고라니 서식지 분석을 위한 기계학습식 종분포모형 비교

송원경 · 김은영[†]

한국환경정책 · 평가연구원

A Comparison of Machine Learning Species Distribution Methods for Habitat Analysis of the Korea Water Deer (*Hydropotes inermis argyropus*)

Wonkyong Song and Eunyoung Kim[†]

Korea Environment Institute

Abstract : The field of wildlife habitat conservation research has attracted attention as integrated biodiversity management strategies. Considering the status of the species surveying data and the environmental variables in Korea, the GARP and Maxent models optimized for presence-only data could be one of the most suitable models in habitat modeling. For make sure applicability in the domestic environment we applied the machine learning species distribution model for analyzing habitats of the Korea water deer(*Hydropotes inermis argyropus*) in the *Sapgyocheon* watershed, *Chungcheong* province. We used the 3rd National Natural Environment Survey data and 10 environment variables by literature review for the modelling. Analysis results showed that habitats for the Korea water deer were predicted 16.3%(Maxent) and 27.1%(GARP), respectively. In terms of accuracy(training/test) the Maxent(0.85/0.69) was higher than the GARP(0.65/0.61), and the Spearman's rank correlation coefficient result of the Maxent($\rho=0.71$, $p<0.01$) was higher than the result of GARP($\rho=0.55$, $p<0.05$). However results could be depended on sites and target species, therefore selection of the appropriate model considering on the situation will be important to analyzing habitats.

Key Words : Ecological Niche Model, Maxent, GARP, Sapgyocheon watershed

요약 : 야생동 · 식물 서식지 보전 연구는 통합적인 생물다양성 관리 전략으로서 주목받고 있다. 국내 종 조사자료 및 환경공간정보 여건을 고려할 때 종출현정보에 최적화된 것으로 알려져 있는 GARP 모형과 Maxent 모형이 서식지 분석에 가장 적합한 것으로 판단된다. 국내 적용가능성을 확인하기 위해 충청도 삽교천 일원을 대상으로 고라니(*Hydropotes inermis argyropus*)에 대한 기계학습식 모형을 적용하였다. 종출현지점은 3차 전국자연환경조사, 환경변수는 문헌조사를 통해 10개를 도출하였다. 분석 결과 Maxent 모형과 GARP 모형은 각각 전체 면적의 16.3%, 27.1%를 고라니 서식지로 예측하였다. 종분포모형 정확도(훈련/검증)는 Maxent 모형(0.85/0.69)이 GARP 모형(0.65/0.61)보다 높게 분석되고 Spearman 순위 상관계수 역시 Maxent 모형($\rho=0.71$, $p<0.01$)이 GARP 모형($\rho=0.55$, $p<0.05$)보다 높게 분석되었다. 이는 대상지의 특성과 대상종에 따라 달라질 수 있으므로 상황에 따라 적절한 모형을 적용하는 것이 바람직할 것으로 판단된다.

접수일(2012년 1월 13일), 수정일(1차 : 2012년 2월 17일), 게재확정일(2012년 2월 18일).

[†] 교신저자: 김은영(eykim@kei.re.kr)

1. 서론

생물다양성이란 육상·해상 및 그 밖의 수중생태계와 이들 생태계가 부분을 이루는 복합생태계 등 모든 분야의 생물체간 변이성을 말하며 이는 종내 다양성, 종간의 다양성 및 생태계 다양성을 포함하는 개념이다(관계부처합동, 2009). 즉, 생물다양성은 생물종 다양성, 생물이 서식하는 생태계 다양성, 생물이 지닌 유전자 다양성을 총체적으로 지칭한다. 전 지구적으로 생물다양성에 대한 중요성이 크게 인식되기 시작하면서 생물다양성을 보전·관리하기 위한 다양한 노력들이 진행 중이다. 우리나라 역시 이러한 세계적 흐름에 편승하여 다양한 제도적 기반을 통해 종다양성 보전에 노력을 기울이고 있다. 특히 기후변화가 생태계 및 인간사회에 세계적인 위기로 인식됨에 따라 생물다양성협약 제4차 국가보고서를 발행하고(대한민국, 2009), 관계부처합동(2010)으로 국가 기후변화 적응대책(2011~2015) 수립을 통해 생태계 분야의 적응 전략을 마련하고 있다. 특히 생물다양성 보전의 핵심이라 할 수 있는 야생동식물 보호관리를 위해 환경부(2010)에서는 제2차 야생동·식물보호기본계획(2011~2015)을 수립하여 야생동식물 서식실태

조사 및 서식지 보호 관리를 주요 정책과제로 선정하고 있는 상황이다.

생물다양성 보전관리 정책은 체계적인 보호지역 지정 및 관리(생태계 차원), 멸종위기종 보전 및 복원(종 차원), 체계적인 유전자원 관리(유전자원 차원) 정책을 통해 추진되고 있다. 이 세 가지 정책은 상호 밀접한 관계를 갖고 추진되기 때문에 그 중요도를 비교하기 어렵지만 효과적인 보전지역 관리를 통해 종다양성 및 유전자 다양성을 보전할 수 있다는 관점에서 생태계 다양성 보전 전략이 보다 장기적으로 추진되어야 하는 정책으로 판단된다. 종 보전을 위한 서식지 관리는 생태계뿐만 아니라 종, 유전자 다양성을 함께 보전할 수 있는 효과적인 생물다양성 보전 전략이 될 수 있다. 이러한 맥락에서 이미 선진국에서는 1970년대 이후 자연환경 관리 정책에 생물종의 서식지 적합성 평가에 따른 보전지역 설정 및 복원을 위한 대상지 선정을 반영하고 있다(이동근과 송원경, 2008).

종분포모형은 종과 환경간의 관계를 경험적인 자료를 통해 비통계적 기법으로 분석하는 방법, 전문가 판단에 의한 분포 예측, 종 분포에 대한 공간적으로 명백한 통계적·경험적 모형 구축의 세 방향으로 진행되고

Fig. 1. The steps in species distribution modeling and predictive mapping (Frank, 2009, modified).

있다(Guisan and Thuiller, 2005; Degraer *et al.*, 2008). 최근에는 통계적 기법과 함께 다양한 기계학습식 종분포모형 방법론이 많이 활용되고 있으며, 이에 따라 종분포모형을 생태적 지위 모형과 서식지 적합성 모형으로 구분하기도 한다(Franklin, 2009). 생태적 지위 모형은 종의 분포와 다양한 환경변수와의 관계를 통해 가장 객관적인 결과를 도출하고 있어 최근 종분포 모형에 많이 활용되고 있다(Franklin, 2009)(Fig. 1). 이러한 종분포모형은 최근 통계적 모형(statistical model)과 기계학습식 모형(machine learning model)으로 발전되고 있는 추세이다. 통계적 모형은 GLM(Generalized Linear Model)과 GAM(Generalized Additive Model)을 비롯하여 MARS(Multivariate Adaptive Regression Splines) 등 다양한 통계 기반 방법론으로 발전되었다(권혁수, 2011). 기계학습식 모형은 의사결정나무(Decision tree), 인공신경망(ANN; Artificial Neural Network), 유전학적 알고리즘(genetic algorithm), 최대 엔트로피(maximum entropy) 등의 방법론을 기반으로 활용범위가 넓어지고 있다. 기계학습식 모형은 높은 분류정확도에도 불구하고 출현/비출현 자료의 활용 한계, 결과에 대한 설명이 어려운 단점이 있다고 제기되었으나(서창완 등, 2008b), 최대 엔트로피 모형으로 개발된 Maxent, 유전학적 알고리즘 모형으로 개발된 GARP(Genetic Algorithm for Rule-set Production) 등의 모형은 출현자료만을 활용할 경우에도 우수한 분류정확도를 도출할 수 있음을 보여주었다(Franklin, 2009).

우리나라는 보호대상 야생동물의 분포 및 서식밀도에 관한 기초 자료가 적고 이 분야의 GIS 적용 연구와 주요 야생동물의 서식지 분포 예측능력은 아직 부족한 상황이다(서창완 등, 2008b). 모형의 결과가 조사자료와 사용하는 모형의 방법에 따라 달라지므로 다양한 모형에 대한 검토와 적합한 모형의 선택이 필요하지만(서창완 등, 2008b), 국내에서는 아직 기계학습식 모형에 대한 검토가 부족한 상황이다. 현재까지 국내에서는 의사결정나무 방법론의 CART(Classification and Regression Tree), ANN 등을 이용한 설악산 산양 서식지 적합성 모형 연구(서창완 등, 2008b), Maxent를 이용한 생물종 분포모형 비교 연구(서창완 등, 2008a), Maxent 모형을 이용한 서식지 잠재력 평가 연구(이동

근과 김호걸, 2010), Maxent 모형을 이용한 종분포모형 적용(권혁수, 2011), Maxent 모형과 ANN을 이용한 종분포모형 연구(송원경, 2011) 등이 종분포모형과 관련된 기계학습식 모형 적용 연구이다.

본 연구에서는 최근 우리나라의 종조사 데이터 형태에 최적화된 기계학습식 종분포모형을 적용하여 서식지 분석 방법을 비교하고 향후 우리나라 환경에서의 종분포모형 적용 가능성을 확인하고자 한다.

2. 연구의 범위 및 방법

1) 연구대상지 및 대상종

본 연구의 대상지로는 산림지역과 농경지가 매트릭스 형태로 복잡하게 나타나는 충청도 서부지역의 충청남도 삽교천 유역 일원(면적: 약 1,668 km², 위도: 36° 54' 11"N ~ 36° 23' 17"N, 경도: 127° 12' 48"E ~ 126° 35' 50"E)으로 선정하였다(Fig. 2). 이 지역은 농경지와 산림지역, 도시지역이 복잡한 형태를 이루고 있으며 주요 야생 포유동물의 서식지인 산림지역이 패치 형태로 존재하여 다양한 환경변수를 고려한 야생동물 서식지 모형 적용에 적합하다.

종분포모형 비교를 위한 대상종은 국내에서 서식환경에 대한 연구가 비교적 많고 산림과 농경지를 서식지로 이용하는 고라니를 선정하였다. 고라니는 소목 사슴과 고라니속에 속하는 동물로 전 세계적으로 두 아종만이 분포하는데 그 중 한 아종이 우리나라에 분포하는 Korean water deer(*Hydropotes inermis argyropus*)이다(Cooke and Farrell, 1983). 고라니는 우리나라에서 제주도와 울릉도 등을 제외하고 전역에 분포하는 야생동물로서 높은 산맥에 접한 산기슭에 서식하는 것으로 알려져 있다(원병휘, 1967). 이처럼 고라니는 우리나라 생태계에 잘 적응하여 분포하고 있으며 서식지 이용 특성이 분명한 종으로서 종분포모형의 적용에 적합한 종으로 판단된다.

2) 자료수집 및 분석

우리나라에서 국가적으로 실시하고 있는 종조사 자료는 전국자연환경조사를 통해 수집된 자료로서 2005년

Fig. 2. Study area.

까지 2차 조사가 완료되어 현재(2006~2013) 3차 조사가 진행 중이다. 3차 전국자연환경조사는 해양을 제외한 전국토를 대상으로 실시하며, 대상지역의 기본단위는 위, 경도를 2' 30" 간격으로 나눈 격자이다(국립환경과학원, 2007). 3차 조사는 격자 단위로 전국을 동일한 간격으로 조사한 자료이기 때문에 2차까지의 조사에 비해 보다 균일한 조사지역 선정이 가능해져서 야생동·식물 종분포모형에 더욱 적합한 기초자료를 제공해 줄 수 있을 것으로 기대된다. 그러나 종 발견지점에 대한 자료만을 제공하고 있기 때문에 출현/비출현 기반의 종분포모형을 활용하기는 한계가 있다. 따라서 종출현 모형에 최적화되어있는 모형을 적용하는 것이 바람직하다.

본 연구에서는 3차 전국자연환경조사 결과를 활용하기 위해 3차 전국자연환경조사 보고서에 포함되어 있는 권역별 보고서에서 대상종별 발견지점 정보를 확인하여 해당 지점을 도면화 하고 이를 종출현지점으로 설정하였다. 본 연구 대상지에서는 총 193개의 고라니 발견지점이 파악되었으며 이는 포유류 중 가장 많은 발견지점으로 우리나라에서의 고라니 서식 환경을 확인할 수 있다.

분석에 활용한 환경변수는 대상종인 고라니의 서식 환경에 대한 문헌조사를 통해 선정하였다. 우리나라와 다른 아종인 Chinese water deer(*Hydropotes*

inermis inermis)는 중국 북동부지역에 서식하며 주로 호수, 강, 사초지, 억새, 갈대밭 등의 습지와 그 부근의 낮은 산림지대나 구릉지에 분포하는 것으로 알려져 있다(Zhang, 2000). 또한 산림지역에 비해 초지를 선호하여 대부분의 시간을 산림지역에 인접한 갈대밭이나 초원에서 생활한다(Zhang, 2000). 우리나라에 서식하는 고라니는 물이 있는 땅을 좋아하며 논밭 근처 낮은 산을 선호하는 반면 가파르고 바위가 많은 험한 산악지대에서는 살지 않거나 수가 적은 것으로 조사되었다(문교부, 1967; 최태영과 최현명, 2007). 또한 고라니 서식지 연구를 통해 고라니는 토지유형별로 침엽수림과 나지, 갈대밭, 20도 이상의 경사지, 남향, 수계에서 가까운 지역, 도로에서 먼 지역을 서식지로 이용하는 것으로 분석되었다(최서윤, 2003). 이는 경사도에 따른 서식지 선호도는 김백준(2011)의 연구와 유사하지만 활엽수림을 선호한다는 점은 차이가 있는 부분이다. 환경부 공간정보서비스에서는 고라니가 높은 산에 연맥된 산기슭에 서식하며 야지에 가깝고 나무가 적당히 무성한 남향의 사면에서 또는 평지 버들밭에서, 억새가 무성한 황무지에서, 산기슭 가까운 논에서 흔히 볼 수 있다고 소개하고 있다. 이처럼 고라니 서식지 특성이 지형, 토지피복 및 식생, 수계 및 도로와의 거리 등과 관계가 있는 것으로 확인되었으므로 본 연구에서는 해당

Table 1. Environmental variables and data processing

Environmental variables	Data processing	Data type	Meta data
Elevation	DEM	Continuous	DEM (Digital elevation model)
Slope	Surface analysis (slope)	Continuous	
Aspect	Surface analysis (aspect)	Categorical	
Topography	TPI (topographic position index)	Categorical	
Land cover	Reclassification of Land cover map	Categorical	Land cover map (v.2)
Distance from forest edges	Distance analysis	Continuous	
Distance from natural grasslands	Distance analysis	Continuous	
Distance from agricultural areas	Distance analysis	Continuous	
Distance from water body	Distance analysis	Continuous	River map
Distance from roads	Distance analysis	Continuous	Transportation digital map

Fig. 3. Spatial distributions of environmental variables applied species distribution models.

환경변수 적용을 위해 DEM(표고, 경사, 향, 지형유형, 습윤도), 토지피복지도(초지 등 토지피복 유형, 산림 가 장자리로부터의 거리, 자연초지로부터의 거리), 수계망 도(수계로부터의 거리), 도로망도(도로로부터의 거리) 를 기초자료로 분석을 진행하였다(Fig. 3). 향과 지형유 형, 토지피복 유형 변수는 범주형 변수, 나머지는 연속 형 변수로 적용하였다(Table 1). 이 중 지형유형은 Weiss(2001)가 구분한 산악지형 10개 유형 기준으로 구분하였다. 문헌조사를 통해 도출된 환경변수는 총 10 개로 정리될 수 있다(Fig. 3). 기초 환경변수는 30 m 격자단위로 통일하였으며 변환 및 분석은 ArcGIS 9.3(ESRI Inc., U.S.A) 소프트웨어를 사용하였다.

3) 중분포모형 적용

기계학습식 모형의 장점은 주어진 자료를 내부 알고 리즘에 따라 분석하여 각각의 방법론에 따라 가장 적합 한 예측 결과를 도출할 수 있다는 점이다. 물론 기계학 습식 모형이 통계 기반의 모형보다 성숙한 통계적 방법 론이 아니기 때문에 모형의 사용에 대한 가이드라인이 적고 예측에 있어서 오류의 양을 측정하는 방법이 적다 는 한계가 있다(Phillips *et al.* 2006). 이러한 한계에도 불구하고 기계학습식 모형은 데이터를 조작하고 원하는 환경변수를 적용하여 다양한 모형을 적용하고 최적의 결과를 도출하는데 효과적이기 때문에 향후 생태계 보 전 분야에서 많이 사용될 수 있는 방법론이라 판단된다.

본 연구에서는 중출현지점 정보를 가장 효과적으로

활용하기 위해 우리나라 종조사 자료의 특성을 모형에 가장 잘 적용할 수 있는 기계학습식 모형 중 GARP와 Maxent 모형을 이용하여 종분포모형을 분석하였다. GARP와 Maxent 모형은 종출현자료만을 활용할 경우에도 가장 우수한 분류정확도를 도출하는 것으로 연구되었다(Franklin, 2009). 즉, 우리나라와 같이 국가적인 종조사 자료가 존재하지만 그 자료가 출현정보로 한정된 경우 GARP와 Maxent 모형의 적용이 가장 효과적인 방법론이 될 수 있다.

GARP 모형은 야생동물의 생태적 지위 모형을 위한 유전학적 알고리즘으로서 종출현지점 자료만이 존재하는 넓은 범위, 저해상도에서 적용이 가능하여 현재 종분포모형에 광범위하게 적용되고 있다(Franklin, 2009). 이 모형은 종출현정보에서 인공지능 체계를 적용하여 긍정과 부정을 구분하는 2진법의 예측을 통해 결과를 예측한다. Maxent 모형 역시 종출현정보만을 활용하여 최고의 서식지 예측 결과를 도출하는 모형이다. Maxent 모형의 기본 개념은 최대 엔트로피 분포 가능성을 찾아내서 대상종의 분포 가능성을 예측하는 것이다. Maxent 모형은 모형개발자가 설정한 변수와 제한요소를 이용하여 확률분포를 예측하는 기계학습모형으로 종의 출현과 변수의 선형, 비모수적인 관계들을 잘 표현하는 장점이 있다(Phillips *et al.*, 2006; 서창완 등, 2008a). 본 연구에서는 두 모형을 동일한 환경변수로 적용하여 국내에서 적용 가능한 결과를 도출할 수 있는지 확인하였다. 모형의 세부 설정은 기존 연구를 통해 기본값을 유지한 상태로 분석을 진행하였다. 분석 결과는 ROC(receiver operating characteristic)의 AUC(area under the ROC)을 통해 검증하고, 이와 함께 Spearman 순위 상관분석을 실시하여 두 모형의 결과를 비교하였다.

3. 결과 및 고찰

1) 종분포모형 적용 결과

적용한 기계학습식 모형인 GARP와 Maxent 모형은 요구하는 자료가 크게 종출현지점 포인트 자료와 환경변수 격자 자료이다. 종출현확률과 모형에 적용된 환경

변수와의 관계를 분석한 결과 표고는 100 m 이상의 지역에서 400 m 지역까지 출현확률이 증가하는 경향을 보여주었다. 경사는 20~25° 범위에서 가장 높은 출현확률을 보였으며, 향은 평지를 제외한 모든 향에서 유사한 출현확률 분포가 확인되었다. 총 10개 유형으로 구분된 지형유형에서는(Weiss, 2001; 장광민 등, 2009), 산악수로(shallow valley)에서 고라니의 출현확률이 높았으며 평지(plain)에서 값이 낮게 나타났다. 토지피복유형별로는 산림지역과 자연초지에서 출현확률이 높게 나타났으며 특히 활엽수림에서의 출현확률이 가장 높게 나타났다. 거리변수 중 산림 가장자리에서 가까울수록, 자연초지에서 가까울수록 고라니의 출현확률이 높게 나타났다. 농경지는 500 m 이내의 거리에서 고라니가 주로 분포하고 있으나 그 이상의 거리에서도 분포확률이 높게 나타났다. 수계에서 2,500 m 이내의 거리에서 고라니 출현확률이 높게 나타났으며, 도로와의 거리는 2,000 m 거리까지 출현확률에 영향을 미치는 것으로 확인되었다. 이러한 결과는 기존의 고라니 서식지 연구와 유사한 경향을 보이는 것으로 확인되었는데 특히 20~25° 경사, 활엽수림을 서식지로 활용한다는 부분은 김백준(2011)의 연구와 동일한 것으로 파악되었다. 또한 고라니가 수변지역 및 초지를 선호하는 기존의 연구에서와 마찬가지로, 수계와 초지에서 가까운 거리, 산악수로지역에서 고라니 출현확률이 높게 나타나고 있음을 확인하였다(최태영과 최현명, 2007). 또한 높고 경사가 급한 산림지역보다 낮은 산림지역에서 주로 서식하는 특성도 본 연구와 동일한 부분이다(최태영과 최현명, 2007). 지형인자 중 대부분의 서식지 선호경향은 최서윤(2003)의 연구와 유사한 결과를 보이고 있으나 고라니가 남향을 선호하는 경향을 보이고 있는 부분은 본 연구와 차이가 있다. 이는 Thomas(1979)의 연구에서 사슴류가 겨울철에 북쪽의 경사지대보다 온도가 높아 눈이 녹은 약간 경사진 남향을 선호한다는 연구 결과를 통해 해석이 가능한 부분이다. 최서윤(2003)의 종조사가 늦은 가을(10월~11월)과 겨울(1월~2월)에 진행되어 고라니가 체온을 유지시킬 수 있는 남사면을 더 많이 이용하였기 때문이었을 것으로 판단된다(최서윤, 2003).

GARP 모형 적용 결과 고라니 서식이 예상되는 종분포확률 0.5 이상인 지역은 전체면적의 27.1%로 확인되었다. Fig. 4(a)에서와 같이 GARP 모형은 종출현확률

Fig. 4. Results of machine learning species distribution modeling.

을 0.5를 기준으로 비교적 명확하게 구분하고 있는 것을 확인할 수 있다. 이는 모형의 특성상 2진법에 따른 분류 기준을 택하기 때문인 것으로 판단된다. 또한 GARP 모형은 총 20번의 내부적 모형을 종합하여 최종 종출현 예상지역을 계산하게 되는데 각각의 모형이 비교적 유사한 예측 결과를 도출하여 이러한 결과가 도출된 것으로 판단된다. 모형의 훈련 정확도(train accuracy)는 0.65, 정확도 검증을 위한 검증 정확도(test accuracy)는 0.61로 확인되었다. Maxent 모형 적용 결과 고라니 서식이 예상되는 종출현확률 0.5 이상인 지역은 전체면적의 16.3%로 확인되었다. Maxent 모형을 통해 예측된 종출현지역은 GARP 모형 적용 결과에 비해 약 10% 정도 낮게 계산되었다. 이러한 차이는 모형의 차이에 의한 것이라 판단되는데 Maxent 모형은 Fig. 4(b)와 같이 종출현확률이 0.25~0.5 사이의 값이 넓게 분포하는 등 중간값이 GARP 모형에 비해 넓은 면적을 차지하는 경향을 보이고 있다. 모형의 검증은 ROC(Receiver Operating Characteristic)의 AUC(Area Under Curve) 값으로 계산되었으며 Maxent 모형의 경우 0.85로 계산되어 본 모형의 설명력이 충분히 높은 것으로 판단되었다(Phillips and Dudik, 2008). 모형의 검증을 위해 종출현지점의 50%를 검증 데이터로 활용하여 정확도를 검증한 결과 0.69

로 양호한 값을 보이는 것으로 확인되었다. 도출된 각각의 모형 결과를 비교하기 위해 서식지 적합성 지수를 20개로 순위화하고 면적대비 종출현지점과의 관계를 Spearman 순위 상관관계로 분석한 결과 Maxent 모형($\rho=0.71$, $p<0.01$)이 GARP 모형($\rho=0.55$, $p<0.05$)보다 상관관계가 더 높은 것으로 분석되었다. 이는 Maxent 모형이 종 출현정보를 서식지 모형에 더 잘 반영하고 있다는 것을 의미한다. Maxent 모형의 결과는 GARP 모형과 비교할 때 0.5 이상의 지역이 약 10% 적게 분석된 것을 확인할 수 있다. GARP 모형을 통한 종본포모형은 중간값이 적어 높은 서식지 적합성을 가지는 것으로 판단되는 지역이 과대평가되는 경향이 있는 것으로 판단된다(Gontier *et al.*, 2010). 두 모형 모두에서 서식지로 확인된 지역은 전체의 11.1%, Maxent 모형, GARP 모형에서만 서식지로 분석된 지역은 각각 16%, 5%로 분석되었다.

2) 국내에서의 적용 가능성 고찰

현재 추진 중인 3차 조사(2006~2013)는 기존의 권역별 조사에서 탈피해 전국을 1/25,000 축척으로 균등 분할하여 도엽별로 연차적인 조사를 추진하고 있으며 이를 통해 전국에 대한 균등한 종조사 자료 구축이 가능하게 되었다. 4차 조사부터는 제1차 국가경쟁력강화위원

회(2008)의 결정에 따라 자연환경조사 주기를 단축하여 5년 주기로 전국 조사가 진행될 계획이다(환경부, 2010). 이러한 종조사 기준의 개선, 주기 단축은 전국자연환경조사가 국내의 서식지 연구에 중요한 기초자료로 활용되는 기반이 될 수 있을 것이라 기대된다.

앞서 자료수집 및 분석 부분에서 언급한 바와 같이 우리나라 종조사 자료는 종출현지점을 조사하는 형태이기 때문에 다양한 통계 기반의 중분포모형을 적용하는데 한계가 있다(서창완 등, 2008b; 송원경, 2011). 따라서 전국자연환경조사 자료를 활용할 경우에는 출현지점에 특화된 모형을 사용하거나 비출현지점을 통계적인 방법을 통해 설정하고 출현/비출현 모형을 적용할 수 있다. 중분포 예측 결과를 활용한 서식지 분석 결과의 신뢰성을 높이기 위해서는 국가적으로 신뢰도 있는 조사 자료를 활용하고 검증된 모형을 사용하여 분석을 진행하는 것이 중요하다. 따라서 국가적으로 진행중인 종조사 결과를 활용하되 해당 조사자료의 특성에 가장 적합한 모형을 적용하는 것이 결과적으로 신뢰성 높은 결과를 분석할 수 있는 방법이 된다.

이전까지 출현자료만을 이용할 경우 생태학 분야에서 GARP가 가장 보편적으로 이용되고 있다(Franklin, 2009). 이는 기계학습식 모형의 특성상 데이터의 입력 및 결과값의 출력이 용이하고 결과를 쉽게 파악할 수 있기 때문인 것으로 판단된다. 이러한 편의성뿐만 아니라 모형의 예측 결과에서도 생태학 분야에서 우수한 결과를 도출하는 것으로 알려져 있다. 특히 넓은 면적을 대상으로 분석이 진행되거나 역사적 기록물을 이용한 저해상도 자료의 분석에서도 폭넓게 활용된다(Franklin, 2009). GARP는 대부분의 생태학 분야에서 중분포모형으로 적용되는데 최근 기후변화로 인한 중분포 영향(Peterson *et al.*, 2002; Martinez-Meyer *et al.*, 2004), 식물을 비롯한 침입종의 지리적 분포(Peterson and Vieglais, 2001; Peterson, 2003) 등에서 많이 활용되고 있다(Franklin, 2009). 그러나 본 모형은 완벽한 블랙박스 모형으로서 각각의 환경변수가 예측모형에 어떻게 기여하는지 파악하는 것이 매우 어렵다. 또한 최대 1,000회 반복으로 원하는 모형 예측 결과 개수를 설정하도록 되어 있어 예측에서 누락되는 지역이 발생하는 경우가 있기 때문에 이 누락값을 어디까지 허용해야 하는지에 대한 문제가 있다. 일부 연구에서는 5%를 기준

값으로 설정하고 있는데(Phillips *et al.*, 2006), 본 연구에서도 이러한 기준값은 적합한 것으로 판단된다.

서창완 등(2008a)은 종조사 자료의 특성에 따라 어떠한 모형을 적용하는 것이 가장 효과적인지 비교하고 있는데 종출현자료만을 이용할 경우 Maxent 모형이 효과적이라고 분석한 바 있다. 본 연구 결과에서도 Maxent 모형의 훈련 정확도와 검증 정확도가 GARP 모형보다 우수한 것으로 확인되었다. Maxent 모형의 장점은 GARP 모형 등 기계학습식 모형의 장점과 동일하게 자료의 입·출력이 용이하고 결과의 해석이 쉽다는 점이다. 뿐만 아니라 본 모형은 각각의 환경변수가 모형의 예측결과와 어떠한 관계가 있는지를 보여주는데 이를 통해 결과의 해석이 부분적으로 가능하다. 즉, Maxent 모형은 기계학습식 모형이지만 통계적인 방법론을 함께 적용하고 있어 통계학습식 모형이라 구분할 수도 있다(Franklin, 2009). GARP 모형과 Maxent 모형을 비교한 연구들이 최근 많이 소개되고 있으며, Phillips *et al.*(2006), Gontier *et al.*(2010)은 Maxent 모형이, Peterson *et al.*(2007)은 GARP 모형이 더 서식지 적합성을 예측하는데 효과적이라고 분석한 바 있다. 본 연구의 결과에서도 Maxent 모형의 훈련 및 검증 정확도 기준으로 활용한 AUC값이 더 크게 계산되었다. Maxent 모형은 비교적 고른 중분포확률값 분포를 보여주고 있는 부분이 극단적인 중분포확률 분포를 갖는 GARP 모형과 다른 점이라 할 수 있다.

기계학습식 중분포모형은 통계적 모형에 비해 데이터 가공, 편집, 요구되는 종조사자료 등에 있어서 강점이 있을 뿐만 아니라 예측력 또한 우수하여 국내의 서식지 예측에 적합한 모형이라 판단된다. 현재 전 세계는 생물다양성 확보를 위한 국가적 경쟁이 진행되고 있다. 우리나라에서 확보하고 있는 생물종 자원의 효과적인 보전 및 관리를 위해 서식지 보전지역 선정 방법론으로서 기계학습식 중분포모형의 활용 가능성이 앞으로 더욱 높아질 것으로 기대된다.

4. 결 론

우리나라에서 국가적으로 수집하고 있는 종조사자료의 효과적인 활용을 위해서는 종출현지점을 이용한 모

형에 최적화되어있는 기계학습식 모형 중 GARP 모형과 Maxent 모형을 적용하는 것이라 판단된다. 기계학습식 종분포모형 중 기존 연구를 통해 활용가능성이 높다고 판단되는 GARP 모형과 Maxent 모형을 국내 자료에 맞게 적용·분석한 결과 두 모형 모두 대상지에서 고라니의 서식지를 예측하는데 적합한 모형인 것으로 확인되었다. 훈련 정확도 및 검증 정확도, Spearman 순위 상관계수는 Maxent 모형이 다소 높게 나타났으나 이는 대상지 특성, 환경변수 종류, 종 선정 등에 따라 다른 결과가 나타날 수 있으므로 어떤 모형이 국내 연구에 더 적합한지 일반화하기는 어렵다. 또한 기계학습식 모형이 블랙박스 모형으로서 내부 구조를 확인하기 어려운 단점이 있으나 분석 결과의 정확도가 다른 모형에 비해 높고 Maxent 모형처럼 환경변수와 출현 확률간의 관계를 확인할 수 있는 모형도 존재하므로 그 활용성이 더욱 높아질 수 있다. 이와 같이 기계학습식 종분포모형은 생물다양성 확보를 위한 국가전략 수립에 가장 효과적인 서식지 보전지역 확보 방법론으로 활용될 수 있을 것으로 판단된다. 특히 우리나라 종조사 자료 및 환경공간정보 여건을 고려할 때 향후 다양한 연구에서 기계학습식 모형에 의한 서식지 보전·관리 연구가 진행될 것으로 기대된다.

참고문헌

- 관계부처합동, 2009. 국가생물다양성 전략 및 이행계획.
- 관계부처합동, 2010. 저탄소 녹색성장 기본법 시행에 따른 국가 기후변화 적응대책(2011~2015).
- 국립환경과학원, 2007. 제3차 전국자연환경조사 지침.
- 권혁수, 2011. 보호지역계획을 위한 생물다양성 통합평가모형 -지리산과 덕유산, 가야산 권역을 대상으로-, 서울대학교 대학원 박사학위논문.
- 김백준, 2011. 한국산 고라니(*Hydropotes inermis inermis*)의 생태·유전학적 특성, 서울대학교 대학원 박사학위논문.
- 대한민국, 2009. 생물다양성협약 제4차 국가보고서.
- 서창완, 박유리, 최윤수, 2008a. 위치자료의 종류에 따른 생물종 분포모형 비교 연구, 한국지형공간정보학회지, 16(4): 59-64.
- 서창완, 최태영, 최윤수, 김동영, 2008b. 설악산 산양을 대상으로 한 야생동물 서식지 적합성 모형에 관한 연구, 한국환경복원녹화기술학회지, 11(3): 28-38.
- 송원경, 2011. 공간그래프 이론을 적용한 삼 서식지 네트워크 모형 개발, 서울대학교 대학원 박사학위논문.
- 원병휘, 1967. 한국동식물도감 제 7권 동물편(포유류), 문교부.
- 이동근, 김호걸, 2010. Maxent 모형을 이용한 서식지 잠재력 평가 -하천으로부터의 거리, 하천의 차수, 토지이용을 중심으로-, 한국환경복원녹화기술학회지, 13(6): 161-172.
- 이동근, 송원경, 2008. 삼의 서식지 적합성 평가를 위한 분석단위 설정 및 보전지역 선정 -충청도 지역을 중심으로-, 한국조경학회지, 36(5): 64-72.
- 최서윤, 2003. 대부도에 서식하는 고라니의 서식지 적합도 평가모형 개발에 관한 연구, 서울대학교 대학원 석사학위논문.
- 환경부, 2010. 제2차 야생동·식물보호기본계획(2011~2015).
- Cooke, A. and L. Farrell, 1983. Chinese water deer, The British Deer Society.
- Degraer, S., E. Verfaillie, W. Willems, E. Adriaens, M. Vincx, and V. Van Lancker, 2008. Habitat suitability modelling as a mapping tool for macrobenthic communities: an example from the Belgian part of the North Sea, *Continental Shelf Research*, 28: 369-79.
- Franklin, J., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*, Cambridge: Cambridge University Press.
- Gontier, M., U. Mörtberg, and B. Balfors, 2010. Comparing GIS-based habitat models for applications in EIA and SEA, *Environmental Impact Assessment Review*, 30: 8-18.
- Guisan, A. and W. Thuiller, 2005. Predicting species distribution: offering more than simple habitat models, *Ecology Letters*, 8: 993-1009.
- Martinez-Meyer, E., A.T. Peterson, and W.W. Hargrove, 2004. Ecological niches as stable

- distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity, *Global Ecology and Biogeography*, 13: 305-314.
- Peterson, A.T., 2003. Predicting the geography of species' invasions via ecological niche modeling, *The Quarterly Review of Biology*, 78(4): 419-433.
- Peterson, A.T., M.A. Ortega-Huerta, J. Bartley, V. V. Sanchez-Cordero, J. Soberon, R.H. Buddemeier, and R.B. Stockwell, 2002. Future projections for Mexican faunas under global climate change scenarios, *Nature*, 416: 626-629.
- Peterson, A.T. and D.A. Vieglais, 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem, *Bioscience*, 51(5): 363-371.
- Peterson, A.T., M. Papes, and M. Eaton, 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent, *Ecography*, 30: 550-560.
- Phillips, S.J., R.P. Anderson, and R.E. Schapire, 2006. Maximum entropy modeling of species geographic distributions, *Ecological Modelling*, 190: 231-259.
- Thomas, J.W., 1979. Wildlife habitats in managed forests. USDA Forest Service.
- Zhang, E., 2000. Daytime activity budget in the Chinese water deer, *Mammalia*, 64: 163-172.
- <http://egis.me.go.kr/>
- <http://library.me.go.kr/>
- <http://www.cs.princeton.edu/~schapire/maxent/>
- <http://www.nhm.ku.edu/desktopgarp/>