

## 기후변화에 따른 법정보호종 분포 예측을 위한 종분포모델 적용 방법 검토\* - *Rodgersia podophylla*를 중심으로 -

유영재<sup>1)</sup> · 황진후<sup>1)</sup> · 전성우<sup>2)</sup>

<sup>1)</sup>고려대학교 오정리질리언스연구원 연구교수 · <sup>2)</sup>고려대학교 환경생태공학부 교수

## A Study on the Application of Modeling to predict the Distribution of Legally Protected Species Under Climate Change - A Case Study of *Rodgersia podophylla* -

Yoo, Youngjae<sup>1)</sup> · Hwang, Jinhoo<sup>1)</sup> and Jeon, Seong-woo<sup>2)</sup>

<sup>1)</sup>Ojeong Resilience Institute, Korea University, Research Professor,

<sup>2)</sup>Division of Environmental Science & Ecological Engineering, Korea University, Professor.

### ABSTRACT

Legally protected species are one of the crucial considerations in the field of natural ecology when conducting environmental impact assessments (EIAs). The occurrence of legally protected species, especially 'Endangered Wildlife' designated by Ministry of Environment, significantly influences the progression of projects subject to EIA, necessitating clear investigations and presentations of their habitats. In perspective of statistics, a minimum of 30 occurrence coordinates is required for population prediction, but most of endangered wildlife has insufficient coordinates and it posing challenges for distribution prediction through modeling. Consequently, this study aims to propose modeling methodologies applicable when coordinate data are limited, focusing on *Rodgersia podophylla*, representing characteristics of endangered wildlife and northern plant species. For this methodology, 30 random sampling coordinates were used as input data, assuming little survey data, and modeling was performed using individual models included in BIOMOD2. After that, the modeling results were evaluated by using discrimination capacity and the reality reflection ability. An optimal modeling technique was proposed

\* 본 연구는 환경부의 재원으로 한국환경산업기술원의 ICT기반 환경영향평가 의사결정 지원 기술개발사업의 지원을 받아 연구되었습니다.(2020002990009)

**First author** : Yoo, Youngjae, Ojeong Resilience Institute, Korea University, Research Professor,  
Tel : +82-2-3290-3543, E-mail : lemonesty@korea.ac.kr

**Corresponding author** : Jeon, Seongwoo, Division of Environmental Science & Ecological Engineering, Korea University, Professor,  
Tel : +82-2-3290-3043, E-mail : eepps\_korea@korea.ac.kr

**Received** : 27 March, 2024. **Revised** : 5 June, 2024. **Accepted** : 3 June, 2024

by ensemble the remaining models except for the MaxEnt model, which was found to be less reliable in the modeling results. Alongside discussions on discrimination capacity metrics(e.g. TSS and AUC) presented in modeling results, this study provides insights and suggestions for improvement, but it has limitations that it is difficult to use universally because it is not a study conducted on various species. By supporting survey site selection in EIA processes, this research is anticipated to contribute to minimizing situations where protected species are overlooked in survey results.

**Key Words:** *Species Distribution Model, Climate Change, Environmental Impact Assessment, Discrimination capacity, Reliability Assessment*

## I. 서론

법정보호종은 환경영향평가시 자연생태환경 분야의 동·식물상 항목에서 중요하게 다루어 진다(Korea Environment Institute, 2006). 환경영향평가 사업대상지에서 법정보호종 서식이 확인될 경우 우선적으로 사업지구 제척 또는 원형보전 방안을 검토하고, 제척 또는 원형보전이 불가능할 경우 대체서식지를 조성하도록 되어 있다(Ministry of Environment, 2023). 발견되는 법정보호종이 동물일 경우 포획·이주를 통해 사업을 진행하는 경우가 종종 있지만, 식물일 경우 설계 변경을 하거나 사업을 진행할 수 없게 되는 경우가 대부분이다(Ministry of Environment, 2011).

법정보호종은 환경영향평가 대상사업의 진행 여부에 영향을 미치기 때문에 서식여부를 명확하게 조사하고 제시할 필요성이 있다. 「환경영향평가서등 작성 등에 관한 규정」 제33조(환경영향평가서의 구성 및 작성 방법)에 따른 [별표 6]에서는 동·식물상 항목의 조사에 대해 현지조사, 문헌조사 및 탐문조사로 구분하여 제시하고 있다. 표본점 조사 방식으로 이루어지는 현지조사의 경우 조사당시 표본점에서 서식이 확인되지 않으면 결과에서 누락될 수 있으며, 문헌조사는 과거 현지조사시 발견되지 않았을 경우 법정보호종이 누락될 수 있고, 탐문조사의 경우 탐문조사대상자의 생물종에 대한 지식수준에 따라

법정보호종이 누락될 수 있다. 이에 따라 법정보호종의 서식지를 공간적으로 예측하고 제공함으로써 효과적으로 관리하기 위한 방법을 제시할 필요성이 있지만, 분포 예상 지역을 포인트로 제공한 일부 사례(Wonju Regional Environment Office, 2014)를 제외하면 공간적인 분포를 제시한 사례는 확인할 수 없었다.

통계학적인 관점에서 모집단 예측을 위해 최소 30개 이상의 데이터가 필요하며(Abramovich & Bayvel, 1997), 종분포모델의 관점에서는 최소 50개 이상의 포인트 사용을 권장하고 있다(Stockwell & Peterson, 2002, Hernandez et al., 2006). 서식적합지의 공간적 예측을 제한하는 대표적인 요소는 현장조사 자료(예: 출현좌표)의 부족이다(Do et al., 2017). 종분포모델을 구동하고 결과를 제시하기 위해서는 앞서 말한 것과 같이 일정 수 이상의 현장 조사 자료가 필요하지만, 국내에서 시행되고 있는 대규모 조사인 전국자연환경조사(NES; National ecosystem survey)조차도 법정보호종에 대한 조사자료는 많지 않은 실정이다. 현재까지 구축이 완료된 제2차 전국자연환경조사부터 제4차 전국자연환경조사 자료를 종합한 결과 총 92종(I 급 13종, II 급 79종)의 멸종위기 육상식물 중 1개 이상의 좌표가 확인된 것은 59종(I 급 4종, II 급 55종)이었으며, 30개 이상의 좌표가 확인된 종은 솔붓꽃과 산뽕나무 단 2종으로 각 32개의 좌표가 확인되었으며 모두 멸종위기야생생물 II 급에 해당된

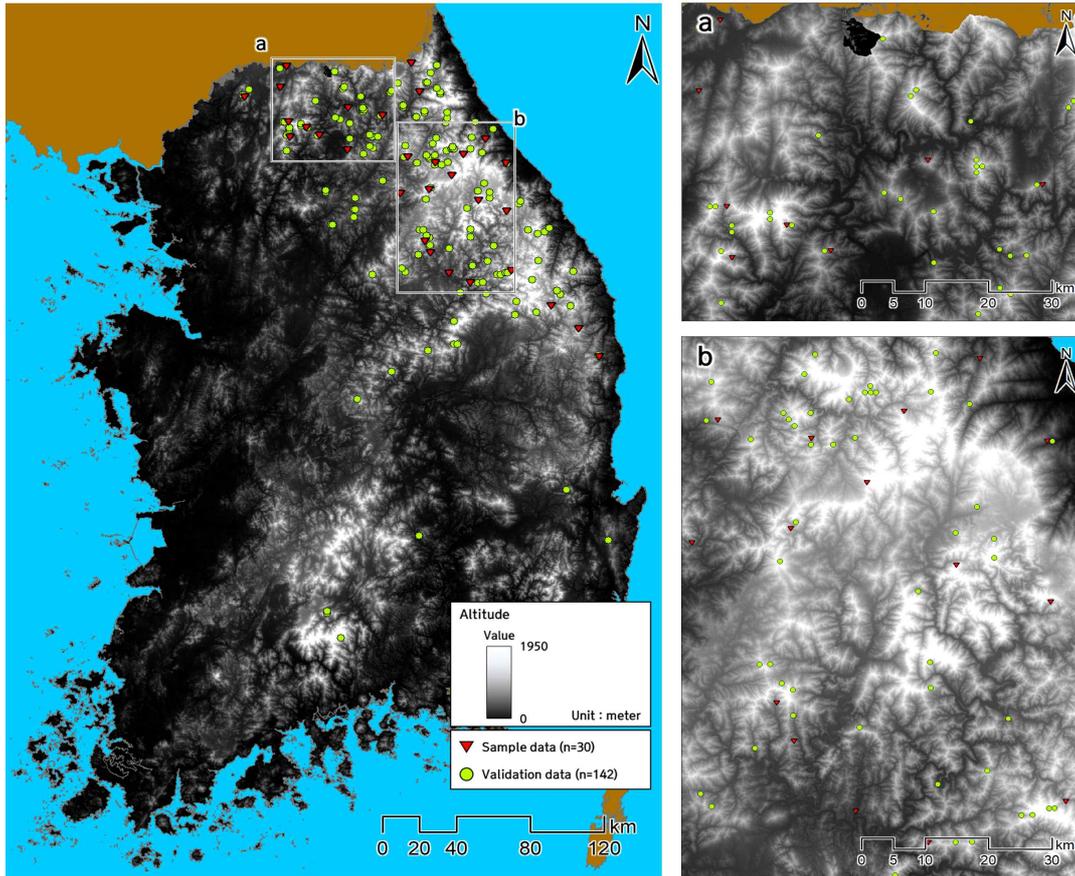


Figure 1. Study area with purified presence data

다. 즉, 멸종위기 육상식물 중 종분포모델의 구동을 시도할 수 있는 종은 단 2종 뿐이며, 통계적으로는 유의할 수 있으나 이마저도 종분포모델의 관점에서는 권장되지 않는다고 할 수 있다. 하지만, 그럼에도 불구하고 종분포모델의 구동을 통해 멸종위기 야생생물의 서식지를 예측하고 보전하기 위한 노력을 기울일 필요성은 존재한다.

전국자연환경조사에서 조사된 멸종위기 육상식물은 앞서 확인된 바와 같이 종분포모델을 구동하기에 충분한 좌표가 조사되어 있지 않을 뿐만 아니라 기후변화에 따른 생물자원 관리를 위해 제시하고 있는 모델링 기법이 부족한 실정이다. 본 연구는 기후변화로 인해 서식지변화를 크

게 겪을 것으로 예상되는 종을 대상으로 하여, 제공되는 종의 좌표 수가 적을 때에도 분포를 정확하게 예측하기 위한 모델링 방법론과 함께, 모델링시 주의해야 할 사항을 제시하는 데 목적이 있다. 즉, 앞서 언급한 바와 같이 종분포 모델의 관점에서는 50개 이상의 포인트가 권장되나, 현실점에서 50개 이상의 좌표를 나타내는 범정보호종이 없는 점을 고려하였을 때, 통계적으로 권장되는 30개를 기준으로 모델링 기법을 개발하고자 하였다. 이를 통해 좌표의 수가 제한적일 때에도 가장 현실을 잘 반영하는 모델링 방법을 제시하고자 하였으며, 동시에 모델링 결과의 신뢰성으로 주로 제시되는 통계적 정확도에 대해 고찰하고자 하였다.

## II. 재료 및 방법

### 1. 연구대상지

대한민국은 유라시아 대륙의 동쪽 끝에 돌출된 반도로, 동쪽은 높고 서쪽은 낮은 지형이며 북쪽으로는 북한과 국경을 접하고 있고 나머지 3면은 바다로 둘러싸여 있다(Figure 1). 봄, 여름, 가을 겨울의 사계절이 나타나며 온화한 기후를 보이는데, 여름은 덥고 습하며 겨울은 춥고 건조하고 봄과 가을은 비교적 짧은 편이다. 본 연구는 대한민국 중에서도 내륙을 대상으로 하였다.

### 2. 연구대상종

본 연구는 기후변화에 의한 서식환경 변화의 영향을 많이 받을 것으로 예상되는 육상식물을 대상으로 하였다. 북방계 식물은 대체로 기온이 낮고 고도가 높은 지역에 서식하는 등 특정한 생태적 지위를 가진다. 이것은 멸종위기 야생생물의 특성과 비슷하다. 멸종위기 야생생물은 대체로 환경적 및 지리적으로 제한적인 분포를 나타내기 때문에 개체수가 적고 조사로써 발견하기가 쉽지 않다.

연구대상종 선정 과정은 다음과 같다. ‘한반도 기후변화 적응 대상식물 300종’(Korea National Arboretum, 2010) 중 기후변화로 인해 서식지가 축소될 것으로 예상되는 북방계 식물 100종을 대상으로 제2차, 제3차 및 제4차 전국자연환경조사 자료를 검토하여 좌표 100개 이상인 11개 종을 선별하였다. 이후, 특정 환경을 선호하는 멸종위기 야생생물의 특성을 가진 1종을 연구 대상으로 선정하였다. 최종적으로 선정된 대상종은 도깨비부채(*Rodgersia podophylla*)로써 1993년 「환경보전법」에 의해 특정야생동·식물(現 「야생생물 보호 및 관리에 관한 법률」 멸종위기 야생생물)로 지정되었다가 개체수 증가로 1998년 해제된 이력이 있다. 현재는 「수목원·정원의 조성 및 진흥에 관한 법률」의 시행규칙

에 의해 희귀식물로 지정된 상태이다. 즉, 해당 종은 과거에 멸종위기종으로써 취급되었으며 현재는 희귀식물로 지정되어 관리되며, 전국자연환경조사 상 출현좌표를 충분히 확보할 수 있음이 확인된 종이다. 도깨비부채의 좌표는 제2차 조사에서 75개, 제3차 조사에서 62개, 제4차 조사에서 51개로 총 188개가 확인되었다(Table 1).

**Table 1.** Number of presence data surveyed for each period of the nationwide natural environment survey on the target species

| Korean name | Scientific name             | Number of coordinates |                     |                     |
|-------------|-----------------------------|-----------------------|---------------------|---------------------|
|             |                             | 2 <sup>nd</sup> NES   | 3 <sup>rd</sup> NES | 4 <sup>th</sup> NES |
| 도깨비부채       | <i>Rodgersia podophylla</i> | 75                    | 62                  | 51                  |

### 3. 연구 방법론

#### 1) 종 분포 자료

전국자연환경조사는 자연환경보전법에 의해 1986년부터 시행된 국가적 차원의 조사이다. 자료의 GIS(Geographic information system) 데이터 베이스가 구축된 제2차 전국자연환경조사 자료부터 2018년 종료된 제4차 전국자연환경 조사 자료를 종합하여 비교하였다. 종분포 모델의 사용시 획득한 좌표를 그대로 사용할 경우 공간자기상관(Spatial autocorrelation)과 함께 비정상성(Non-stationarity)이 발생한다(Miller, 2012). 공간자기상관을 최소화 하기 위해 데이터 정제 과정이 필요하다는 것은 오래 전부터 언급되어 왔다(Haining, 1990). 공간 샘플링 방법은 크게 임의 추출(Random sampling), 층화 추출(Stratified sampling), 계통 추출(Systematic sampling)로 구분되는데, 전국자연환경조사 자료의 경우 전국을 대상으로 하여 임의추출된 자료임을 가정하였을 때, 공간적 거리가 가까운 경우 가중치가 부여됨에 따라 공간자기상관이 발생할 수 있다.

본 연구에서는 환경변수가 가지는 격자단위에 맞추어 각 격자당 하나의 좌표만 남기는 계통 추출 방법을 사용하여 공간자기상관을 최소화 하

**Table 2.** Environmental variables used for the prediction of *R. podophylla*.

| Category    | Variable name             | Raw data   |
|-------------|---------------------------|--|
| Climatic    | Bio1                      | Korea meteorological administration<br>- Max temperature<br>- Min temperature<br>- Precipitation |
|             | Bio2                      |  |
|             | Bio4                      |  |
|             | Bio12                     |  |
|             | Bio13                     |  |
|             | Bio14                     |  |
| Topographic | Aspect                    | WorldClim2.1<br>- Digital Elevation Model (DEM)  |
|             | Distance from forest edge | Ministry of environment<br>- Land cover map  |
|             | Distance from valley      | Water resources management information system<br>- Stream order map                              |
| Pedological | Bed rock                  | Korea forest service<br>- Forest soil map  |
|             | Soil texture              |  |
|             | Soil humidity             |  |
|             | Soil organic matter       |  |
| Forest      | Forest type               | Korea forest service<br>- Forest type map  |
|             | Age class                 |  |
|             | Crown density class       |  |

고자 하였다(Fortin et al., 1990). 이에 따라 총 188개 좌표에 대해 데이터 정제를 수행하여 172개의 출현좌표를 획득하였다. 모델의 예측 성능에 대한 객관적인 추정치를 얻기 위해서는 모델에 사용되지 않은 독립적인 데이터를 사용여야 한다(Pearce & Ferrier, 2000). 독립적인 데이터를 위해 정제된 172개의 출현좌표 중 30개의 출현좌표를 임의추출하여 모델링에 활용하였으며, 나머지 142개의 출현좌표는 모델의 현실 반영 정도를 평가하기 위한 검증자료로 활용하였다(Figure 1).

## 2) 환경변수

환경변수는 총 16개로 지형변수, 토양변수, 산림변수와 기후변화의 영향을 반영하기 위한 기후변수로 구성된다(Table 2). 지형변수, 토양변수, 산림변수는 선행연구 검토를 통해 선정하였다(Park et al., 2011; Korea National Arboretum, 2015; Yoon et al., 2015; Chicago Botanic Garden, 2016). 지형변수는 향, 산림경계로부터의 거리, 계곡으로부터의 거리로, 토양변수는 기

반암, 토성, 토양 습도, 토양 유기물 함량으로, 산림변수는 임상, 경급, 수관밀도로 구성된다.

데이터는 가장 해상도가 낮은 기후변수(30 arc seconds)에 맞추어 통일된 기준으로 제작하였다. 지형변수 중 산림경계로부터의 거리와 계곡으로부터의 거리는 ArcMap 프로그램의 Euclidean distance 기능을 사용하여 원 데이터로부터 제작하였다. 기후요인에 해당되는 변수는 기상청의 SSP시나리오를 기반으로 한 2000-2019년 자료를 월평균한 후 R의 Biovars 기능을 이용하여 생성하였다. 생물기후변수(Bioclimatic variables)의 경우 최저기온, 최고기온, 강수량 3개의 기후자료로부터 파생되는 것이므로 변수간 상관성이 높아 공선성(collinearity)이 발생할 수 있다(Lee et al., 2015). 변수간 상관성이 높은 경우 특정 변수에 가중치가 부여되어 연구결과에 편향을 만들 수 있다. 공선성을 제거하기 위한 방법으로 피어슨 상관계수(Pearson correlation coefficient) 및 주성분분석(Principal component analysis)을 이용한 변수의 차원 축소 방법을 흔히 사용한다(Dormann, 2013; De Marco, 2018). 이 방법을 통해

최종적으로 Bio1 (Annual mean temperature), Bio2 (Mean diurnal range), Bio5 (Max temperature of warmest month), Bio12 (Annual precipitation), Bio13 (Precipitation of wettest month), Bio14 (Precipitation of driest month) 6개의 변수를 선정하였다. 위에서 언급되지 않은 나머지 변수는 shape 파일의 경우 ArcMap의 Polygon to raster 기능을 사용하여 제작하였으며, raster 파일의 경우 기후변수의 격자에 맞추기 위하여 ArcMap의 Resample 기능을 이용하여 제작하였다.

### 3) 모델링 방법

본 연구에서 제시하고자 하는 모델은 종의 수가 적을 때에도 비교적 정확한 예측이 가능한 모델이었다. SDM(Species distribution model)은 오랜 기간 사용되어 온 모델로 소수종의 예측에 관한 다수의 연구가 존재했는데, MaxEnt와 Random Forest가 단일모델로서는 출현좌표가 적은 종을 예측하는데 있어 정확도가 높은 것으로 나타났다(Williams et al., 2009; Buechling & Tobalske, 2011; Gogol-Prokurat, 2011; Qin et al., 2017; Gao et al., 2022). 종분포를 예측하기 위한 단일 모델들이 등장한 이후 모델링 결과를 통합하는 앙상블 기법이 제시되었다. 앙상블은 모델을 통합하여 단일모델에서 발생할 수 있는 불확실성을 감소시켜주기 때문에 더 일관성 있는 결과를 제시하므로 더 뛰어난 성능을 나타낸다(Araújo & New, 2007; Hao et al., 2020). 모델을 앙상블 할 때에는 평균 또는 가중평균을 이용함으로써 예측 성능을 향상시킬 수 있는데(Marmion et al., 2009), 본 연구에서는 평균을 이용하여 앙상블을 수행하였다.

BIOMOD2는 앙상블 모델링을 수행할 수 있게 해주는 가장 잘 알려진 도구로써 R에서 사용 가능하며, 미래 기후 예측 기능을 통해 미래 기후에 대한 예측을 생성할 수 있다. 본 연구에서는 종 분포를 추정하기 위해 BIOMOD2에서 제공하는 10가지 모델 중 8가지 모델(generalized

linear model [GLM], generalized boosting model [GBM], classification tree analysis [CTA], artificial neural network [ANN], flexible discriminant analysis [FDA], multiple adaptive regression splines [MARS], random forest [RF] and maximum entropy [MaxEnt])을 사용하였다. GAM은 자유도를 나타내는 k 값에 대한 기준이 모호하기 때문에 제외하였으며, SRE는 결과값이 0, 0.5, 1로 연속형 수치로 제공되지 않기 때문에 임계값 기반의 서식지 분류 방법을 적용할 수 없으므로 제외하였다. 각 단일 모델에 대하여 학습자료와 평가자료를 8:2로 5개의 그룹으로 분할하였으며, 각 그룹별 5회 반복 모델링을 수행하여 개별모델당 25회의 모델링을 수행하였다. 25회의 모델링 결과 중 TSS(True skill statistic) 값이 0.6 이상인 모델을 평균하여 단일 모델의 결과를 생성하였다. 앙상블 모델은 단일 모델의 통계적 정확도 비교와 함께 검증자료를 이용한 현실 반영 능력을 검토하여 제시하였다.

### 4) 모델 정확도 평가

정확도의 평가는 널리 이용되고 있는 방법인 TSS, AUC(Area under the curve)를 활용하였다(Chapman et al., 2019; Fernandes et al., 2019; Hao et al., 2020). TSS는 민감도와 특이도의 합이 최대가 될 때의 임계값을 기반으로 생성되며, 특정 임계값에 종속된 성능 평가가 필요할 때 Kappa 대신 사용한다. 값의 범위는 -1부터 1까지로 나타나며, 0은 랜덤으로 선택된 모델이나 다름없음을 나타내고, 1에 가까울수록 임계값에 한정하여 모델의 설명력이 높다고 판단할 수 있다(Allouche et al., 2006). ROC(Receiver operating characteristic)는 모든 임계값에 대한 모델 성능을 보여주는 그래프로써, 주로 AUC 값을 평가에 사용한다(Lobo et al., 2008; Peterson et al., 2008). AUC 값은 주로 0.5부터 1까지로 나타나는데, 0.5인 경우 모델이 종의 분포와 비분포를 구분할 수 있는 능력이 없음을 의미하고, 1에 가

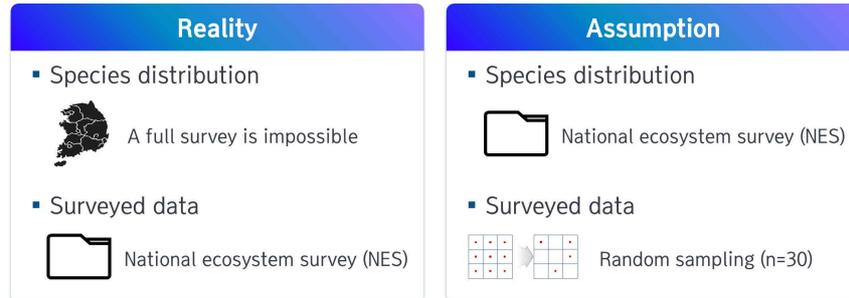


Figure 2. Assumptions defined for this study

까울수록 종의 분포와 비분포를 구분하는 능력이 뛰어나다는 것을 의미한다(Baldwin, 2009; Franklin et al., 2009).

AUC는 비출현자료 대신 임의비출현자료(Pseudo-absence)를 사용하는 경우 AUC theory에 위배되며, 부작용 및 작위 오류(Omission and commission errors)를 사용하는 Kappa와 같은 단일지표들은 동일한 문제를 나타낸다(Jiménez-Valverde, 2012). 본 연구는 비출현자료의 부재로 임의비출현자료를 사용하여 모델링을 수행하므로, TSS를 주요 평가지표로 설정하고 서식지와 비서식지를 구분하는 임계값 또한 TSS의 임계값(Cut-off value)을 사용하였다. 앞서 언급한 문제점이 있지만, AUC 또한 여전히 널리 쓰이는 평가지표 중 하나이므로, 함께 이용하여 연구결과가 전체적으로 설명력이 있는지를 확인하는데 보조적으로 사용하고자 하였다.

##### 5) 모델 현실 반영 능력 평가

SDM의 정확도를 측정하기 위해서 제시되는 방법은 주로 두 가지로, 신뢰도(Reliability)와 판별 능력(Discrimination capacity)가 제시된다(Pearce & Ferrier, 2000). 후자가 전자에 비해 더 중요하게 여겨지는 경향이 있으며(Ash & Shwartz, 1999), 검증을 위한 별도의 데이터가 없는 경우 이러한 평가방법이 가장 유용하게 사용될 수 있다. 현재 SDM 분야에서 진행되고 있는 대부분의 연구는 후자에 해당되는 판별 능력의 평가 지표 값을 맹신하는 경향이 있다. AUC와

TSS와 같은 평가 지표 값은 입력자료를 기반으로 모델을 평가함으로써 분류 성능을 보장하지만, 실제 분포를 현실적으로 잘 반영하는지에 대해서는 보장할 수 없다. 따라서, 본 연구에서는 얼마나 현실을 잘 반영하는지를 알아보기 위하여 모델에 활용하지 않은 142개의 검증자료를 이용하여, 모델링 결과로 도출되는 확률분포지도의 값을 추출하여 비교하였다.

전국자연환경조사 자료는 실제 분포에 대한 표본추출 데이터로 이해할 수 있다. 본 연구는 전국자연환경조사 자료의 종 출현 지점에 대한 자료를 실제 분포로 가정하고, 임의추출한 30개의 출현좌표를 표본추출데이터로 가정하였다(Figure 2). 전국자연환경조사 자료를 해당 종의 실제 분포로 가정하였기 때문에 임의추출한 표본추출데이터를 사용한 모델의 결과가 현실을 잘 반영했다면, 검증자료인 142개 출현좌표에 추출되는 확률값이 높아야 한다. 위 가정을 확장하면, 어떠한 모델이 단 30개의 자료를 기반으로 모집단을 잘 예측해 낸다면 더 많은 자료인 전국자연환경조사 자료를 사용해 모델링을 수행하였을 경우 실제 분포 또한 잘 예측할 수 있다고 할 수 있다.

### III. 결과 및 고찰

#### 1. 모델별 결과 비교

각 모델이 통계적으로 얼마나 서식지를 잘 예측해 내었는지를 알아보기 위해 판별 능력 측면에서 결과를 비교하고자 하였다. 각 모델을 25회

**Table 3.** Evaluation results obtained by single models

| Model  | Metric | Value | Cut-off | Sensitivity | Specificity |
|--------|--------|-------|---------|-------------|-------------|
| GLM    | TSS    | 0.863 | 568     | 100.000     | 86.094      |
|        | ROC    | 0.960 | 576     | 100.000     | 86.912      |
| GBM    | TSS    | 0.853 | 403     | 100.000     | 85.276      |
|        | ROC    | 0.969 | 402     | 100.000     | 85.276      |
| CTA    | TSS    | 0.816 | 776     | 90.000      | 91.616      |
|        | ROC    | 0.940 | 774     | 90.000      | 91.616      |
| ANN    | TSS    | 0.861 | 577     | 100.000     | 86.094      |
|        | ROC    | 0.965 | 580     | 100.000     | 86.503      |
| FDA    | TSS    | 0.784 | 424     | 93.333      | 84.867      |
|        | ROC    | 0.937 | 426     | 93.333      | 85.276      |
| MARS   | TSS    | 0.897 | 647     | 96.667      | 93.047      |
|        | ROC    | 0.982 | 644     | 96.667      | 93.047      |
| RF     | TSS    | 0.892 | 402     | 100.000     | 89.162      |
|        | ROC    | 0.985 | 400     | 100.000     | 89.162      |
| MaxEnt | TSS    | 0.793 | 391     | 86.667      | 92.653      |
|        | ROC    | 0.944 | 391     | 86.667      | 92.653      |

반복하였을 때 TSS 값이 0.6 이상인 모델의 수는 GLM 12개, GBM 24개, CTA 16개, ANN 19개, FDA 20개, MARS 15개, RF 25개, MaxEnt 13개로 나타났으며, 각 모델에서 TSS값이 0.6 이상인 모델을 이상불하여 대표모델을 생성하였다. TSS는 MARS, RF, GLM, ANN, GBM, CTA, MaxEnt, FDA 순으로 나타났고, AUC는 RF, MARS, GBM, ANN, GLM, MaxEnt, CTA, FDA 순으로 나타났다(Table 3). TSS의 값은 모든 모델에서 0.7 이상으로 나타나 출현 및 비출현 자료에 대한 정확도가 모두 높은 것으로 나타났고 AUC의 값 또한 모든 모델에서 0.9 이상으로 나타나, 30개라는 비교적 적은 출현좌표를 사용했음에도 불구하고 모델링 결과가 통계적으로 신뢰성있는 것으로 나타났다.

각 모델이 현실을 잘 반영하였는지를 확인하기 위해 142개의 검증자료를 이용하여 확률분포 지도의 예측값을 추출하고 관련 결과를 나타내었다(Table 4; Figure 3). 중분포모델의 결과인 확률분포지도의 해석과 관련하여 과거 많은 연구에서 확률값을 구간별로 나누어 사용하였는데, 50%를 뜻하는 0.5를 기준으로 서식 확률의 높고 낮음을 판단하였다(Song et al., 2012; Kim et al., 2016, Kim et al., 2022). 본 연구에서 활용한 BIOMOD2의 확률분포지도는 0부터 1000까지의 값으로 나타나므로, 500 이상의 값을 나타내어야 서식할 가능성이 높다고 할 수 있다. 앞선 가정에 따라 검증자료는 종의 실제 분포로 가정하였던 전국자연환경조사 데이터에서 모델링에 사용된 30개의 출현좌표만 제거된 것이므로, 현실을 잘

**Table 4.** Details of the probability distribution map values extracted as validation data

| Quantile                                     | GLM         | GBM         | CTA         | ANN         | FDA         | MARS       | RF         | MaxEnt     | Ens         |
|--|-------------|-------------|-------------|-------------|-------------|------------|------------|------------|-------------|
| Q1   | 544         | 486         | 590         | 453         | 487         | 381        | 399        | 294        | 529         |
| Median                                       | 694         | 643         | 770         | 697         | 638         | 601        | 527        | 489        | 707         |
| Q3   | 872         | 745         | 823         | 834         | 796         | 791        | 625        | 586        | 797         |
| Number of points ( $\geq 500$ ) [percentage] | 111 [78.2%] | 104 [73.2%] | 117 [82.4%] | 100 [70.4%] | 103 [72.5%] | 88 [62.0%] | 81 [57.0%] | 64 [45.1%] | 108 [76.1%] |

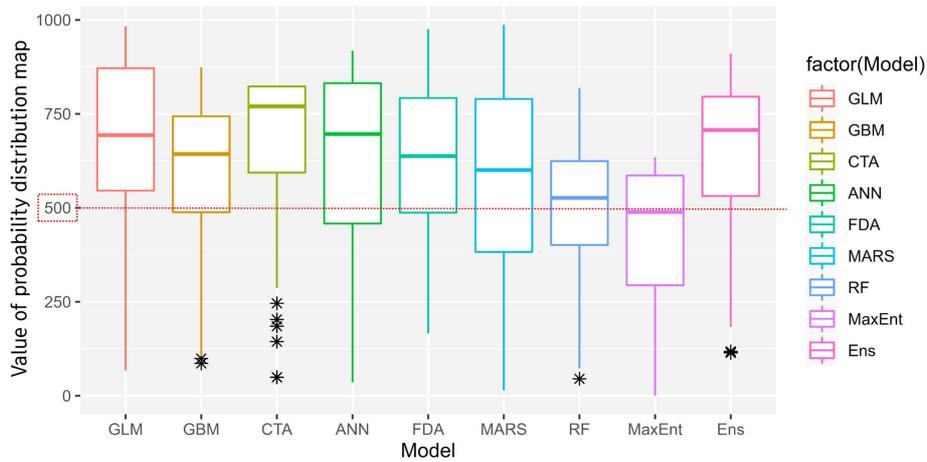


Figure 3. Boxplot showing how well reality was reflected by the models

반영하였다면 높은 값에 밀집된 분포를 나타내야 한다. CTA 모델의 경우 중앙값은 770으로 나타났으며, 검증자료로 추출한 값 중 500 이상의 값을 나타내는 지점은 전체 142개 지점 중 117지점(82.4%)으로 나타나 단일 모델 중 가장 실제와 가까운 예측을 한 것을 확인하였다. MaxEnt 모델의 경우 489로 가장 낮은 중앙값을 나타내었으며, 500 이상의 값을 나타내는 지점 또한 64지점(45.1%)으로 나타나 실제 분포를 예측하는 능력은 가장 떨어지는 것으로 확인되었다.

모든 모델이 0.9 이상의 AUC 값을 나타내어 통계적 정확도 측면에서 신뢰성이 있다고 나타난 것과 달리, 실제 분포 자료를 기반으로 모델별 확률분포지도 값을 추출하여 현실반영능력을 검토하였을 때에는 정확도가 적게는 45% 수준에서 높게는 82% 수준으로 비교적 낮은 것을 확인할 수 있었다. 특히, MaxEnt의 경우 TSS가 0.793, AUC가 0.944로 통계학적 측면에서 매우 높은 분류정확도를 보여주었지만, 검증자료를 기반으로 평가한 결과 중앙값은 489였으며 실제

분포를 예측한 정확도는 45.1%에 그쳐 가장 낮은 현실반영 능력을 나타내었다. MARS와 RF는 AUC 및 TSS에서 높은 분류정확도를 나타냈고, FDA는 수치적으로 가장 낮은 분류정확도를 나타냈지만, 현실반영 정도를 평가한 결과는 높은 분류 정확도를 나타낸 MARS나 RF 보다 FDA가 더 우수한 것을 확인하였다. 즉, 실제 분포를 잘 반영하는가를 나타내는 현실 반영 능력의 경우 통계적 정확도를 나타내는 지표값과 독립적임을 확인할 수 있었다.

판별 능력의 지표인 TSS 및 AUC가 모든 단일 모델에서 높은 값으로 나타나 변별력이 없는 것으로 나타나, 현실 반영 정도를 고려하여 Median 값이 500 미만으로 나타난 MaxEnt를 제외하고 나머지 7개의 모델을 앙상블함으로써 기후변화에 따른 *R. podophylla*의 분포를 예측하기 위한 모델을 설정하였다. 앙상블 모델은 CTA를 제외한 모델들보다 더 현실을 잘 반영하는 것으로 나타났으며, TSS 및 AUC의 값 또한 모두 통계적으로 높은 분류정확도를 나타내었다(Table 5). 또한

Table 5. Evaluation results of ensemble model

| Model    | Metric | Value | Cut-off | Sensitivity | Specificity |
|----------|--------|-------|---------|-------------|-------------|
| Ensemble | TSS    | 0.862 | 517     | 96.667      | 89.366      |
|          | ROC    | 0.976 | 519     | 96.667      | 89.571      |

검증자료를 활용한 실제 분포 예측시 76.1%의 정확도를 나타내어 현실반영 능력이 높은 것을 확인할 수 있었다.

각 개별모델과 앙상블모델이 나타내는 도깨비

부채의 분포를 확인해 본 결과 대부분의 모델에서 백두대간을 중심으로 고지대에 출현하는 것으로 나타났으나, ANN 모델의 경우 비교적 산발적인 서식지 분포를 나타내었다(Figure 4). 중

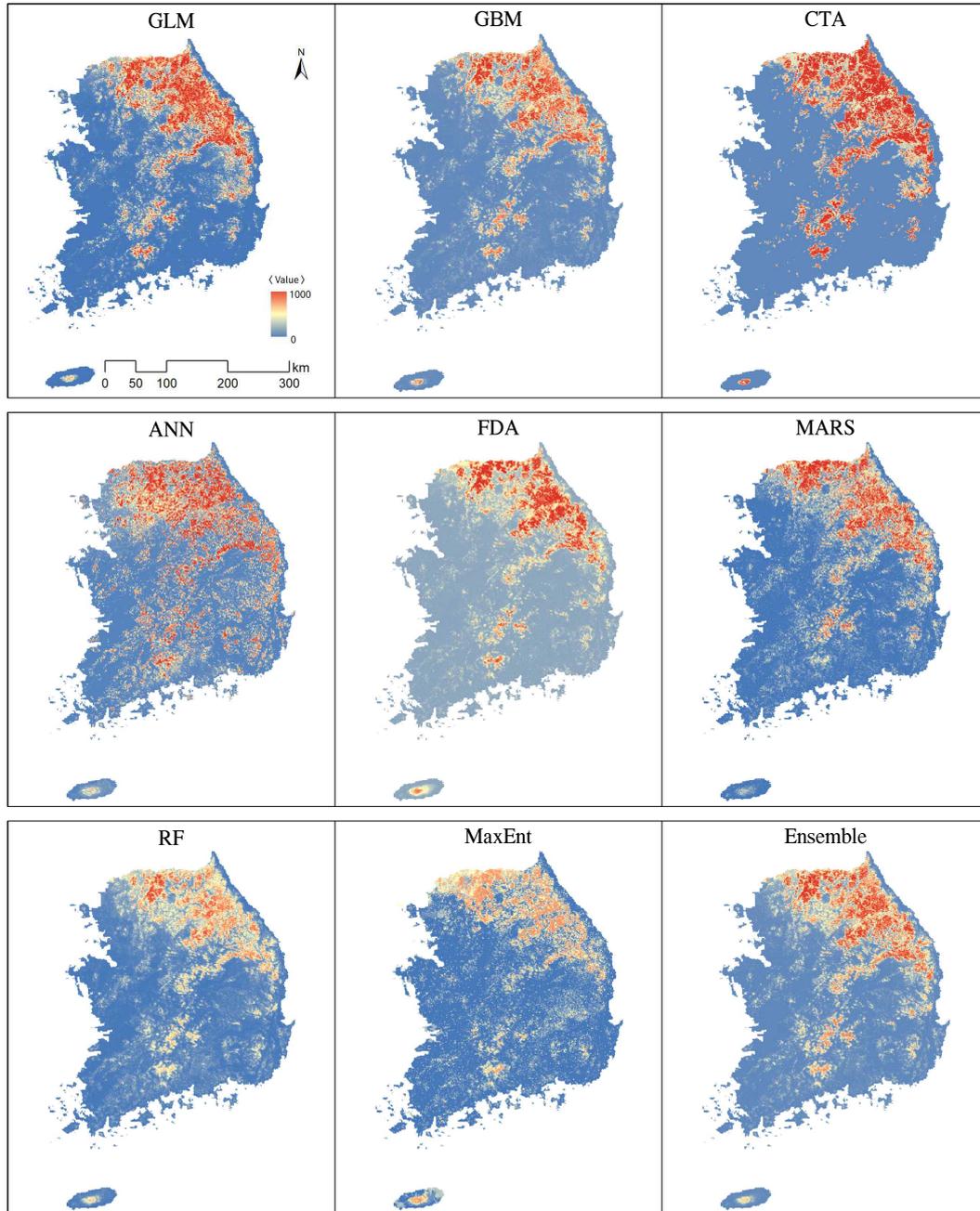


Figure 4. The probability distribution map of *R. podophylla* by model

분포모델의 경우 모델 내에서 판별 능력을 평가하는 AUC 및 TSS와 같은 지표의 값이 최대가 되는 민감도(Sensitivity)와 특이도(Specificity)의 교차점을 사용한다(Johnson et al., 2023; Jung et al., 2023). 일부 모델의 경우 서식적합여부를 구분하는 임계값인 Cutoff value를 낮추어 서식지를 과대추정함으로써, 통계적 정확도를 높이는 것으로 판단된다. 이는 모델링적 접근 방법을 사용할 때 종 분포의 예측이 과대추정될 수 있다는 선행연구와도 일치한다(Crimmins et al., 2013).

AUC 및 TSS의 값으로 결정되는 판별 능력 측면에서는 모든 모델이 통계적으로 예측 능력이 매우 뛰어난 것으로 나타나 이를 통한 모델의 선별은 어려웠다. 반면에, 검증자료를 사용한 현실반영 능력은 통계적 정확도를 나타내는 판별 능력의 지표값에 의존적이지 않았다. 따라서, 좌표 수가 제한적이고 좁은 환경범위를 갖는 종에 대한 모델링 결과 검증시 판별능력 지표만을 신뢰하는 것은 위험할 수 있다. 따라서 모델링에 사용하지 않은 독립된 데이터로 모델링 결과의 신뢰도에 대한 검증을 추가적으로 수행하는 것이 바람직할 것으로 판단된다.

#### IV. 결 론

본 연구는 법정보호종 및 북방계 식물에 대한 모델링 방법을 제시하고 추가적인 검증의 필요성을 제시하였다는 점에서 의의가 있다. 첫 번째로, 모델링 결과로써 정량적으로 제시되는 TSS, AUC 등 판별 능력을 나타내는 지표값의 높고 낮음이 현실을 잘 반영하는지의 여부를 결정하는 것은 아니므로, 독립된 데이터로 추가적인 검증이 필요하다는 점을 확인하였다. 두 번째로, BIOMOD2에서 제공하는 10개 모델 중 8개 모델 결과를 검증 자료와 비교 검토한 결과 MaxEnt 모델은 *R. podophylla*의 분포를 설명하기에 신뢰도가 떨어진다는 점을 확인하였다. 세 번째로, MaxEnt를 제외한 7개 모델을 이용해 앙상블 모

델을 만든 결과는 TSS, AUC 등 판별 능력 면에서 개선되었으나 CTA 단일모델에 비해 현실 반영 능력이 떨어지는 것으로 확인되었다. 하지만, 앙상블 모델의 경우 불확실성의 감소와 일관성 있는 결과를 제시함으로써 더 안정적인 결과를 제시할 것으로 예상된다. 다만, 데이터 취득에서 발생하는 한계로 하나의 데이터 세트를 분할하여 사용하였기 때문에, 더 정확한 검증을 위해서는 완전히 독립적으로 취득한 데이터를 사용할 필요성이 있다는 것이 본 연구의 한계점이다.

기후변화 시나리오를 사용하는 모델은 그 특성상 검증이 불가능하다. 기후변화 시나리오는 그 자체로 불확실성을 포함하고 있기 때문에, 이를 바탕으로 SDM을 수행하는 것은 더 큰 불확실성을 가져오는 것 처럼 보일 수 있다. 하지만 이러한 불확실성에도 불구하고 종의 보전을 위해 모델링을 수행하여 기후변화에 대비할 필요성이 있다. 기후변화는 생물종의 서식지 변화에 영향을 미치는 주요한 요인 중 하나이다. 북방계 식물의 경우 주로 서늘한 기후에 서식하기 때문에 기후변화로 인해 가장 큰 서식지 변화를 겪을 것으로 예상된다. 특히, 대한민국의 경우 3면이 바다로 둘러싸여 있고, 북쪽은 북한과 접해 있어 북상해가는 생물자원을 연속적으로 관리하기에 어려움이 있다. 따라서, 본 연구를 통해 기대하는 것은 크게 두 가지이다. 첫 번째는 좌표수가 제한적인 종에 대한 연구를 진행 할 때 모델링 결과인 TSS, AUC 등 판별능력지표 뿐만 아니라, 독립데이터 또는 현장검증을 통해 모델링 결과에 신뢰성을 더함으로써 향후 진행되는 연구 결과의 신뢰성 또한 향상 되기를 기대한다. 다른 하나는, 환경영향평가시 주요하게 다루어지는 법정보호종의 분포를 예측하고 조사지점을 제시하는 방법론으로 활용함으로써 법정보호종이 조사결과에서 누락되는 경우를 최소화 하는데 기여할 수 있을 것으로 기대된다.

## References

- Abramovich, F. & Bayvel, P. (1997). Some statistical remarks on the derivation of BER in amplified optical communication systems. *IEEE Transactions on Communications*, 45(9), 1032-1034. <https://doi.org/10.1109/26.623065>
- Allouche O.· Tsoar A. & Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araujo M. B. & New M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, 22(1), 42-47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Ash A. & Shwartz M. 1999. R2: a useful measure of model performance when predicting a dichotomous outcome. *Statistics in Medicine*, 18(4), 375-384. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990228\)18:4<375::AID-SIM20>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0258(19990228)18:4<375::AID-SIM20>3.0.CO;2-J)
- Baldwin R. A. 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11(4), 854-866. <https://www.mdpi.com/1099-4300/11/4/854>
- Buechling, A. & Tobalske, C. 2011. Predictive Habitat Modeling of Rare Plant Species in Pacific Northwest Forests. *Western Journal of Applied Forestry*, 26(2), 71-81. <https://doi.org/10.1093/wjaf/26.2.71>
- Chapman D.· Pescott O. L.· Roy H. E. & Tanner R. 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography*, 46(5), 1029-1040. <https://doi.org/10.1111/jbi.13555>
- Chicago Botanic Garden. 2016. A Comparative Trial Report on Rodgersias and Astilboides
- Crimmins S. M.· Dobrowski S. Z. & Mynsberge A. R. 2013. Evaluating ensemble forecasts of plant species distributions under climate change. *Ecological Modelling*, 266, 126-130. <https://doi.org/https://doi.org/10.1016/j.ecolmodel.2013.07.006>
- De Marco P. J. & Nobrega C. C. 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLOS ONE*, 13(9), e0202403. <https://doi.org/10.1371/journal.pone.0202403>
- Do MS· Lee JW· Jang HJ· Kim DI· Park JW & Yoo JC. 2017. Spatial Distribution Patterns and Prediction of Hotspot Area for Endangered Herpetofauna Species in Korea. *Korean Journal of Environment and Ecology*, 31(4), 381-396. <https://doi.org/10.13047/KJEE.2017.31.4.381>
- Dormann C. F.· Elith J.· Bacher S.· Buchmann C.· Carl G.· Carre G.· Marquez J. R. G.· Gruber B.· Lafourcade B.· Leitaó P. J.· Munkemüller T.· McClean C.· Osborne P. E.· Reineking B.· Schroder 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*, 36(1), 27-46. [/https://doi.org/10.1111/j.1600-0587.2012.07348.x](https://doi.org/10.1111/j.1600-0587.2012.07348.x)
- Fernandes R. F.· Scherrer D. & Guisan A. 2019. Effects of simulated observation errors on the performance of species distribution models. *Diversity and Distributions*, 25(3), 400-413. <https://doi.org/10.1111/ddi.12868>
- Fortin M.J.· Drapeau P. & Legendre P. 1990. Spatial autocorrelation and sampling design

- in plant ecology. In Progress in theoretical vegetation science. 209-222. Springer.
- Franklin J.· Wejnert K. E· Hathaway S. A· Rochester C. J. & Fisher R. N. 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. Diversity and Distributions, 15(1), 167-177. <https://doi.org/10.1111/j.1472-4642.2008.00536.x>
- Gao X.· Liu J. & Huang Z. 2022. The impact of climate change on the distribution of rare and endangered tree *Firmiana kwangsiensis* using the Maxent modeling. Ecology and Evolution, 12(8), e9165. <https://doi.org/10.1002/ece3.9165>
- Gogol-Prokurat M. 2011. Predicting habitat suitability for rare plants at local spatial scales using a species distribution model. Ecological Applications, 21(1), 33-47. <https://doi.org/10.1890/09-1190.1>
- Haining R. 1990. Spatial Data Analysis in the Social and Environmental Sciences. Cambridge University Press. <https://doi.org/10.1017/CB09780511623356>
- Hao T.· Elith J.· Lahoz-Monfort J. J. & Guillera-Aroita G. 2020. Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. Ecography, 43(4), 549-558. <https://doi.org/10.1111/ecog.04890>
- Hernandez P. A.· Graham C. H.· Master L. L. & Albert D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29(5), 773-785.
- Jimenez-Valverde A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography, 21(4), 498-507. <https://doi.org/10.1111/j.1466-8238.2011.00683.x>
- Johnson S.· Molano-Flores B. & Zaya D. 2023. Field validation as a tool for mitigating uncertainty in species distribution modeling for conservation planning. Conservation Science and Practice, 5(8), e12978. <https://doi.org/10.1111/csp2.12978>
- Jung JB· Park GE· Kim HJ· Huh JH & Um YR. 2023. Predicting the Habitat Suitability for *Angelica gigas* Medicinal Herb Using an Ensemble Species Distribution Model. Forests, 14(3), 592. <https://doi.org/10.3390/f14030592>
- Kim, CY· Kim, WM· Song, WK & Choi, JY. 2022. A study on the range of native seed habitat analysis using Seedzones and MaxEnt. Journal of the Korean Society of Environmental Restoration Technology, 25(1), 57 - 74. <https://doi.org/10.13087/KOSERT.2022.25.1.57>
- Kim, TG· Yang, DH· Cho, YH· Song, KH & Oh, JG. 2016. Habitat Distribution Change Prediction of Asiatic Black Bears (*Ursus thibetanus*) Using Maxent Modeling Approach. Korean Journal of Ecology and Environment. Korean Society of Limnology. <https://doi.org/10.11614/ksl.2016.49.3.197>
- Korea Environment Institute. 2006. A plan to introduce biodiversity components in environmental assessment
- Korea Forest Service & Korea National Arboretum. 2010. 300 Target Plats Adaptable to Climate Change in the Korean Peninsula. Korea forest service, Republic of Korea
- Korea National Arboretum. 2015. Guide for propagataion and cultivation of useful

- resource palnts
- Lee YH· Oh YJ· Hong SH· Na CS· Na YE· Kim CS & Sohn SI. 2015. Predicting the suitable habitat of invasive alien plant *Conyza bonariensis* based on climate change scenarios. *Journal of Climate Change Research*, 6(3), 243-248.
- Lobo J. M.· Jimenez-Valverde A. & Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2), 145-151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Marmion M.· Parviainen M.· Luoto M.· Heikkinen R. K. & Thuiller W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15(1), 59-69. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- Miller J. A. 2012. Species distribution models: Spatial autocorrelation and non-stationarity. *Progress in Physical Geography: Earth and Environment*, 36(5), 681-692. <https://doi.org/10.1177/0309133312442522>
- Ministry of Environment. 2011. A study on the preparation of guidelines for the creation of alternative habitats for major species in conducting environmental impact assessment
- Ministry of Environment. 2023. Guide to the Preparation, etc. of Environmental Impact Assessment Report, etc.
- Park JM· Seo HS & Lee JS. 2011. A Study on the Spatial and Environmental Characteristics of Forest Biology using GIS: A Case Study of Baekdudaegan area, Gyeongsangbuk-do and Chungcheongbuk-do. *Journal of forest and environmental science*, 27(3), 169-181.
- Pearce J. & Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133(3), 225-245. [https://doi.org/10.1016/S0304-3800\(00\)00322-7](https://doi.org/10.1016/S0304-3800(00)00322-7)
- Pearce J. & Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133(3), 225-245. [https://doi.org/10.1016/S0304-3800\(00\)00322-7](https://doi.org/10.1016/S0304-3800(00)00322-7)
- Peterson A. T.· Pape M. & Soberon J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, 213(1), 63-72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>
- Qin A.· Liu B.· Guo Q.· Bussmann R. W.· Ma F.· Jian Z.· Xu G. & Pei S. 2017. Maxent modeling for predicting impacts of climate change on the potential distribution of *Thuja sutchuenensis* Franch. an extremely endangered conifer from southwestern China. *Global Ecology and Conservation*, 10, 139-146. <https://doi.org/10.1016/j.gecco.2017.02.004>
- Song, WK & Kim, EY. 2012. A Comparison of Machine Learning Species Distribution Methods for Habitat Analysis of the Korea Water Deer (*Hydropotes inermis argyropus*). *Korean Journal of Remote Sensing*, 28(1), 171-180. <https://doi.org/10.7780/KJRS.2012.28.1.171>
- Stockwell D. R. B. & Peterson A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148(1), 1-13. [https://doi.org/10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X)
- Williams J. N.· Seo CW· Thorne J.· Nelson J. K.· Erwin S.· O'Brien J. M. & Schwartz M.

- W. 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15(4), 565-576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x>
- Wonju Regional Environment Office. 2014. Endangered wildlife (land plants) status data for verification of the natural ecological environment field of Environmental impact assessment report
- Yoon YS· Kim KA & Yoo KO. 2015. Environmental Characteristics of *Astilboides tabularis* (Hemsl.) Engl. Habitats. *Korean Journal of Plant Resources*, 28(1), 64-78.