-0.25	-0.11
bio4	1	-0.08	-0.84	0.93	-0.14	-0.74	0.74	-0.12	0.35	0.13	0.1	-0.29	0.06	-0.14	-0.55
bio5	-0.08	1	0.51	-0.01	0.95	0.65	-0.69	-0.11	0.08	-0.76	-0.48	0.5	-0.37	0.45	0.82
bio6	-0.84	0.51	1	-0.86	0.6	0.91	-0.96	0.09	-0.35	-0.53	-0.39	0.35	-0.31	0.42	0.89
bio7	0.93	-0.01	-0.86	1	-0.14	-0.68	0.71	-0.17	0.45	0.18	0.17	-0.12	0.13	-0.23	-0.55
bio8	-0.14	0.95	0.6	-0.14	1	0.69	-0.75	-0.1	-0.01	-0.8	-0.53	0.42	-0.44	0.52	0.88
bio9	-0.74	0.65	0.91	-0.68	0.69	1	-0.94	0.03	-0.18	-0.56	-0.39	0.47	-0.29	0.4	0.9
CI	0.74	-0.69	-0.96	0.71	-0.75	-0.94	1	-0.02	0.25	0.64	0.42	-0.47	0.34	-0.44	-0.96
dis_river	-0.12	-0.11	0.09	-0.17	-0.1	0.03	-0.02	1	-0.11	0.05	-0.01	-0.08	0.03	0.02	0
dis_sea	0.35	0.08	-0.35	0.45	-0.01	-0.18	0.25	-0.11	1	0.21	0.14	0.15	0.16	-0.2	-0.21
dem	0.13	-0.76	-0.53	0.18	-0.8	-0.56	0.64	0.05	0.21	1	0.56	-0.27	0.6	-0.57	-0.76
orm	0.1	-0.48	-0.39	0.17	-0.53	-0.39	0.42	-0.01	0.14	0.56	1	-0.1	0.57	-0.9	-0.53
radiation	-0.29	0.5	0.35	-0.12	0.42	0.47	-0.47	-0.08	0.15	-0.27	-0.1	1	-0.05	0.05	0.46
slope	0.06	-0.37	-0.31	0.13	-0.44	-0.29	0.34	0.03	0.16	0.6	0.57	-0.05	1	-0.63	-0.42
soil_depth	-0.14	0.45	0.42	-0.23	0.52	0.4	-0.44	0.02	-0.2	-0.57	-0.9	0.05	-0.63	1	0.53
WI	-0.55	0.82	0.89	-0.55	0.88	0.9	-0.96	0	-0.21	-0.76	-0.53	0.46	-0.42	0.53	1

Without mark all variables are significant $p<0.001$

Appendix 2. Relative Importance of all environmental variables by sampling methods in all sampling methods

Relative importance of variables in area sampling method (Needle leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.188	0.095	0.114	0.056	0.134	0.056	0.117	0.257
Slope	0.132	0.118	0.064	0.062	0.089	0.055	0.077	0.048
Radiation	0.182	0.160	0.059	0.063	0.030	0.062	0.040	0.044
Distance from water	0.121	0.033	0.013	0.015	0.024	0.014	0.024	0.287
Distance from sea	0.137	0.098	0.087	0.039	0.076	0.035	0.076	0.532
Soil Depth	0.126	0.094	0.068	0.021	0.127	0.022	0.284	0.280
Soil Organic Matter Content in Layer A	0.030	0.074	0.000	0.017	0.162	0.023	0.416	0.001
Warmth Index	0.201	0.085	0.131	0.018	0.055	0.011	0.067	0.078
Isothemality	0.237	0.435	0.192	0.098	0.222	0.179	0.145	0.026
Min Temperature of Coldest Month	0.226	0.201	0.343	0.071	0.321	0.142	0.366	0.043
Precipitation of wettest month	0.098	0.072	0.040	0.025	0.046	0.025	0.054	0.130
Precipitation of driest month	0.138	0.053	0.034	0.017	0.002	0.024	0.044	0.006

Relative importance of variables in random sampling method (Needle leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.107	0.136	0.163	0.061	0.106	0.079	0.159	0.174
Slope	0.113	0.160	0.179	0.043	0.054	0.051	0.088	0.042
Radiation	0.155	0.185	0.140	0.046	0.127	0.110	0.120	0.064
Distance from water	0.090	0.084	0.009	0.026	0.028	0.028	0.037	0.401
Distance from sea	0.077	0.041	0.030	0.031	0.015	0.019	0.040	0.684
Soil Depth	0.191	0.157	0.208	0.057	0.342	0.099	0.718	0.254
Soil Organic Matter Content in Layer A	0.011	0.198	0.000	0.023	0.161	0.048	0.520	0.001
Warmth Index	0.123	0.129	0.186	0.031	0.192	0.051	0.354	0.065
Isothemality	0.112	0.146	0.024	0.046	0.064	0.058	0.070	0.008
Min Temperature of Coldest Month	0.064	0.051	0.036	0.023	0.036	0.010	0.118	0.013
Precipitation of wettest month	0.096	0.184	0.109	0.051	0.141	0.102	0.138	0.135
Precipitation of driest month	0.092	0.049	0.017	0.032	0.028	0.028	0.061	0.015

Relative importance of variables in stratified sampling method (Needle leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.275	0.479	0.509	0.101	0.213	0.198	0.182	0.391
Slope	0.208	0.162	0.059	0.064	0.069	0.087	0.155	0.072
Radiation	0.116	0.097	0.035	0.034	0.069	0.061	0.075	0.050
Distance from water	0.172	0.045	0.010	0.033	0.035	0.044	0.052	0.318
Distance from sea	0.081	0.023	0.035	0.028	0.020	0.022	0.049	0.521
Soil Depth	0.180	0.127	0.121	0.058	0.456	0.057	0.623	0.239
Soil Organic Matter Content in Layer A	0.023	0.035	0.006	0.024	0.282	0.042	0.470	0.005
Warmth Index	0.173	0.057	0.121	0.028	0.316	0.051	0.446	0.068
Isothemality	0.101	0.033	0.012	0.025	0.078	0.027	0.070	0.005
Min Temperature of Coldest Month	0.092	0.018	0.063	0.018	0.185	0.013	0.254	0.026
Precipitation of wettest month	0.135	0.144	0.056	0.032	0.057	0.075	0.048	0.162
Precipitation of driest month	0.128	0.069	0.121	0.061	0.097	0.037	0.072	0.018

Relative importance of variables in area sampling method (Broad leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.369	0.686	0.313	0.134	0.275	0.247	0.327	0.590
Slope	0.144	0.036	0.026	0.044	0.065	0.054	0.060	0.040
Radiation	0.089	0.065	0.013	0.011	0.020	0.014	0.049	0.041
Distance from water	0.136	0.113	0.054	0.054	0.066	0.047	0.096	0.268
Distance from sea	0.076	0.030	0.022	0.027	0.034	0.022	0.076	0.252
Soil Depth	0.061	0.011	0.021	0.014	0.093	0.006	0.145	0.072
Soil Organic Matter Content in Layer A	0.016	0.038	0.024	0.015	0.171	0.018	0.350	0.001
Warmth Index	0.071	0.053	0.162	0.030	0.129	0.012	0.215	0.045
Isothemality	0.054	0.020	0.016	0.013	0.028	0.006	0.037	0.007
Min Temperature of Coldest Month	0.104	0.056	0.073	0.013	0.120	0.014	0.137	0.014
Precipitation of wettest month	0.357	0.148	0.053	0.075	0.174	0.100	0.158	0.160
Precipitation of driest month	0.197	0.103	0.039	0.037	0.096	0.076	0.106	0.020

Relative importance of variables in random sampling method (Broad leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.204	0.209	0.228	0.087	0.090	0.077	0.153	0.341
Slope	0.279	0.424	0.341	0.099	0.355	0.223	0.257	0.073
Radiation	0.044	0.021	0.009	0.020	0.042	0.012	0.049	0.052
Distance from water	0.073	0.072	0.022	0.049	0.029	0.038	0.035	0.381
Distance from sea	0.068	0.047	0.023	0.029	0.062	0.026	0.083	0.571
Soil Depth	0.073	0.116	0.019	0.053	0.259	0.058	0.594	0.164
Soil Organic Matter Content in Layer A	0.005	0.004	0.006	0.010	0.107	0.008	0.302	0.002
Warmth Index	0.099	0.012	0.066	0.036	0.070	0.014	0.153	0.068
Isothemality	0.065	0.038	0.054	0.035	0.115	0.030	0.117	0.004
Min Temperature of Coldest Month	0.117	0.190	0.185	0.059	0.238	0.075	0.343	0.013
Precipitation of wettest month	0.075	0.041	0.021	0.019	0.037	0.022	0.077	0.117
Precipitation of driest month	0.073	0.029	0.006	0.030	0.030	0.027	0.071	0.014

Relative importance of variables in stratified sampling method (Broad leaved forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.096	0.330	0.087	0.071	0.074	0.068	0.062	0.282
Slope	0.261	0.379	0.302	0.086	0.238	0.180	0.159	0.070
Radiation	0.095	0.052	0.018	0.026	0.024	0.017	0.061	0.039
Distance from water	0.068	0.035	0.012	0.037	0.042	0.033	0.051	0.364
Distance from sea	0.224	0.094	0.030	0.031	0.121	0.062	0.108	0.622
Soil Depth	0.096	0.078	0.086	0.028	0.124	0.028	0.346	0.142
Soil Organic Matter Content in Layer A	0.009	0.039	0.004	0.016	0.390	0.031	0.491	0.000
Warmth Index	0.103	0.054	0.179	0.047	0.071	0.022	0.233	0.061
Isothemality	0.053	0.052	0.019	0.033	0.048	0.027	0.069	0.007
Min Temperature of Coldest Month	0.090	0.079	0.135	0.058	0.156	0.061	0.247	0.010
Precipitation of wettest month	0.078	0.090	0.023	0.023	0.043	0.026	0.069	0.105
Precipitation of driest month	0.052	0.047	0.004	0.036	0.013	0.030	0.055	0.014

Relative importance of variables in area sampling method (Mixed forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.342	0.771	0.519	0.160	0.615	0.326	0.630	0.567
Slope	0.163	0.077	0.035	0.067	0.139	0.052	0.052	0.036
Radiation	0.031	0.035	0.041	0.017	0.099	0.024	0.063	0.038
Distance from water	0.059	0.031	0.052	0.034	0.011	0.044	0.026	0.247
Distance from sea	0.044	0.062	0.094	0.043	0.015	0.060	0.041	0.353
Soil Depth	0.047	0.032	0.010	0.018	0.361	0.019	0.565	0.120
Soil Organic Matter Content in Layer A	0.015	0.002	0.031	0.011	0.141	0.011	0.354	0.000
Warmth Index	0.029	0.007	0.101	0.010	0.134	0.004	0.134	0.060
Isothemality	0.025	0.034	0.039	0.015	0.026	0.017	0.031	0.005
Min Temperature of Coldest Month	0.074	0.054	0.141	0.039	0.106	0.064	0.074	0.009
Precipitation of wettest month	0.050	0.048	0.109	0.038	0.027	0.057	0.110	0.072
Precipitation of driest month	0.060	0.054	0.043	0.017	0.003	0.021	0.039	0.010

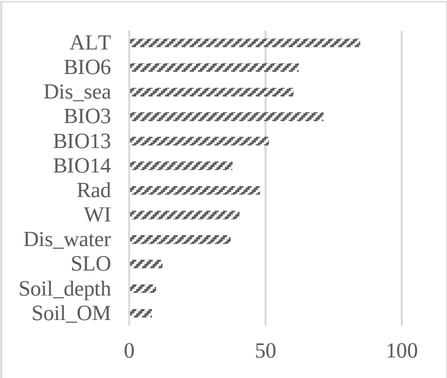
Relative importance of variables in random sampling method (Mixed forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.137	0.089	0.210	0.044	0.131	0.078	0.157	0.291
Slope	0.306	0.282	0.180	0.054	0.159	0.131	0.092	0.062
Radiation	0.078	0.070	0.011	0.026	0.033	0.047	0.040	0.053
Distance from water	0.076	0.036	0.000	0.021	0.056	0.026	0.044	0.468
Distance from sea	0.128	0.102	0.052	0.037	0.071	0.056	0.097	0.631
Soil Depth	0.162	0.281	0.121	0.035	0.283	0.086	0.665	0.218
Soil Organic Matter Content in Layer A	0.020	0.068	0.014	0.015	0.072	0.031	0.465	0.000
Warmth Index	0.127	0.099	0.336	0.036	0.101	0.051	0.385	0.077
Isothemality	0.129	0.084	0.059	0.021	0.033	0.031	0.067	0.006
Min Temperature of Coldest Month	0.159	0.068	0.165	0.025	0.112	0.030	0.245	0.008
Precipitation of wettest month	0.096	0.159	0.096	0.040	0.145	0.052	0.099	0.097
Precipitation of driest month	0.152	0.060	0.080	0.030	0.031	0.044	0.111	0.005

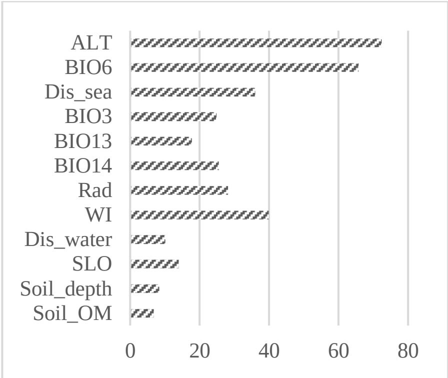
Relative importance of variables in stratified sampling method (Mixed forest)

	MAX-ENT	CTA	FDA	RF	GLM	GBM	GAM	ANN
Altitude	0.078	0.176	0.090	0.034	0.000	0.059	0.083	0.297
Slope	0.183	0.178	0.201	0.059	0.254	0.147	0.147	0.034
Radiation	0.124	0.115	0.095	0.039	0.056	0.065	0.079	0.047
Distance from water	0.071	0.093	0.006	0.033	0.000	0.047	0.024	0.284
Distance from sea	0.102	0.049	0.034	0.024	0.062	0.027	0.052	0.607
Soil Depth	0.197	0.354	0.380	0.055	0.498	0.115	0.834	0.301
Soil Organic Matter Content in Layer A	0.023	0.090	0.017	0.029	0.037	0.054	0.355	0.001
Warmth Index	0.069	0.059	0.023	0.014	0.000	0.011	0.167	0.050
Isothemality	0.080	0.059	0.016	0.023	0.000	0.022	0.012	0.002
Min Temperature of Coldest Month	0.040	0.025	0.046	0.016	0.007	0.026	0.095	0.003
Precipitation of wettest month	0.079	0.034	0.021	0.018	0.013	0.016	0.021	0.109
Precipitation of driest month	0.051	0.074	0.013	0.023	0.055	0.031	0.046	0.010

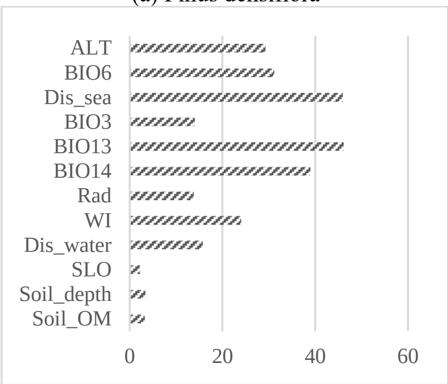
Appendix 3. Importance of Environmental Variables by Major Forest Species in Random Forest Model



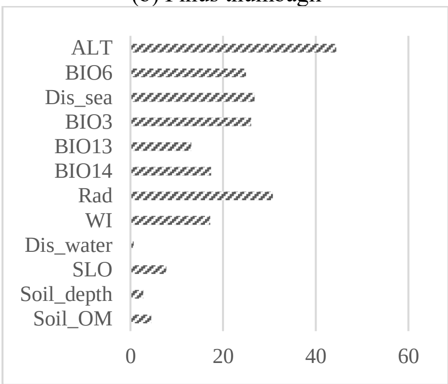
(a) Pinus densiflora



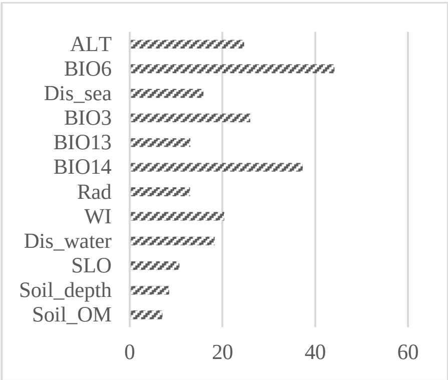
(b) Pinus thumbagii



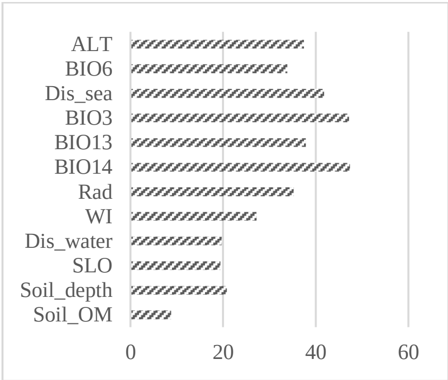
(c) Quercus app.



(d) Deciduous broadleaved forest

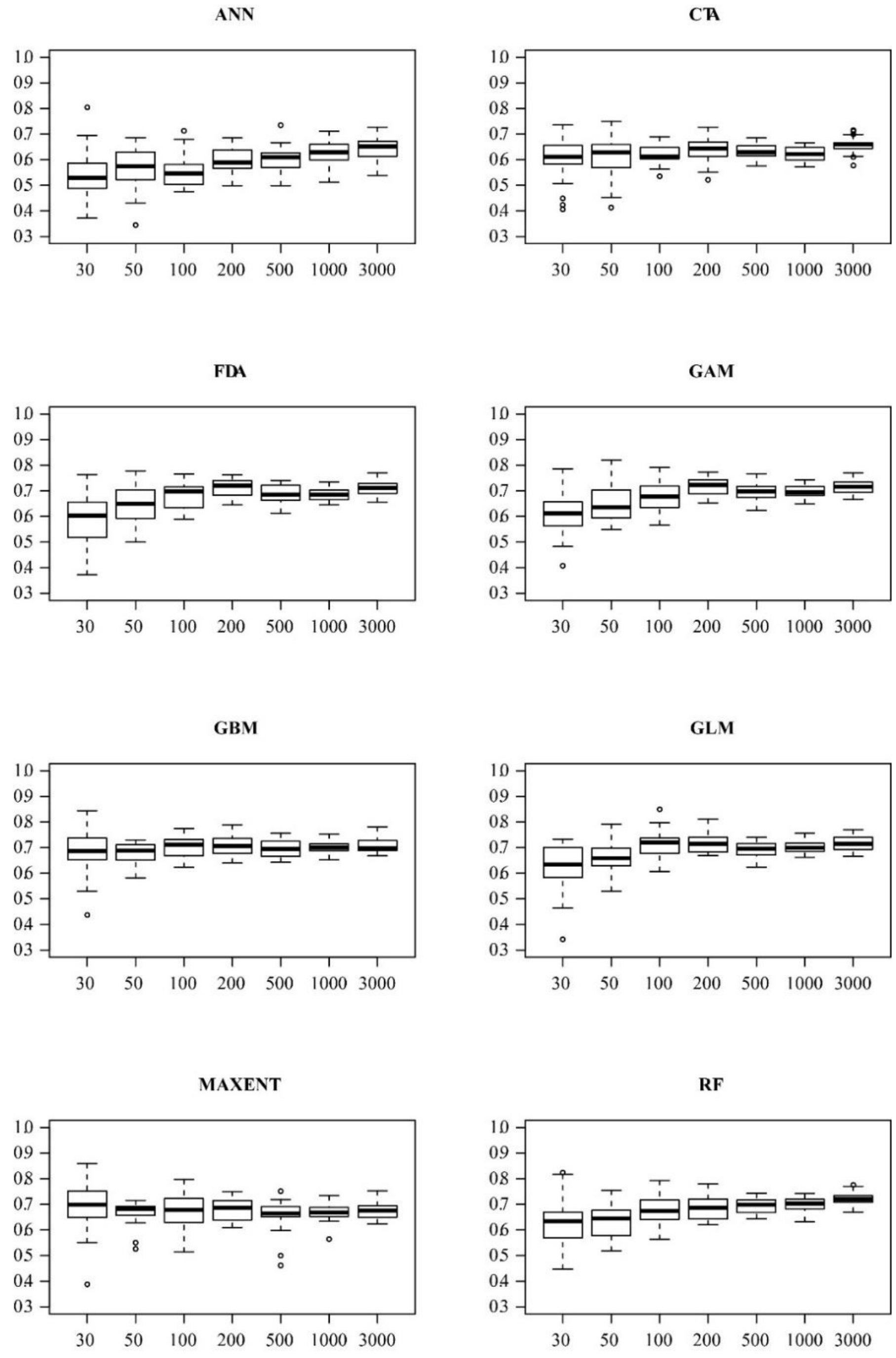


(e) Evergreen broadleaved forest

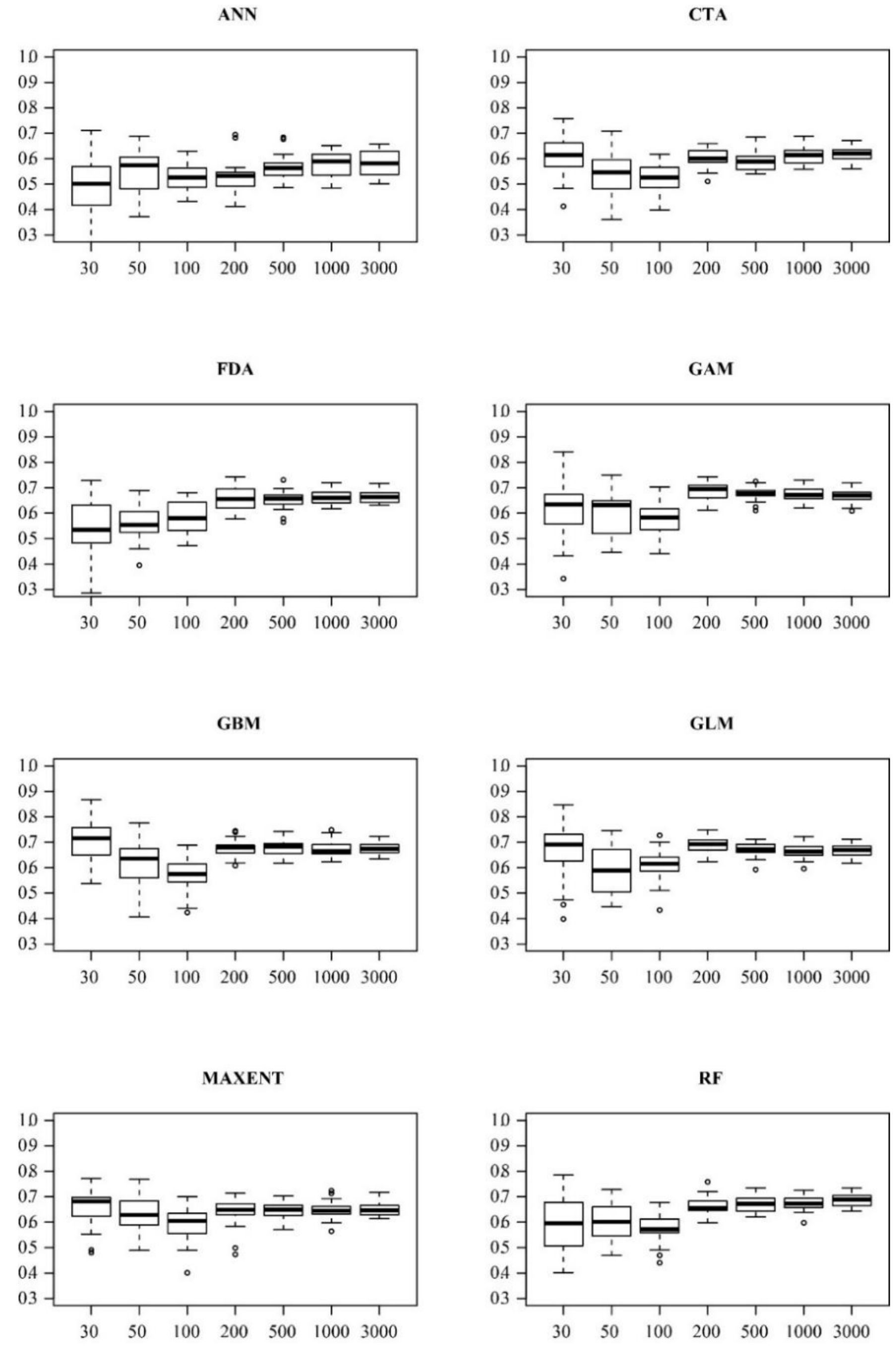


(f) Mixed forest

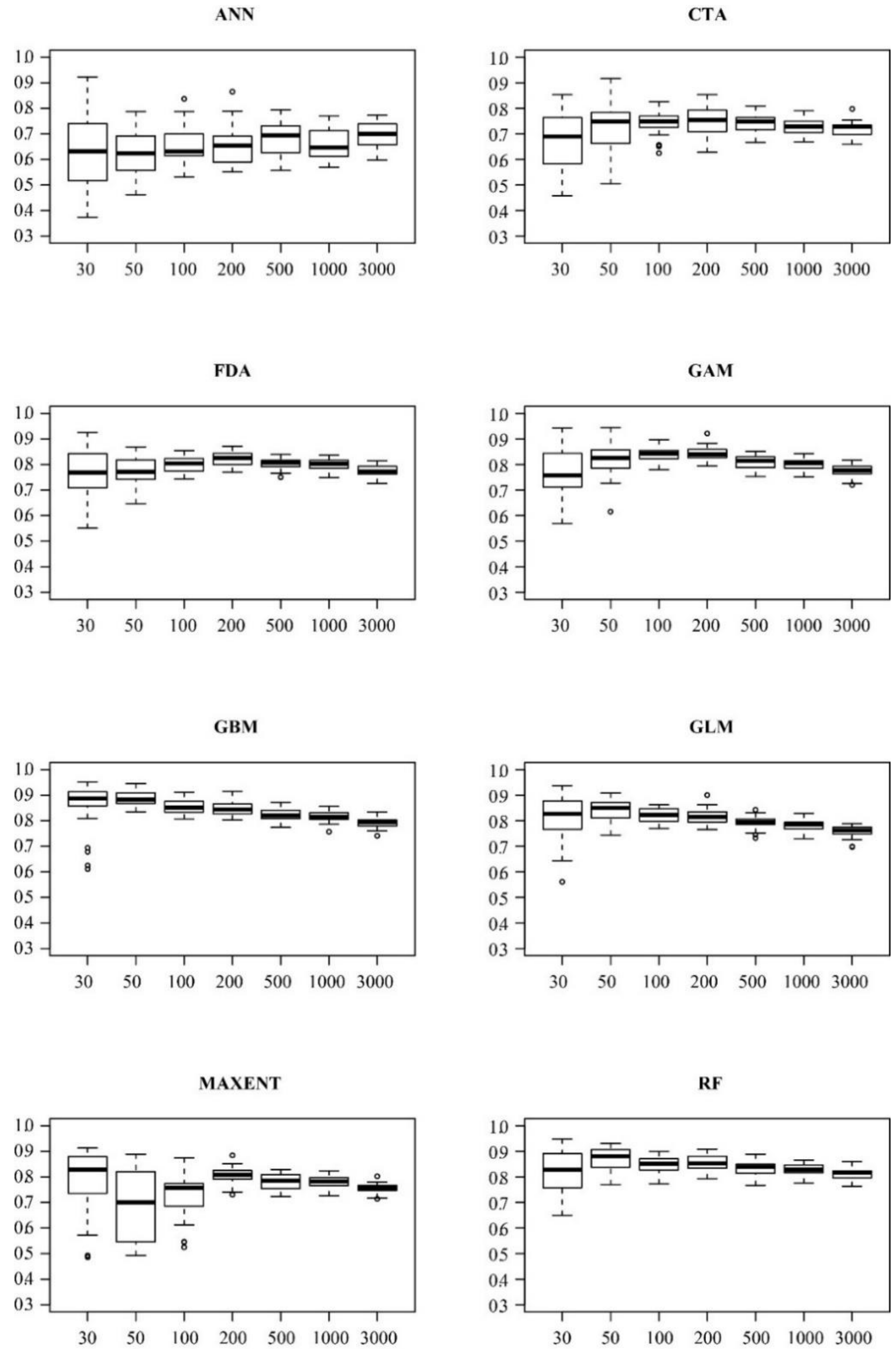
Appendix 4. The performance of SDMs by sampling size (random sampling; x-axis: AUC, y-axis: sample size)



Appendix 5. The performance of SDMs by sampling size (stratified sampling; x-axis: AUC, y-axis: sample size)



Appendix 6. The performance of SDMs by sampling size (area-weighted sampling; x-axis: AUC, y-axis: sample size)



국문 초록

기후변화를 고려한 산림 수종 분포변화 예측의 불확실성 평가

성 선 용

서울대학교 대학원
협동과정 조경학
지도교수: 이 동 근

기후변화에 따른 산림분야의 부정적인 영향이 증대할 것으로 예상되고 있으며 이를 저감하기 위한 다양한 대책들이 제시되고 있다. 제한된 시간과 자원을 가지고 기후변화에 의한 부정적인 영향을 저감하고 효과적으로 대응하기 위해서는 무엇보다도 기후변화에 따른 정확한 영향평가가 필요하다. 기후변화에 정확한 영향평가를 위해서는 영향평가에서 수반될 수 있는 불확실성에 대해서 이해하고 이를 반드시 정량화 하는 것이 필요하다.

IPCC 제4차 보고서에서부터 언급되기 시작한 불확실성에 대한 개념은 제5차 영향평가 보고서에서도 명기되어 있으며 의사결정에 도움을 주는 개념으로서 사용되고 있다. 우리나라에서도 기후변화의 영

향을 평가함에 있어서 불확실성을 고려하기 위하여 이를 정량화 하기 위한 노력을 기울이고 있다. 그러나 이들 연구는 아직 초기단계로서 제한적인 범위 내에서 진행되고 있으며 다양한 측면에서의 불확실성을 고려하지 못하고 있다.

따라서, 본 연구에서는 기후변화에 따라서 발생할 수 있는 불확실성의 원인을 검증하기 위하여 산림의 표본 추출 방법에 대한 효과성을 검증하고 1) 표본 추출 방법 및 표본 개수에 따른 불확실성 2) 모형 알고리즘 종류에 따른 통계적 유의성의 차이와 공간분포의 불확실성 3) 단일 종분포모형과 통계 다중 종 분포모형에서의 주요 수종간 경쟁에 따른 불확실성으로 나누어서 분석을 진행하였으며, 4) RCP시나리오의 기온과 강수량 변화에 따른 종 분포 범위의 변화를 살펴보았다.

표본 추출 방법에 따른 효과 분석을 위하여 3가지의 샘플링 방법과 7개의 샘플 개수를 활용하여 모수와 근접한지 one-way t-test를 활용하여 분석하였다. 그 결과 층화-무작위추출 표본 추출 방법이 가장 잘 모수를 재현하는 것으로 나타났으며, 표본 개수는 200개 이상이 된다면 종 분포모형을 활용하여 산림의 잠재적 분포를 예측하는데 있어 정확도에 유의미한 변화가 없는 것으로 나타났다.

모형의 알고리즘에 따른 불확실성을 분석하기 위하여 널리 사용되고 있는 종 분포 모형 중 통계기반모형과 기계학습기반모형 8개를

활용하여 현재의 식생분포를 모의한 결과, 모형의 알고리즘에 따라서 정확도에 차이가 있는 것으로 나타났으며, 대체적으로 기계학습모형 중 RF모형의 정확도가 높게 나타났으며, GLM 및 GAM과 같은 통계기반의 모형은 경우 양호한 정확도를 나타냈다. 공간적 불확실성을 평가하기 위하여 현재의 산림면적과 면적이 가장 유사하도록 모형의 확률 임계치를 조정하여 공간적 분포를 비교한 결과 모형에 따른 차이가 크게 나타났으며 이를 해결하기 위하여 앙상블 모형과 같은 공간적 불확실성을 고려할 수 있는 방법이 필요성을 확인할 수 있었다.

단일종을 고려한 종분포모형과 달리 다중 종 분포모형에서는 단일 수종에서 확인할 수 없었던 여러 수종의 적합도를 평가 Random Forest 알고리즘과 GAEZ 분류를 활용하여 수행하였다. 그 결과 단일 수종만 고려한 경우보다 다중 종을 고려한 경우의 RCP 시나리오의 범위에 따라 불확실성이 더 광범위하게 나타났다. 하지만 단일종만을 고려한 종 분포 모형에서는 불확실성은 낮지만 기후대에 변화에 따른 수종을 고려하고 있지 못하기 때문에, 향후 기후변화의 범위에 따른 다른 기후대의 수종도입을 고려할 수 있는 방안에 대한 추가적인 고찰이 필요하다.

본 연구를 통하여 향후 기후변화에 따른 산림의 관리전략을 수립함에 있어서 산림의 수종변화를 예측할 때 단계별로 발생할 수 있는 불확실성의 원인을 분석하고 이를 고려한 환경계획을 세우는데 기여

할 수 있을 것으로 판단된다. 이는 불확실성의 큰 요소의 반영 우선순위와 불확실성을 줄이고 효과적인 산림의 관리를 위한 모니터링 및 조사시점을 선정하는데 활용될 것으로 기대할 수 있다.

주요어: 불확실성, 수종분포변화, 산림 표본추출 방법, BIOMOD2, Random Forest, **산림관리**

학번: 2013-30714