



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩 士 學 位 論 文

GMM과 Logistic regression model을
활용한 제주노루 서식지 예측

嶺南大學校 大學院

生 命 科 學 科

生 命 科 學 專 攻

金 俊 赫

指導教授 張 甲 洙

2017 年 8 月





碩 士 學 位 論 文

GMM과 Logistic regression model을
활용한 제주노루 서식지 예측

指導教授 張 甲 洙

이 論文을 碩士學位 論文으로 提出함

2017年 8月

嶺南大學校 大學院

生命科學科 生命科學專攻

金 俊 赫



金俊赫의 碩士學位 論文을 認准함

審 査 委 員 유 시 옥 (印)

審 査 委 員 장 갑 수 (印)

審 査 委 員 석 호 영 (印)

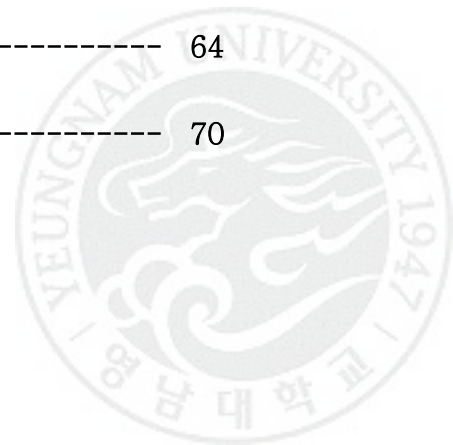
2017 年 8 月

嶺南大學校 大學院



목 차

I. 서론	1
II. 연구사	3
III. 재료 및 방법	4
1. 연구대상종	4
2. 연구대상지	7
3. 연구방법	11
IV. 결과 및 고찰	18
1. 한남시험림 환경 변수	18
2. MaxEnt 모델링결과	30
3. GMM 결과	37
4. Logisitic regression 결과	34
5. 현장검증	39
V. 결론	52
VI. 참고문헌	64
영문요약	70



<List of Table>

Table 1. roe deer GPS data from HEF-----	5
Table 2. List of environmental variables-----	10
Table 3. Selection of the proper variables derived from 14 environmental factors in previous study-----	14
Table 4. Selection of the proper variables derived from 23 environmental factors in this study-----	14
Table 5. Set data according to frequency-----	17
Table 6. Result of Percent contribution and Permutation importance previous study-----	33
Table 7. Result of Percent contribution and Permutation importance in this study-----	33
Table 8. A bundle of variables according to a characteristic-----	37
Table 9. Percent and Accumulative eigenvalues-----	40
Table 10. Result of logistic regression-----	45



List of Figures

Fig. 1. GPS tracking data from HEF	6
Fig. 2. Study area	7
Fig. 3. Hannam experimental forest subtropical forest	8
Fig. 4. MaxEnt 3.3.3k	12
Fig. 5. Environmental variables of DEM and hillshade in HEF	19
Fig. 6. Environmental variables of easpcet and slope in HEF	19
Fig. 7. Environmental variables of D_agriculture and D_forest(-) in HEF	21
Fig. 8. Environmental variables of D_road and D_stream in HEF	21
Fig. 9. Environmental variables of D_oreum and basin_map in HEF	23
Fig. 10. Environmental variables of somildo and kung in HEF	23
Fig. 11. Environmental variables of yung and lmsang in HEF	25
Fig. 12. Environmental variables of deciduous and evergreen in HEF	25
Fig. 13. Environmental variables of conifer and hect_each in HEF	27
Fig. 14. Environmental variables of count_under and TM_Jan in HEF	27
Fig. 15. Environmental variables of TM_Apr and Jan_NDVI in HEF	29
Fig. 16. Environmental variables of Apr_NDVI in HEF	29
Fig. 17. ROC curve in HEF	30
Fig. 18. Jackknife of regularized training gain	34
Fig. 19. Previous study prediction map of roe deers in HEF	35

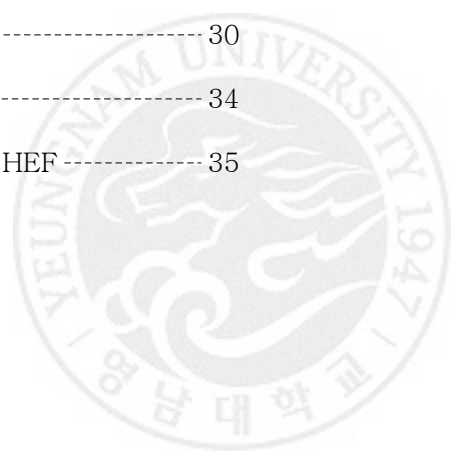


Fig. 20. In this study prediction map of roe deers in HEF -----	35
Fig. 21. result of GGM using parameter feature -----	37
Fig. 22. Prediction map of roe deers in HEF -----	39
Fig. 23. result of PCA -----	41
Fig. 24. result of GMM using PCA -----	43
Fig. 25. Result of AIC and BIC of logistic regression -----	44
Fig. 26. result of logistic regerssion -----	47
Fig. 27. Filed test area -----	49
Fig. 28. Field test result -----	50



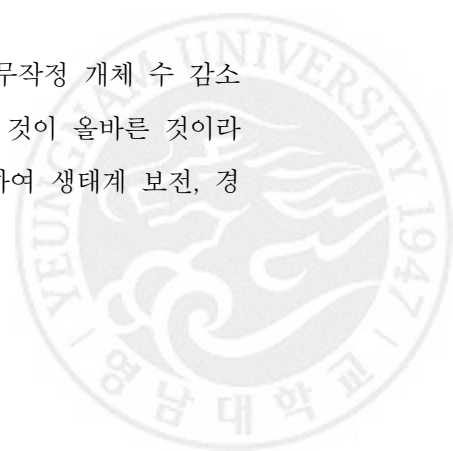
I. 서론

본 연구의 대상종인 제주 노루는 1980년 대에 시행된 제주노루 보호정책에 의해 개체 수가 급격히 증가하였다. 그 배경에는 제주노루를 제주도의 상징종으로 만들겠다는 제주도민과 제주시의 의지가 있는데, 먹이원 제공, 포획 도구 수거 등 보호 운동과 민간단체, 국립공원관리사무소의 노력으로 개체 수가 수천마리로 증가한 것으로 보고되었다. 또한 이를 이용하여 제주도의 관광자원으로 부상하기 까지 이르렀다(윤성일, 2003, 전대욱과 김도훈, 2011)

하나 이런 개체 수의 증가로 새로운 문제가 발생하였는데, 농업, 임업, 환경 등 다양한 분야에서 문제가 발생하였으며, 특히 농가에 대한 피해가 지속적으로 발생하였다. 이는 제주도에 제주노루의 상위포식자인 호랑이, 표범, 늑대 등의 천적이 오래 전에 사라지고, 온화한 기후인 제주도에 노루의 먹이원이 풍부해져 개체 수가 증가하였기 때문이다(김아름, 2016). 이런 문제들로 인해 제주특별자치도 시의회에서는 제주노루를 2013년 7월에서 2016년 6월까지 유해동물로 지정하여 5000여마리를 포획하고, 또한 유해동물로 2019년 6월 30일까지 재지정하기에 이르렀다.

한편, 제주노루의 포획으로 인하여 노루 개체수가 감소하자 보호해야 한다는 환경단체의 의견이 나오기 시작하였다. 피해농가에서는 노루에 대한 추가포획을 찬성하는 입장이지만, 환경단체의 경우 인간과 제주노루가 공생할 수 있는 방안을 찾아야 한다며 주장하고 있다. 또한 한 때 멸종위기에서 2만 마리 넘게 개체수가 증가하였던 제주노루가 현재는 7천 600여 마리 있는 것으로 추정되고 있는데, 언론에서는 신뢰할 수 없는 개체 수 통계라며 논란이 나오고 있다. 노루의 경우 산림에서 서식하고 활동하기에 정확한 개체 수 파악에 많은 어려움이 있다(김병수 등, 2007).

여러 논란이 있는 제주노루에 대해 인간의 이익을 위하여 무작정 개체 수 감소를 해야 하는 것이 아닌 인간과 공생할 수 있는 방법을 찾는 것이 올바른 것이라고 할 수 있다. 특히, 제주노루의 가치 및 공생할 근거에 대하여 생태계 보전, 경



제적 이익 등 다양한 시각에서 접근하여야 한다. 다양한 시각을 통한 접근이 없이 개체 수 감소 혹은 개체 수 보전을 시행하면 경제적 피해 혹은 제주노루의 멸종이라는 문제를 직면할 수 있기 때문이다.

이에 본 연구에서는 노루에 대한 이해를 위하여 공간적인 시각에서 접근하였다. 이미 선진국에서는 1970년대 후반 자연 환경 관리정책의 일환으로 생물종에 따른 서식지 평가를 실시하여 보전지역 및 복원 대상지를 선정하는 등 서식지 가치 평가를 생물종 관리에 적극 활용하고 있으며(U.S. Fish and Wildlife Service, 1980), 최근에는 GIS(geographic information system)기법과 원격탐사기법(remote sensing), 원격무선추적기법(radio-tracking), 공간통계모형 등의 기술 향상과 결합으로 서식지 평가 관련 연구가 활발하게 진행되고 있다(Schadt et al., 2002; 박용수, 2014). 특히, 종 분포 모형(SDM: Species Distribution Model)을 이용하여 노루의 분포 서식지를 확인하고자 다양한 종 분포 모형을 활용하였다. 연구 대상지인 한남시험림(HEF: Hannam Experimental Forest)에서 GPS추적 데이터를 이용하고 3가지 모형을 이용하여 분석을 하였다. 획득한 데이터의 경우 출현좌표(presence-only data)만 있는 것을 감안하여 presence-only data를 사용함에 있어 가장 합리적이고 효율적인 결과를 보여준다고 알려진 MaxEnt를 활용하였다. 또한, 파라미터 값들을 클러스터링 하여 기계학습 후 확률분포를 보여주는 가우시안 혼합 모델(GMM: Gaussian mixture model)을 이용하여 분석을 실시하였으며, 로지스틱 회귀를 이용하여 추가적으로 분석을 실시하였다. 이 결과들을 바탕으로 종 분포 모형을 구축하였으며, 한계점도 제시하였다.

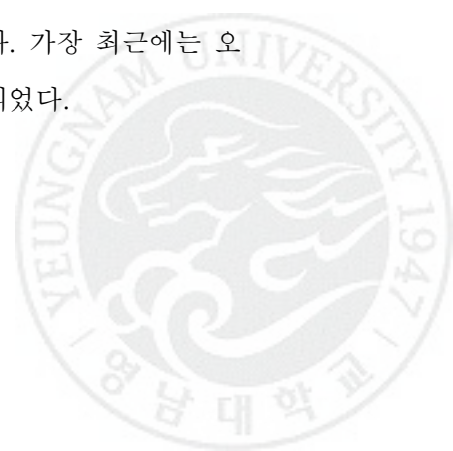


II. 연구사

노루에 대한 연구는 다양한 분야에서 수행되었다. 분류체계의 연구로는 1978년 Sokolov에 의해서 한 종으로 알려진 *Capreolus Pygargus* Gray가 *Capreolus pygargus pygargus*와 *Capreolus pygargus tianschanicus*로 분류가 나누어진 것이 처음으로 시도되었으며, 1999년 Danilkin에 의하여 형태학, 세포핵학, 교잡, DNA 지문분석 등을 이용하여 분류체계를 확립하였다. 이 뿐만 아니라 노루 및 노루 아종들을 포함하여 노루들의 두개골 변화에 대해서 연구하기도 하였다(Irina N. Sheremetyeva, 2008).

2004년에는 A. Coulon에 의해서는 “Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach” 제목으로 노루의 유전자 흐름에 대해 연구가 이루어졌다. 2008년에는 Jiang에 의해 중국 북동지방에 있는 완다산을 대상지로 인강의 행동이 노루에 어떤 영향을 주는 지 연구를 수행하였다. 2013년에는 Vladimir A Trifonov에 의해 노루의 B chromosomes에서의 단백질 전사에 대한 연구도 이루어 졌다.

국내에서도 다양한 연구가 이루어졌는데 2001년에는 김지은, 김문홍에 의해 노루가 선호하는 한라산의 자생식물에 대해 조사가 수행되었다. 연구대상지인 한남시험림에서 이루어진 연구도 있는데, 김은미 등(2013)에 의해 노루 뿔이 성장 과정에 따른 수목피해에 대해 연구가 있었으며, 김은미 등(2013)에 의해 암노루 출산전후 행동권 범위에 대한 연구도 수행되었다. 뿐만 아니라 김도훈과 홍영교(2006, 2008), 김병수 등(2007), 오장근(2010) 등에 의해 제주노루이 개체수 관리방안, 시스템 다이내믹스 모델링에 관한 연구가 이루어졌다. 가장 최근에는 오장근(2015)에 의해 제주도 전역 노루 개체 수 조사가 수행되었다.



III. 재료 및 방법

1. 연구대상종

노루는 유럽노루와 시베리아 노루 2종으로 각각 유럽에 분포하는 종과 아시아 북부에 분포하는 종으로 2종이 존재하는 것으로 알려져 있다. 본 연구에서는 제주도에 서식하고 있는 노루(*Capreolus pygargus tianschanicus*, 제주노루)를 대상 종으로 지정하였으며, 이 제주노루는 우제목(*Artiodactyla*) 사슴과(*Cervidae*) 노루속(*Capreolus*)에 속하는 시베리아노루(*Capreolus pygargus*)의 아종으로 알려져 있다. 학자들에 따라 종의 분류학적 위치가 다른 점으로 보아 분류학적인 부분에서 연구가 더 필요한 것으로 보이며, 본 연구에서는 *Capreolus pygargus tianschanicus*라고 주장한 윤성일(2003), 오장근(2004), 김병수 등(2007)을 따라 명명하였다(Pradeep Adhikari et al., 2016).

박찬열 등(2008)년에 의하면 노루는 동아시아 지역(러시아, 중국)에 분포하는 것으로 알려져 있으며, 한반도 전역에서 서식하고 있으며, 제주도 지역에서 가장 많은 수가 서식하는 것으로 알려져 있다. 해발 200m 이상 중산간 지대에서부터 한라산 정상까지 분포하며, 숲과 가까운 목장지대에서 많은 개체가 관찰되며, 겨울에는 먹이를 찾으러 고지대의 노루가 저지대로 내려오기도 한다고 언급하고 있다.

오장근(2004)과 최대영과 최현명(2007)에 따르면, 숲이나 초지, 관목림이 있는 개간지를 선호하여 서식하고 있으며, 풀밭과 숲의 경계지역, 깊은 산의 평평한 지형을 선호한다고 말하고 있다. 박찬열 등(2008)에 따르면 노루는 땅 위에 자라는 풀이나 나무의 새순, 잎을 즐겨먹는 것으로 알려져 있다. Erik P. Ensing 등(2014)에 따르면 노루는 저녁과 새벽, 황혼이 지는 시간에 높은 활동성을 가지는 것으로 알려져 있다. 주로 산림지역 내에 서식하는 산림성 우제목으로 알려진 노루는 (Fruziński and Łabudzki, 1982) 서식지 내 먹이자원의 질과 양, 물, 잠자리 및 휴식처 등 서식지 구성 요소의 특성에 따라 많은 차이를 보인다고 알려져 있다 (Leopold, 1993).

분석에서 사용된 좌표를 2012~2013년동안 국립산림과학원 난대·아열대산림연구소에서 제공받은 것으로 6마리의 노루를 대상으로 CDMA방식으로 얻은 좌표데이터이다(table 1). 포획한 개체는 한남시험림에서 기존에 서식하는 개체들로 전체 지역이 아닌 포획이 가능한 지역에서 포획한 노루이기에 한남 시험림 전체에 대해 상대적으로 편향된 경향을 보이고 있다.

Table 1. roe deer GPS data from HEF

개체	기간	성별	시간간격(시)
6023_M	2012.04.24~2013.04.21	M	6
5947_F	2012.05.09~2012.07.17	F	1
4930_F	2012.07.12~2012.12.05	F	2
4028_F	2013.07.09~2013.07.21	F	6
4022_M	2013.06.15~2013.06.28	M	6
2606_F	2013.04.04~2013.05.07	F	6



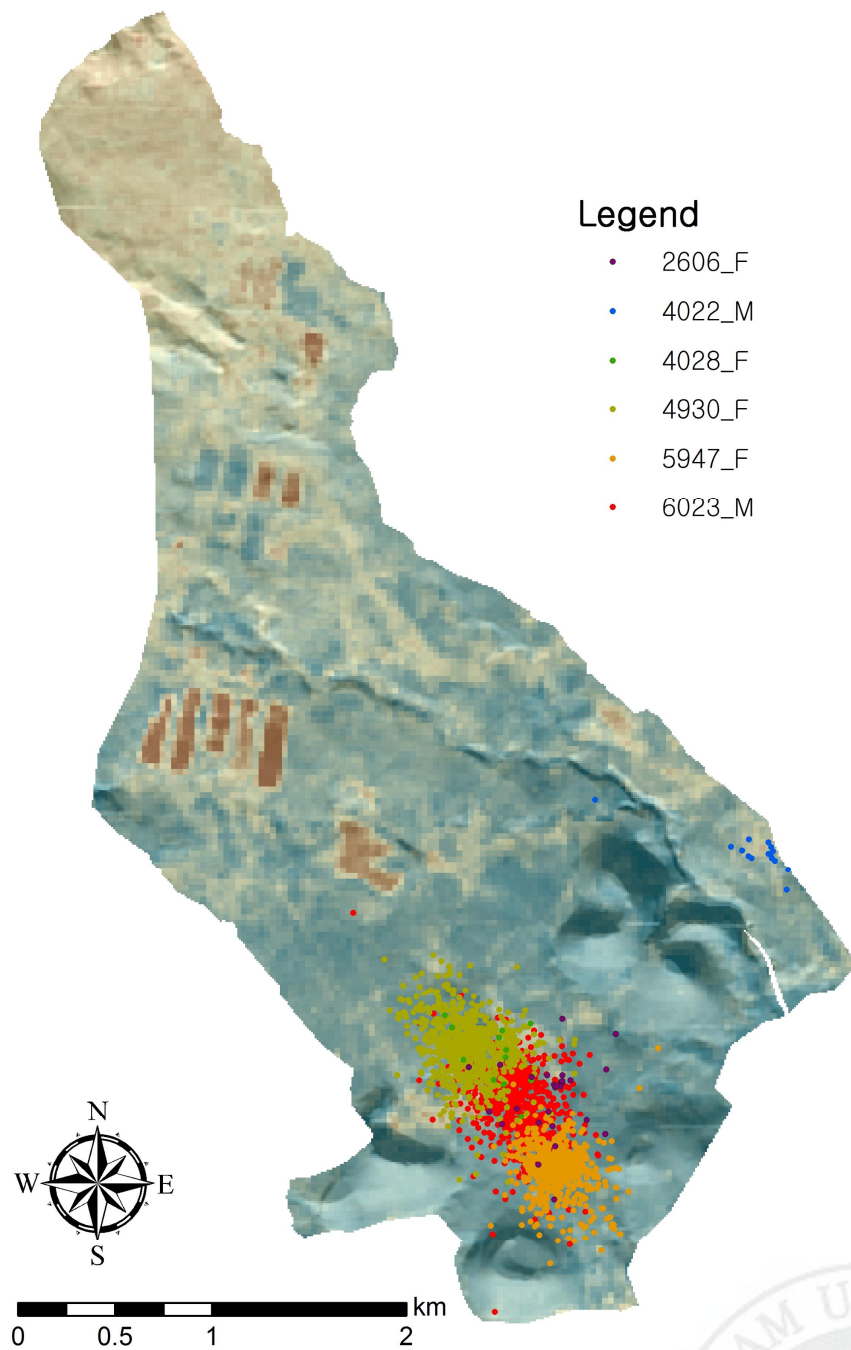


Fig. 1. GPS tracking data from HEF



2. 연구대상지

한남 시험림(Hannam experimental Forest, HEP)은 약 1,203ha의 넓이를 가지는 지역으로 지속가능한 산림경영 SFM(Sustainable Forest Management)을 위해 조성된 모델림이다(산림청). 남제주군 중산간 지역 해발 약 300m ~ 750m에 해당되며 연평균 3,000mm의 강우량을 보이는 지역이며, 온대, 난대, 아열대 기후를 가지고 있으며 한라산 남동사면에 위치하고 있다. 현재 국립산림과학원에서 ‘난대림의 지속가능한 산림경영 시범 산림’으로 조성하기 위해 관리하고 있는 지역이다. 또한 국제적인 산림관리기구인 FSC(Forest Stewardship Council)로부터 산림경영인증을 받았다.

한남 시험림은 인공조림지로 주 구성은 삼나무로 이루어져있으며(원현규, 2014), 극상림에 해당하는 서어나무 등 여러 종이 서식하는 것으로 알려져 있다. 뿐만 아니라 다양한 하층식생 및 은엄폐가 가능한 지형들이 분포하고 있어 초식동물인 노루, 육식성 포식자인 오소리, 족제비 등 14종의 포유류와 80종의 조류 중 매, 팔색조, 참매 등 천연기념물 14종, 삼광조, 붉은해오라기 등 멸종위기야생동물 12종이 살고, 비바리뱀 등 파충류 7종, 양서류 5종 등이 서식하는 것으로 알려져 있다

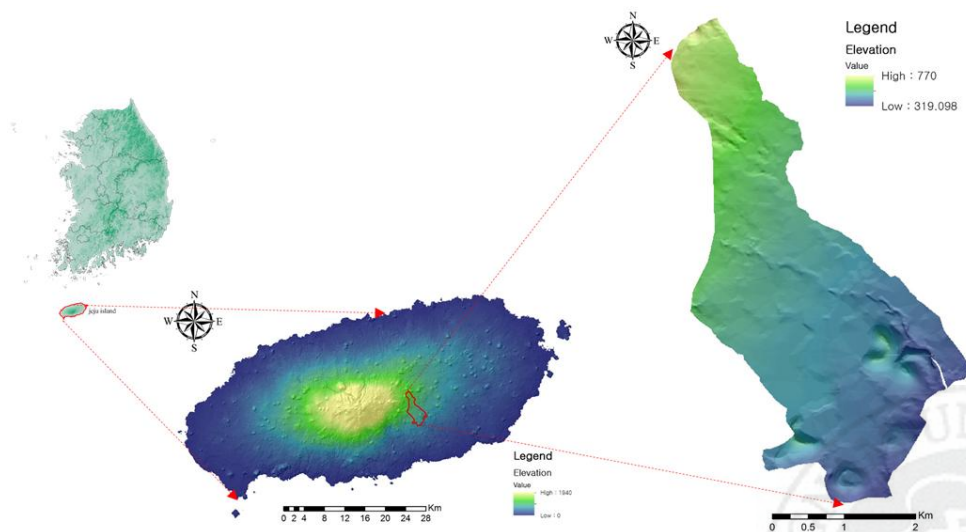


Fig. 2. Study area

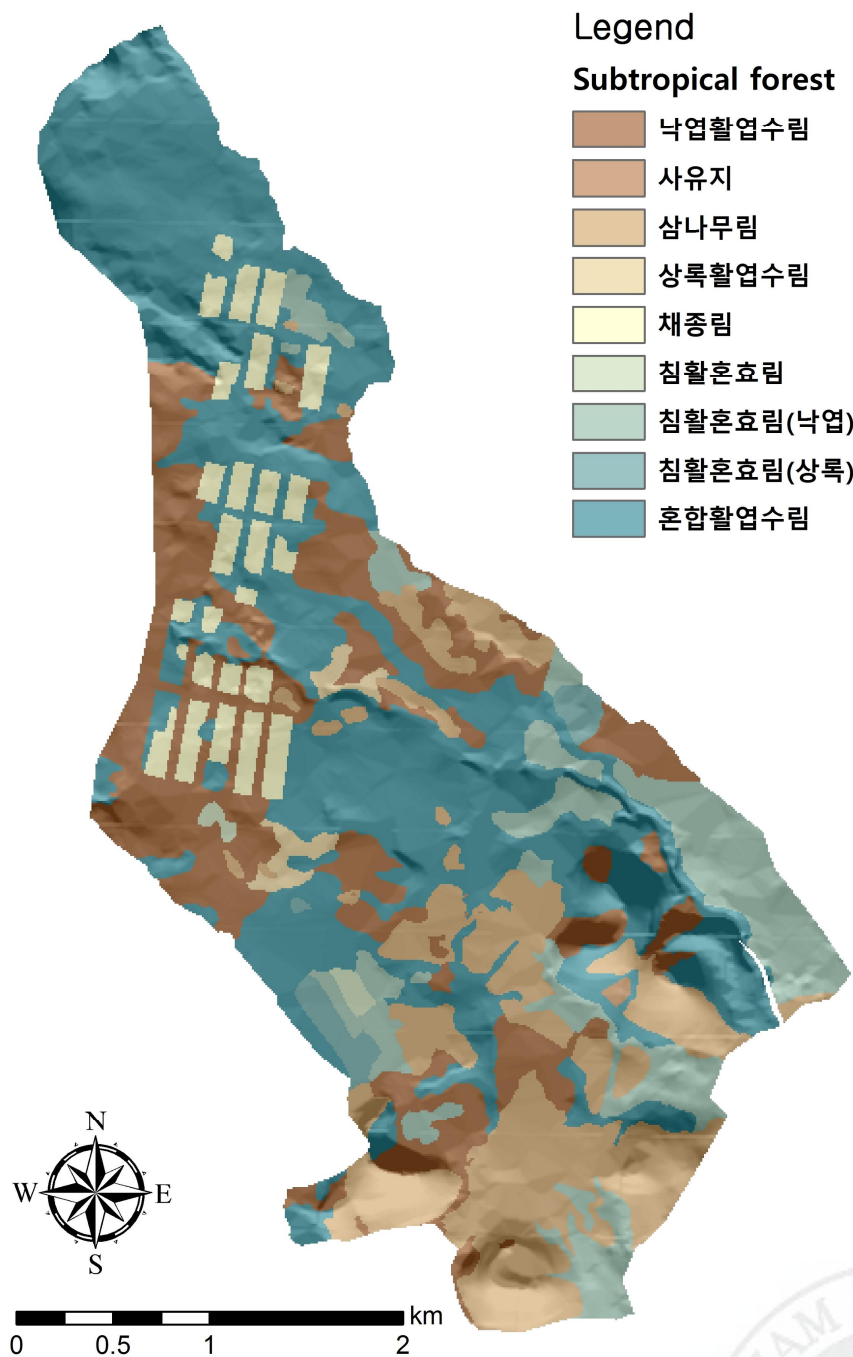


Fig. 3. Hannam experimental forest subtropical forest



한남시험림의 환경변수들은 23가지를 사용하였다. 등고선 자료를 이용하여 만들 수 있는 4가지 변수, 고도(DEM: Digital elevation model), 향(aspect), 경사도(slope), 음영기복도(hillshade)를 만들었다. Landsat 8으로부터 얻은 위성영상 4가지 변수, 1월 온도, 1월 정규식생지수, 4월 온도, 4월 정규식생지수를 사용하였으며, 노루의 활동, 서식, 은폐 등에 영향을 줄 것으로 고려되는 거리 변수 5가지, 경작지로부터의 거리, 숲 내부로의 거리, 도로로부터의 거리, 하천으로부터의 거리, 오름으로부터의 거리를 사용하였다. 또한 한남시험림 디지털임상도 자료에서 얻을 수 있는 임상변수 10가지를 사용하였다.

환경변수의 선정은 문헌조사 및 현장답사를 통해 이루어 졌으며, 고도, 향, 경사, 음영기복도 등의 지형적 인자(Anderson *et al.*, 2003; Pearson *et al.*, 2007; Guisan와 Zimmermann, 2000; Zaniwski *et al.*, 2002; Hernandez *et al.*, 2006; Phillips *et al.*, 2006; Stockwell와 Peterson, 2002; 권혁수 등, 2012; 김지연 등, 2012)와 거리변수(조해진 등, 2015; 김지연 등, 2012)를 선택하였으며, 선행연구에서 밝혀진 오름이 노루의 서식 및 번식에 긍정적 영향을 기여한다는 점을 고려하여 오름으로부터의 거리를 추가적으로 넣었다(오장근, 2015). 숲 내부의 특성을 보여주는 변수들은 활엽수림, 침엽수림, 혼효림 등이 노루에게 어떤 영향을 주는지 확인하기 위해 사용하였고(Danilkin and Hewison, 1996, 박용우, 2014), 뿐만 아니라 식생의 활력도를 파악할 수 있는 NDVI와 온도변수를 넣었으며, 임목 축적, 하층식생의 수, 우점종의 기여도 등 노루에 서식환경에 영향을 주는 것을 파악하기 위한 변수를 추가적으로 변수로 활용하였다.



Table 2. List of environmental variables

variable	description	Unit	variable	Description	unit
DEM	Digital elevation model	m	Yung	Age-class in digital forest type map	· (0~6)
Hillshade	Hillshade (azimuth180, altitude45)	· (0~255)	Imsang	Forest physiognomy in digital forest type map	· (0~6)
Aspect	aspect (0°~360°)	° (Degree)	Deciduous	Percent of Deciduous in digital forest type map	%
Slope	Slope (0°~90°)	° (Degree)	Evergreen	Percent of evergreen in digital forest type map	%
D_agri	Distance from agricultural area	m	Conifer	Percent of conifer in digital forest type map	%
D_forest-	Distance from forest boundary (to inside)	m	Hect_each	trees growing each in the forest	ha
D_road	Distance from road	m	Count_under	Number of lower vegetation in the area	·
D_stream	Distance from stream	m	TM_jan_band10	Thermal band of Landsat-8 image taken in January	° (Degree)
D_oreum	Distance from oreum	m	TM_apr_band10	Thermal band of Landsat-8 image taken in april	° (Degree)
Basin_map	Water classification using DEM	·	Jan_NDVI	Normalized Difference Vegetation Index in January	· (-1~1)
somildo	The ratio between the projected area of each tree in a given forest and the corresponding forest area	· (0~100)	Apr_NDVI	Normalized Difference Vegetation Index in April	· (-1~1)
Kung	Diameter-class in digital forest type map	· (0~4)			



3. 연구방법

(1) MaxEnt

MaxEnt는 불완전한 정보로부터 예측이나 추론을 시행하는 범용 기계 학습방법으로, Maximum Entropy라는 이론을 바탕으로 하여 개발된 프로그램이다. 이런 이론을 바탕으로 확률 분포를 찾아, 목표로 설정된 연구대상지의 확률 분포를 추정한다(Phillips *et al.*, 2006). 뿐만 아니라 모델 개발자가 설정한 환경 변수 및 제약 조건을 사용하여 확률 분포를 예측해 줄수 있게 해주며, 변수의 선택 및 비모수적인 관계를 표현할 수 있는 장점을 가지고 있다(Phillips *et al.*, 2006)

처음 Maximum Entropy가 소개된 것은 Berger, *et al* (1996)에 의해 “A Maximum Entropy Approach to Natural Language Processing”를 통하여 소개되었다. MaxEnt의 경우 출현자료(presence-only data)만을 사용하여 종 출현에 대한 모델링을 할 수 있는데, 이는 MaxEnt의 장점이자 다른 모델들과 차별화 된 점이다.

MaxEnt의 경우 과대평가된 결과를 보여주는 다른 모델들과 비교할 때 합리적인 결과를 도출해 줄 뿐만 아니라 적은 수의 출현 데이터로 오버 피팅을 최소화하는 효율적인 알고리즘으로 최상의 확률 분포를 보여준다. 비록 종 분포모형에 있어 비출현의 자료가 없이 분석된다는 불완전함이 있지만, 출현 데이터만을 가지고 분석하기에는 가장 적합한 모델이라고 설명할 수 있다.(Phillips *et al.*)

MaxEnt를 활용한 국외 및 국내 연구는 다양하게 수행되었다. Phillips *et al.* (2007)에 의해 “Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation” 이란 주제로 수행되었으며, 이보다 앞선 2006년에는 “Maximum entropy modeling of species geographic distributions”란 주제로 MaxEnt와 GARP 모델 뿐만 아니라 다른 종 분포모형

과 비교연구를 수행하였다.

국내에서도 MaxEnt를 이용한 종 분포모형이 많은 연구가 이루어 졌다. 이용호 등(2015)에 의해서는 기후변화 시나리오에 의한 외래식물 실패초의 서식지 분포 예측에 대한 연구가 수행되기도 하였으며, 전성우 등(2014)에 의해서는 MaxEnt를 이용하여 “생태계 서비스 가치평가를 위한 멸종위기 포유류의 종 분포 연구”라는 주제로 수행되었다. 조해진 등(2015)은 MaxEnt를 이용하여 참매의 서식지 예측에 대한 연구가 이루어졌으며, 송원경(2015)은 수원청개구리의 번식지 서식지 분석을 수행하였다. 이제민(2017)은 MaxEnt를 활용하여 삶의 행동권 연구를 수행하였다.

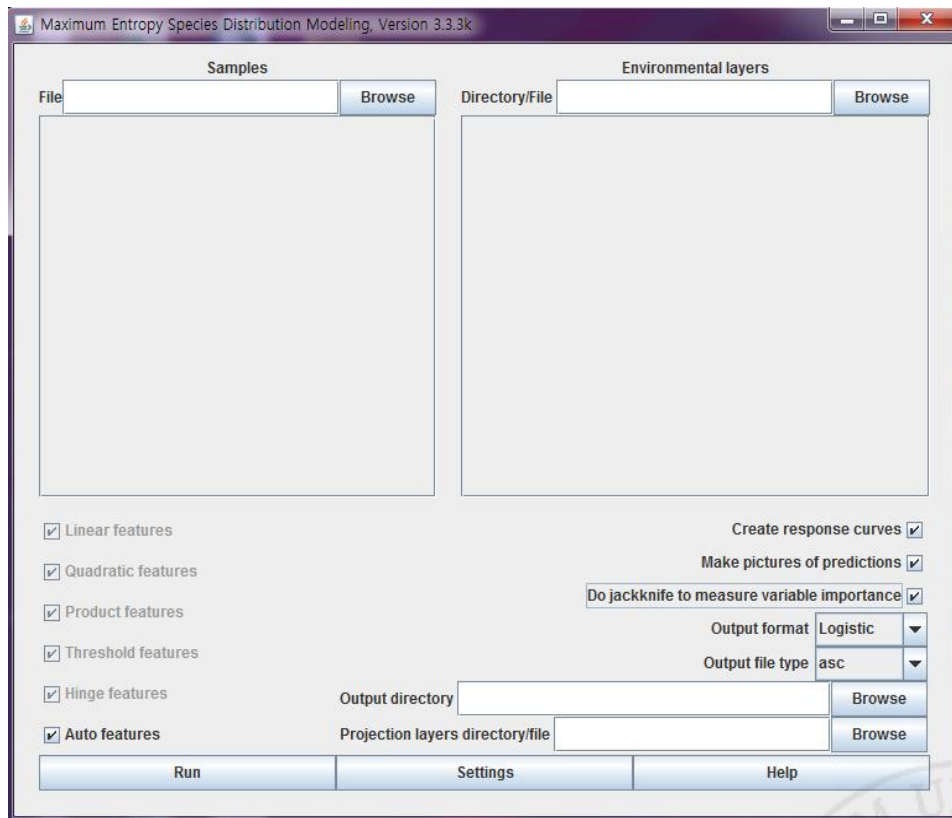


Fig. 4. MaxEnt 3.3.3k

MaxEnt의 특성을 이용하여 중 분포 모형이 아닌 자연재해에 대한 연구도 이루어졌다. 환경변수들과 특정 사건이 이루어진 것을 출현으로 이용한 데이터를 활용하였는데, 김호걸·이동근 등(2013)에 의해 산사태 발생가능성 예측에 대한 연구가 수행되었다. 김효민 등(2013)에 의해서는 MaxEnt를 이용한 서울시 도시홍수 적응능력 취약지역 선정에 관한 연구가 수행되기도 하였다.

본 연구는 2016년에 동일 대상종과 동일 대상지역에 대하여 시행된 선행연구와 비교하기 위해 선행연구에서 사용된 변수 14개를 포함하여 23개의 변수를 사용하였다. 25%의 test sample을 사용하고 100회 반복을 실시하였으며, 반복 샘플 세트가 평가에 활용되는 방식을 택하였다. 추가적으로 사용된 변수들을 포함한 분석으로 통하여 변수들의 기여도 변화, 예측지도의 변화 등을 확인하고자 한다.

MaxEnt를 이용해서 2가지 방법으로 분석을 실시하였는데, 첫 번째는 선행연구와 비교하기 위해, contribution이 낮은 변수를 제거하여 최종적인 변수를 선택하여 분석을 실시하였다. 이를 통해 최종적으로 선택되는 변수들의 비교하고, 변수의 추가적인 사용으로 인해 결과의 변화를 확인하였다. 최종적으로 선택된 변수를 이용하여 중 분포 모형을 구축하였다. 구축된 중 분포 모형의 경우, ROC curve의 AUC값이 유의미하다고 판단되는 0.8 이상의 값을 보여 모형의 적합도가 적절하다고 할 수 있다.

두 번째는 contribution이 낮은 변수들을 제거하지 않고 분석을 실시하여는데, 본 연구에서 시행되는 다른 모델들과 prediction 맵을 비교하기 위해서이다. 이 또한 0.8 이상의 값을 보여 모형의 적합도가 적절하다고 할 수 있다.



Table 3. Selection of the proper variables derived from 14 environmental factors in previous study

Hannam	AUC	variables included
14 variables	0.917	elevation, aspect, hillshade, slope, Jan_NDVI, Apr_NDVI, TM_Jan, TM_Apr, d_agri, d_stream, d_forest(-), d_road, d_oreum, yung
8 variables	0.909	aspect, hillshade, slope, Jan_NDVI, Apr_NDVI, d_stream, d_forest(-), yung
6 variables	0.908	aspect, hillshade, slope, d_stream, d_forest(-), yung

Table 4. Selection of the proper variables derived from 23 environmental factors in this study

variable	AUC	variables included
23 variables	0.897	dem, aspect, hillshade, slope, Jan_NDVI, Apr_NDVI, TM_Jan, TM_Apr, d_agri, d_stream, d_forest(-), d_road, d_oreum, yung, kung, hect_each, imsang, conifer, evergreen, deciduous, somildo, count_under, basin_map
8 variables	0.901	dem, d_oreum, d_forest(-), d_stream, d_agriculture, d_road, somildo, basin_map



(2) 가우시안 혼합 모델(GMM)

가우시안 혼합 모델(GMM)은 주어진 데이터의 집합 분산 분포 밀도를 단 하나의 확률밀도 함수로 모델링 하는 방법을 개선한 밀도 추정방법으로 평균과 분산의 파라미터를 가지는 여러 개의 가우시안 함수로 데이터의 분포를 모델링 하는 방법이다(전미진, 2012). GMM은 밀도 분포의 매끄러운 모양을 근사시킬 수 있기 때문에 최근에는 좋은 결과를 얻기 위해 이미지 처리에서 자주 사용하고 있는 방법이다(Abhishek, 2013).

GMM의 특성을 이용한 연구는 많은 분야에서 활용되고 있다. 좌뇌 MRI 사진을 혼합시키기 위해 GMM을 사용한 사례도 있으며(Huaifer *et al.*, 2013), 뇌의 MRI 사진 분할을 위해서 GMM이 사용되기도 하였다(Zexuan *et al.*, 2015). 움직이는 물체를 추적하기 위해 GMM을 사용한 사례도 있으며(Ahhishek and Prashant, 2013), 온라인 예측 및 프로세서 모니터링을 위한 적응형 소프트웨어 개발에 GMM을 이용하기도 하였다(Ratko *et al.*, 2013). 최근 한국에서는 GMM을 이용하여 “손목 착용형 웨어러블 기기의 가속도 센서를 사용한 사용자 인증”이란 주제로 연구를 하였으며(김용관과 문종섭, 2017), GMM의 특성을 이용하여 한국 신생아 출생체중 데이터 보정이란 연구도 이루어 졌다(신형식, 2017).

본 연구에서는 주어진 데이터의 집합 분산 분포 밀도를 단 하나의 확률밀도 함수로 모델링 하는 방법을 개선한 밀도 추정방법으로 평균과 분산의 파라미터를 가지는 여러 개의 가우시안 함수로 데이터의 분포를 모델링하는 방법에 착안하여 종 분포 모형 연구에 적용해보기로 하였다. GMM을 바탕으로 다차원의 변수를 분석하여 기계학습 후 확률을 표현해주는 원리를 이용하여 Matlab을 사용하였다. 출현데이터에서 추출한 환경변수들을 GMM을 이용하여 기계학습을 시행한 후, 추정된 확률을 비출현지역에 적용시켜 출현확률을 표현하고자 하였다. 사용된 변수는 MaxEnt와 동일하게 23가지를 사용하였으며, 전체 변수를

사용하여 나온 결과와 변수의 특성에 따른 예측결과를 확인하기 위해 변수들의 특징(parameter feature)으로 구분하여 결과를 제시하였다.

또한, PCA(Principle component analysis)를 이용하여 사용된 변수들을 축소하였다. PCA는 변수들간의 상관관계 및 유사도 등을 분석하여 고차원의 많은 데이터를 저차원으로 줄이는 기술로 ArcGis의 principal componet tool을 이용하여 분석하였다. 분석 결과 중 Percent of EigenValues와 Accumulative of EigenValues가 높은 상위 5개를 선택하였고, 이를 이용하여 추가적으로 GMM 분석에 활용하였다.

(3) 로지스틱 회귀

로지스틱 회귀(Logistic regression)는 종종 maximum-entropy로 문헌에서 분류되며 기계 학습에 사용되기도 하며(Berger et al. 1996) 종 분포 모형에 사용되기도 한다(Phillips et al. 2006). 로지스틱 회귀와 같은 선형방법이 최근에 들어 종 분포 모형에 많이 활용되고 있는 실정이다(Osborne and Tigar, 1992; Green et al., 1994; Austin and Meyers, 1996). 로지스틱 회귀는 서식지 적합성 분석으로 많이 사용되고 있는데 연속형 자료와 범주형 자료를 다 포함시킬 수 있는 장점을 가지고 있으며, 자료의 통계적 제약에 덜 영향받고 출현 및 비출현 자료를 이용할 수 있다는 점에서 유용하다(송원경 등, 2013).

로지스틱 회귀의 장점들을 고려하여 노루의 종 분포 모형을 만들기 위해 로지스틱 회귀를 사용하였으며, 환경변수의 유의성과 검증은 불필요한 변수를 단계적으로 선택과 제거를 실시하였다. 뿐만 아니라 출현빈도를 종속변수로 설정하여, 노루 출현에 대해 방형구를 10m x 10m 지정하고, 그에 따른 빈도를 도출하였다. 도출된 출현빈도를 이용하여 threshold를 설정하였는데, 출현자료를 방형구 내에 나타나는 빈도에 따라 출현과 비출현으로 설정하였다(Table 4). 분석결과로 나온 모델들의 적합성을 판단하기 위해 AIC와 BIC를 이용하였다. AIC와 BIC는 값이 낮을수록 모델이 적합하다고 판단하고 있는데, 5가지의 빈

도를 설정하여 각 빈도마다 노루의 출현에 대해 표현해주는 것을 비교하였다.

Table 5. Set data according to frequency

Data	Explanation
1	Set the number of occurrences in the cell to "presence" more than 2 times
2	Set the number of occurrences in the cell to "presence" more that 3 times
3	Set the number of occurrences in the cell to "presence" more that 4 times
4	Set the number of occurrences in the cell to "presence" more that 5 times
5	Set the number of occurrences in the cell to "presence" more that 6 times



IV. 결과 및 고찰

1. 한남시험림 환경 변수

한남시험림에 서식하는 노루의 서식지 모델링을 위해 총 23개 변수가 사용되었고 각 환경변수에 대한 설명은 다음과 같다.

(1) 고도(DEM: Digital elevation model)

한남시험림의 고도는 약 320m~770m에 위치하고 있으며, 고도가 높은 지역으로 갈수록 노란색으로 나타나고 있다. 제주도의 산간 지방에 서식하는 것으로 알려진 노루의 경우, 인간의 활동 및 거주, 생활환경 등 고도에서 영향을 받을 것으로 고려하여 사용하였다(Fig. 6a).

(2) 음영기복도(hillshade)

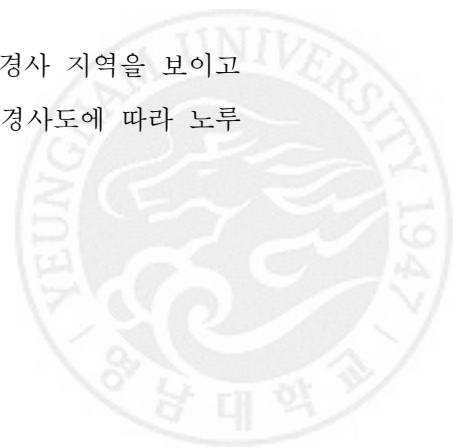
음영기복도는 0~249의 범위를 보여주고 있으며, 값이 높은 지역을 흰색, 낮은 지역을 검은색으로 표현하였다. 남쪽에서 45°의 각도로 햇빛이 비춘다고 가정하였을 때의 변수이며, 밝은 색일수록 빛이 많이 들어와 밝고, 값이 낮을수록 그늘이 저 어두운 곳을 나타내고 있다(Fig. 6b).

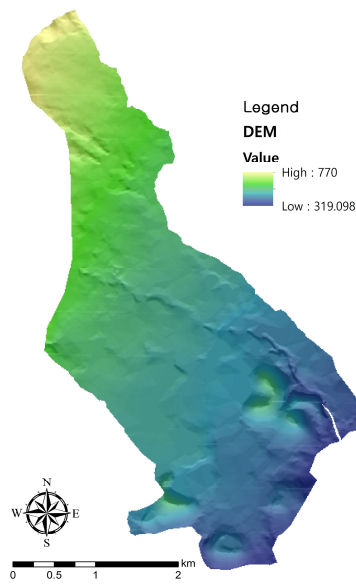
(3) 향(aspect)

향은 경사면이 향하는 수평의 방향을 나타낸 것으로 방향에 따라 색을 지정하였으며, 평지는 가장 어두운색을 보이며, 1°부터 시계방향으로 45° 간격으로 밝아지는 것으로 해석하였다.

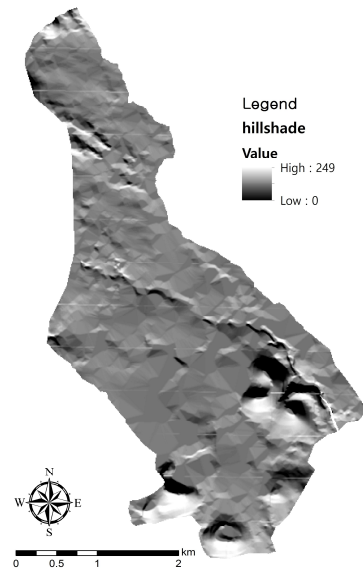
(4) 경사도(slope)

경사도는 0°~52°의 범위가 분포하고 있다. 대부분 낮은 경사 지역을 보이고 있으며, 오름 인근 지역에서 높은 경사도를 보여주고 있다. 경사도에 따라 노루의 활동에 영향을 줄 것으로 판단되어 변수로 활용하였다.



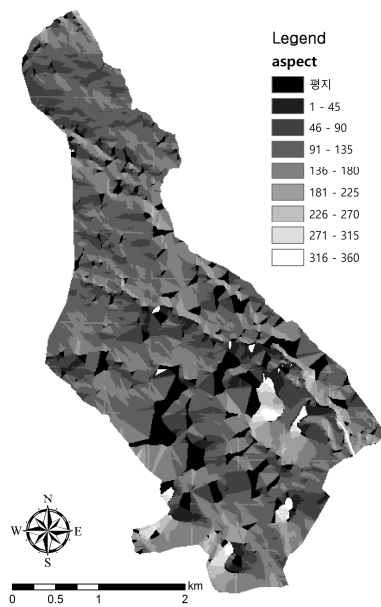


a. DEM

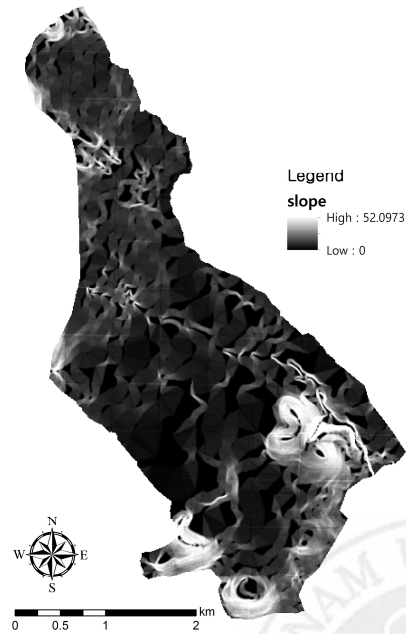


b. hillshade

Fig. 5. Environmental variables of DEM and hillshade in HEF



a. aspect



b. slope

Fig. 6. Environmental variables of easpcet and slope in HEF

(5) 경작지로부터의 거리(D_agriculture)

경작지로부터의 거리는 0m~1785m의 범위를 가지며, 한남시험의 북쪽에 해충원이 위치하고 있으며, 한남시험림을 벗어난 동쪽 및 남쪽에 농경지가 위치한다. 갈색으로 가까울수록 경작지에 가까운 것을 의미하며, 청록색에 가까울수록 경작지로부터 먼 곳을 의미한다(Fig. 7a).

(6) 숲 내부로의 거리(D_forset(-))

숲 내부로의 거리는 -3824m~0m의 범위를 가지고 있으며, 음수의 값으로 갈수록 숲 깊숙한 곳을 의미한다. 북쪽으로 갈수록 초록색이 나타나고 있는 것으로 보아, 숲 경계선으로부터 떨어져 있는 깊숙한 숲을 나타내고 있다. 연구 대상지인 한남시험림의 경우 대상지 전체가 숲으로 구성되어 있어, 숲 외부로의 거리는 분석에 활용할 수가 없고, 숲 내부로의 거리에 따라 노루의 활동 및 출현에 영향을 줄 것으로 판단되어 활용하였다(Fig. 7b).

(7) 도로로부터의 거리(D_road)

도로로부터의 거리는 177m~ 2,931m의 범위를 가지고 있으며, 파란색에 가까울수록 도로로 근접하며, 색이 흰색에 가까울수록 도로로부터 먼 곳을 의미하고 있다. 도로에 따라 인간의 접근으로 인한 교란, 차량 주행에 따른 로드킬, 등 인위적인 간섭이 발생하는 원인으로 판단하여 분석에 활용하였다(Fig. 8a).

(8) 하천으로부터의 거리(D_stream)

하천으로부터의 거리는 0m~1198m 범위를 가지고 있으며, 한남 시험림을 가르는 3개의 하천이 분포하고 있다. 색이 진할수록 하천에 인접함을 의미하며, 흰색에 가까울수록 하천으로부터 떨어져있음을 의미하고 있다. 시험림의 하천의 경우 대부분 건천인 것으로 파악되나, 우천시 혹은 노루의 취음에 대한 부분을 파악하기 위해 변수로 활용하였다(Fig. 8b)



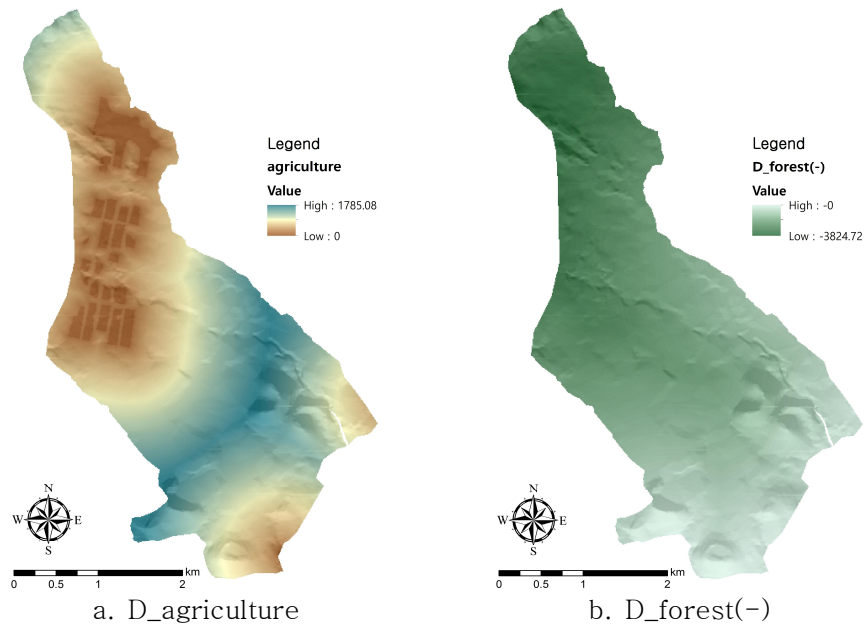


Fig. 7. Environmental variables of $D_{agriculture}$ and $D_{forest(-)}$ in HEF

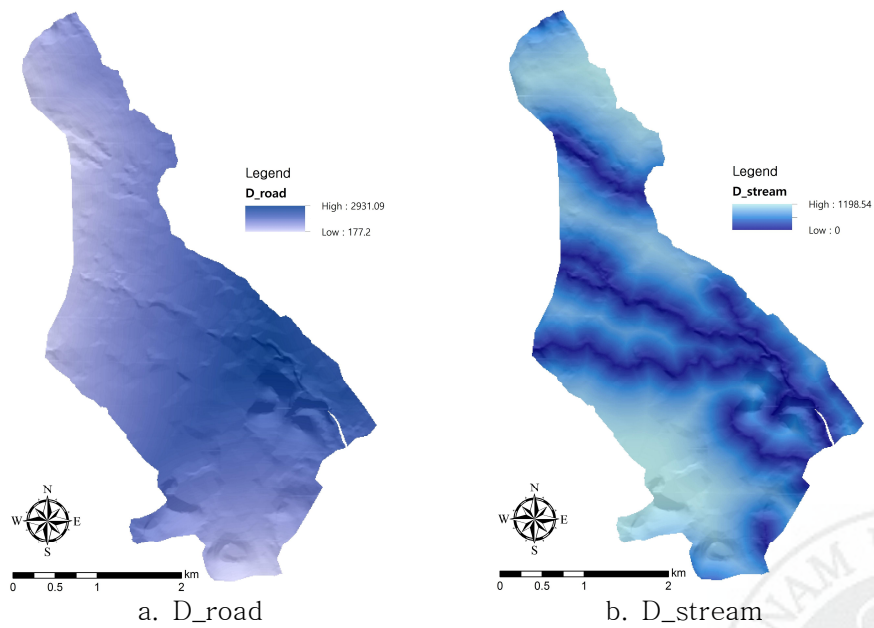


Fig. 8. Environmental variables of D_{road} and D_{stream} in HEF

(9) 오름으로부터의 거리(D_oreum)

오름은 제주도에서 볼 수 있는 독특한 지형으로 기생화산이 형성한 지형이다. 제주도 내에 고라니 서식과 관련이 있는 것으로 판단하여 변수로 지정하였으며, 0m~ 1611m의 범위를 가지고 있다. 붉은 색에 가까울수록 오름과 가까움의 의미하며, 흰색에 가까울수록 오름과 먼 지역을 의미하고 있다.

(10) 소유역권(basin_map)

AirGis의 basin tool로 만든 소유역권 지도로, 고도, 향, 경사도 등을 고려하여 수계의 흐름을 판단하고, 유사한 값과 흐름의 방향을 고려하여 만들었다. 노루의 출현 및 활동이 유역권에 영향을 받는 것을 파악하기 위해 사용하였으며, 색깔에 따라 1차에서 4차까지 구분하였다.

(11) 소밀도(somildo)

소밀도는 일정한 임지에 각 수관을 투영한 면적과 해당 산림면적과의 비를 나타내는 것으로 울폐도 또는 폐쇄도라고도 칭한다. 한남시험림 내 수관이 임분을 덮고 있는 정도를 파악하기 위해 사용하였으며, 소밀도에 따라 노루의 활동이 영향을 받는 것을 파악하기 위해 사용하였다(Fig. 10a)

(12) 경급(Kung)

경급은 임목의 흉고직경의 크기에 따라 나눈 것으로 치수는 흉고직경 6cm미만의 임목이 50%이상 생육하는 임분이고, 소경목은 흉고직경 6~16cm의 임목이 50%이상 생육하는 임분이다. 중경목은 흉고직경 18~28cm의 임목이 50%이상 생육하는 임분이며, 대경목은 흉고직경 30cm이상의 임목이 생육하는 임분이다. 본 연구에서는 경급에 따라 노루의 활동에 영향을 주는 것을 파악하기 위해 사용하였다(Fig. 10b)



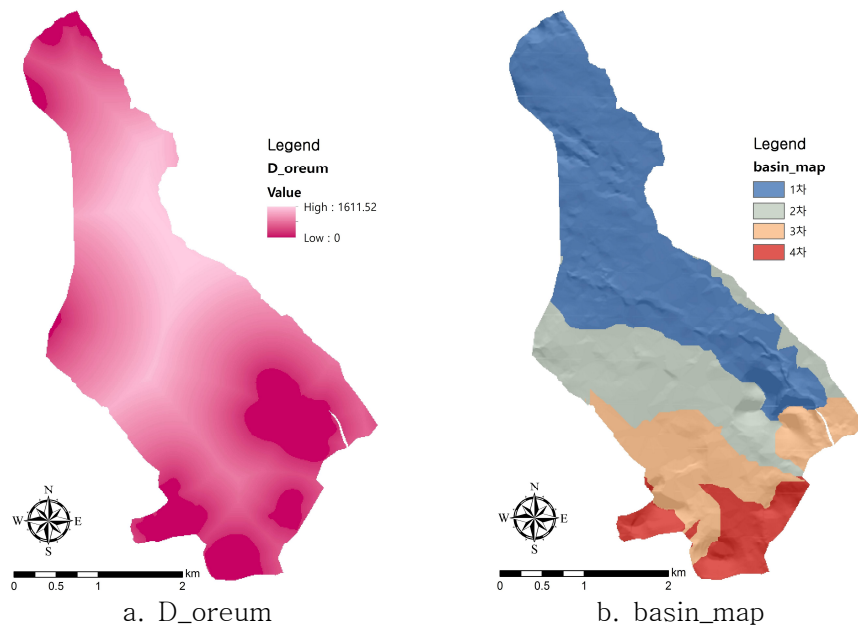


Fig. 9. Environmental variables of D_oreum and basin_map in HEF

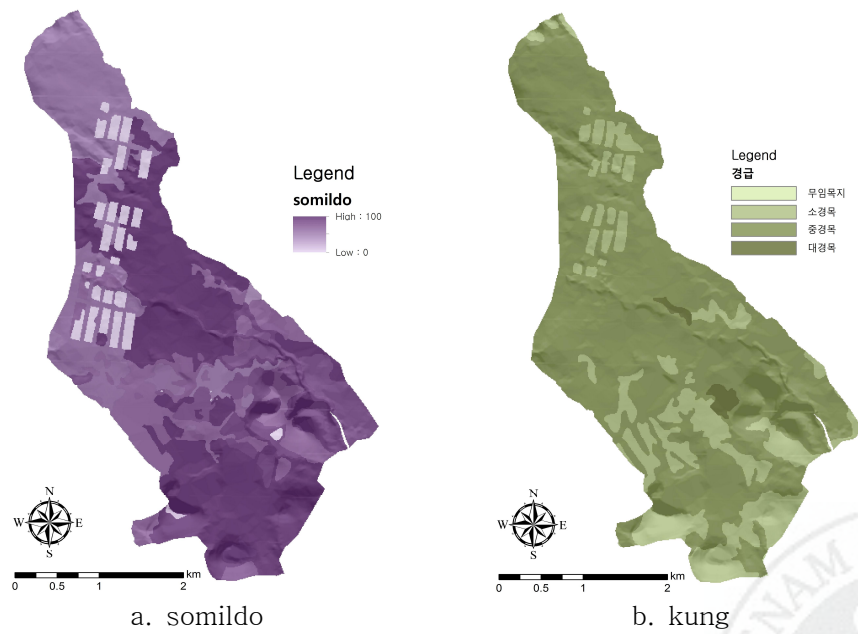


Fig. 10. Environmental variables of somildo and kung in HEF

(13) 영급(yung)

영급은 수목의 임령을 간격지어 표현한 것으로 10년 단위로 영급을 나누고 있다. 1~10년까지를 1영급을 시작해서 10년 간격으로 영급을 나누고 있는데, 본 연구에서는 수목의 영급이 노루의 활동이 영향을 주는 것을 파악하기 위해 사용하였다(Fig. 11a).

(14) 임상(Imsang)

임상의 구분은 수종에 의하여 침엽수림, 활엽수림, 혼효림, 죽림 등으로 구분한다. 본 연구에서는 임상의 구분을 세분화 하여 삼나무림, 상록활엽수림, 침활혼효림, 침활혼효림(낙엽), 침활혼효림(상록), 혼합활엽수림, 채종림 등으로 구별하였다. 임상에 따라 노루의 활동이 영향을 받는 것을 파악하기 위해 사용하였으며, 한남시험림 임소반도 자료에서 추출하였다(Fig. 11b).

(15) 낙엽활엽수(deciduous)

낙엽활엽수 변수는 한남시험림 내에 분포하고 있는 낙엽활엽수의 비중을 나타낸 것으로 색이 진할수록 그 비중은 높으며, 색이 연할수록 비중이 낮다. 낙엽활엽수는 대표적으로 신갈나무, 상수리나무, 갈참나무, 버드나무 등이 있으며 잎이 넓고 가을이면 낙엽이 진다. 낙엽활엽수의 분포와 비중이 노루의 활동에 영향을 주는 것을 파악하기 위해 사용하였다(Fig. 12a)

(16) 상록활엽수(evergreen)

상록활엽수 변수는 한남시험림 내에 분포하고 있는 낙엽활엽수의 비중을 나타낸 것으로 색이 진할수록 그 비중이 높으며, 색이 연할수록 비중이 낮다. 상록활엽수는 대표적으로 구상나무, 소나무, 향나무 등이 있으며, 잎의 모양이 침형으로 뾰족한 형태를 띠고 있다. 상록활엽수의 분포와 비중이 노루의 활동에 영향을 주는 것을 파악하기 위해 사용하였다(Fig. 12b)



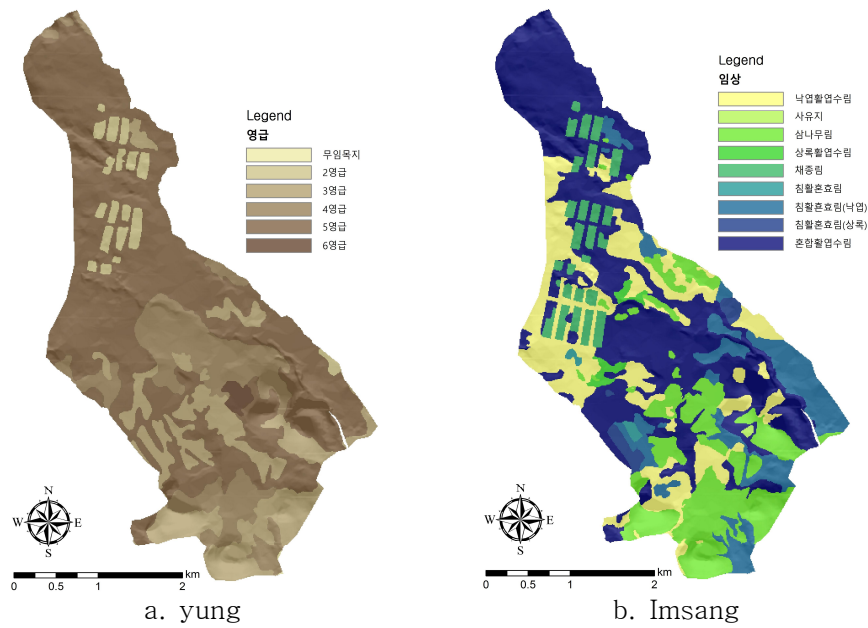


Fig. 11. Environmental variables of yung and Imsang in HEF

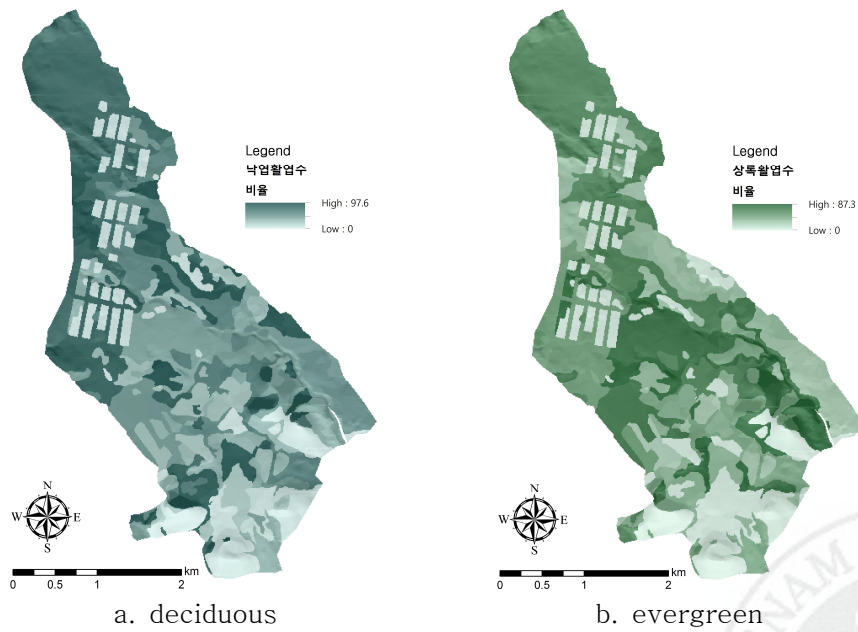


Fig. 12. Environmental variables of deciduous and evergreen in HEF

(17) 침엽수(conifer)

침엽수 변수는 한남시험림 내에 분포하고 있는 침엽수의 비중을 나타낸 것으로 색이 진할수록 그 비중이 높으며, 색이 연할수록 그 비중이 낮은 것으로 나타난다. 침엽수는 대표적으로 소나무, 메타세콰이어, 잎갈나무, 개잎갈나무 등이 있으며, 잎이 대개 바늘같이 뾰족한 형태를 띠고 있다. 침엽수의 분포와 비중이 노루의 활동에 영향을 주는 것을 파악하기 위해 사용하였다(Fig. 13b)

(18) Hect_each

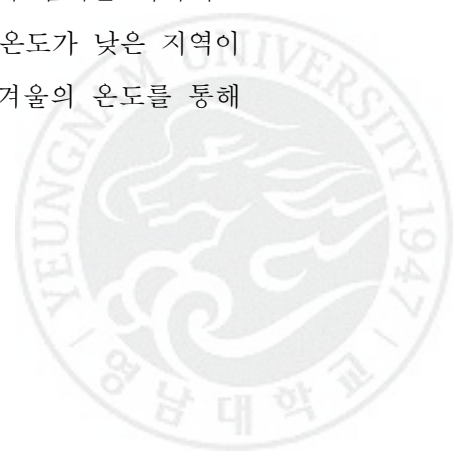
Hect_each는 단위 면적당 임목이 차지하고 있는 축적을 나타내는 변수이다. 한남시험림 내에 분포하고 있는 지역마다 임목이 차지하고 있는 축적이 각기 다르다. 임목 축적이 높을수록 색은 초록색에 가까우며, 임목 축적이 낮을수록 갈색에 가깝게 나타난다. 임목 축적이 높을수록 시야의 확보가 어려운 점이 있어, 노루는 은·엄폐를 이용한 서식활동을 할 것으로 판단된다. 이를 공간분석개념에서 접근하여 확인하기 위해 변수로 활용하였다.

(19) count_under

count_under는 한남시험림내에 분포하고 있는 우점종이 아닌, 하층식생이 분포하고 있는 것을 표현한 것으로 색이 진할수록 하층식생의 종 수가 많은 것으로 표현하고 있다. 노루의 경우 시야에 들어오는 대부분의 하층식생과 나무순, 나뭇잎, 나무줄기 등을 먹는 초식성을 보이고 있다. 다양한 먹이원이 노루의 활동에 영향을 주는 것을 확인하기 위해 변수로 사용하였다

(20) TM_Jan

TM_Jan은 1월의 온도를 나타낸 것으로 약 7.9℃~14℃의 범위를 나타내고 있다. 색이 붉을수록 온도가 높은 지역이고 초록색 일수록 온도가 낮은 지역이다. 식생의 성장이 가장 낮고 동물들의 활동이 가장 적은 겨울의 온도를 통해 노루의 활동을 확인하기 위해 변수로 사용하였다.



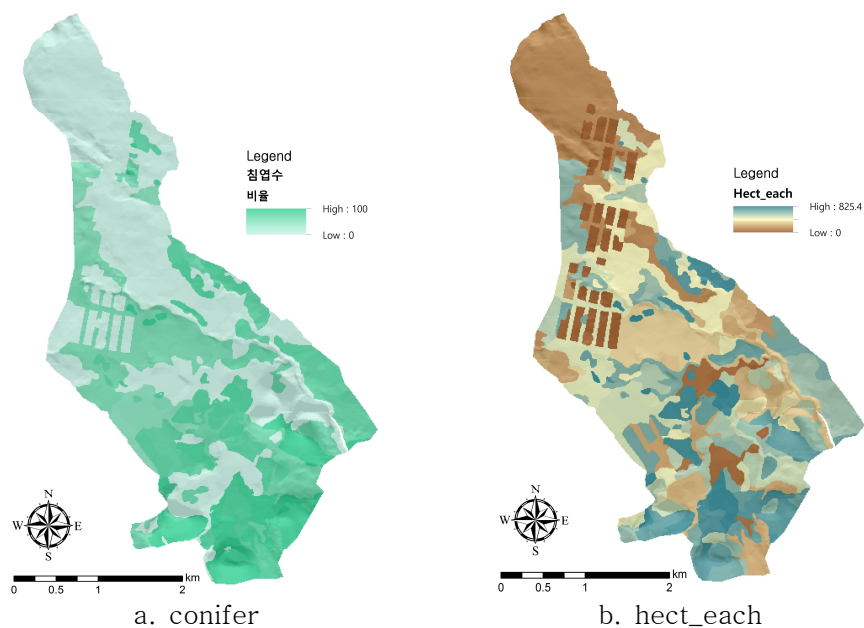


Fig. 13. Environmental variables of conifer and hect_each in HEF

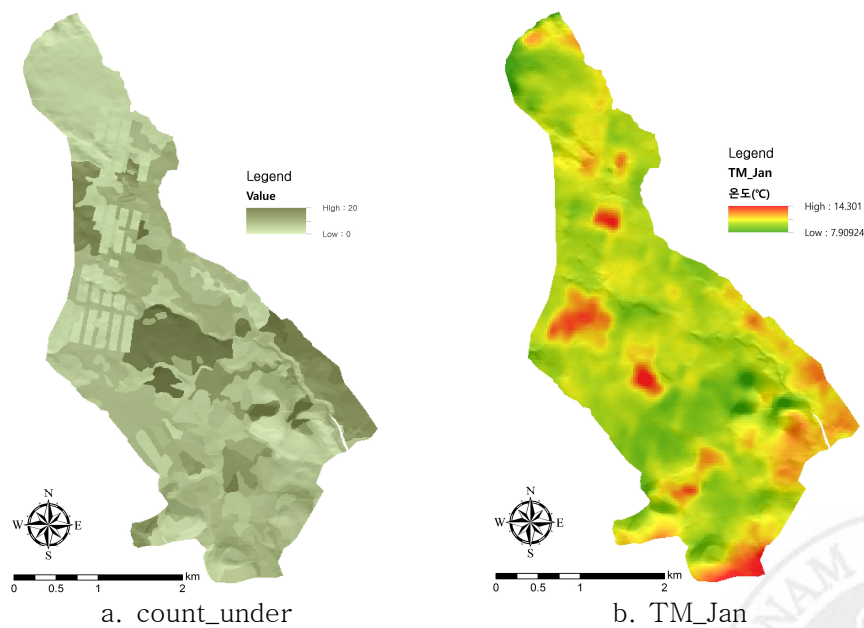


Fig. 14. Environmental variables of count_under and TM_Jan in HEF

(21) TM_Apr

TM_Apr은 4월의 온도를 나타낸 것으로 약 $14.5^{\circ}\text{C}\sim 22.4^{\circ}\text{C}$ 의 범위를 나타내고 있다. 색이 붉을수록 온도가 높은 지역이고 초록색 일수록 온도가 낮은 지역이다. 식생의 성장이 시작하고 동물이 동면기에서 깨어나 본격적인 활동을 보이는 봄의 온도를 통해 노루의 활동을 확인하기 위해 변수로 사용하였다.

(22) Jan_NDVI

Jan_NDVI는 1월의 정규식생지수를 나타낸 것으로 약 $-0.09\sim 0.04$ 의 값을 나타내고 있다. 색이 어두울수록 식생이 우거진 지역이며, 밝을수록 그 반대의 지역이다. 1월의 한남시험림의 경우 남쪽으로 내려올수록 숲이 우거져 있음을 NDVI 지수로 확인할 수 있다. 식생이 우거짐에 따라 노루의 활동에 영향을 주는 것을 확인하기 위해 변수로 활용하였다.

(23) Apr_NDVI

Apr_NDVI는 4월의 정규식생지수를 나타낸 것으로 약 $-0.04\sim 0.09$ 의 값을 나타내고 있다. 색이 어두울수록 식생이 우거진 지역이며, 밝을수록 그 반대의 지역이다. 4월의 한남시험림의 경우 북쪽 일부지대를 제외하곤 숲이 우거져 있음을 NDVI 지수로 확인할 수 있다. 식생이 우거짐에 따라 노루의 활동에 영향을 주는 것을 확인하기 위해 변수로 활용하였다.



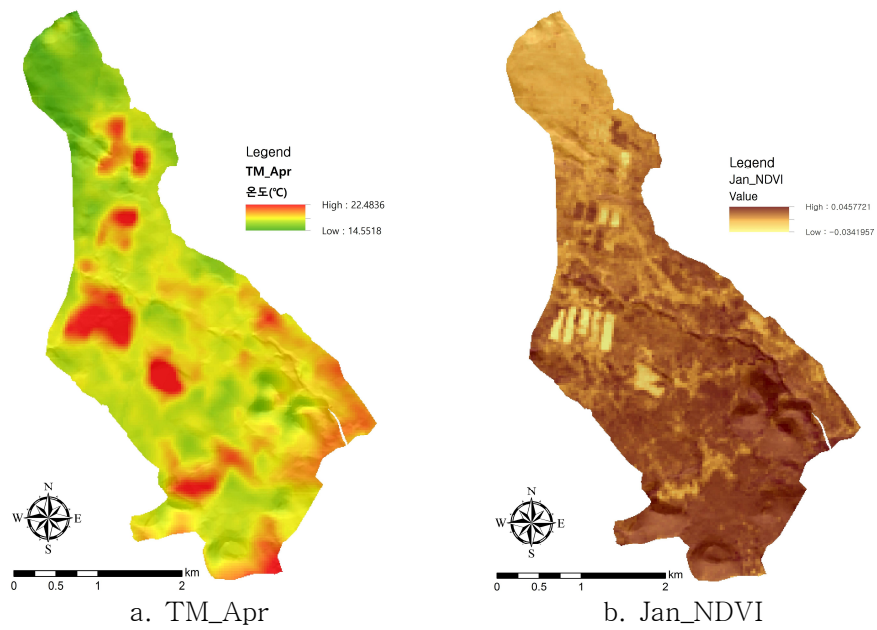


Fig. 15. Environmental variables of TM_Apr and Jan_NDVI in HEF

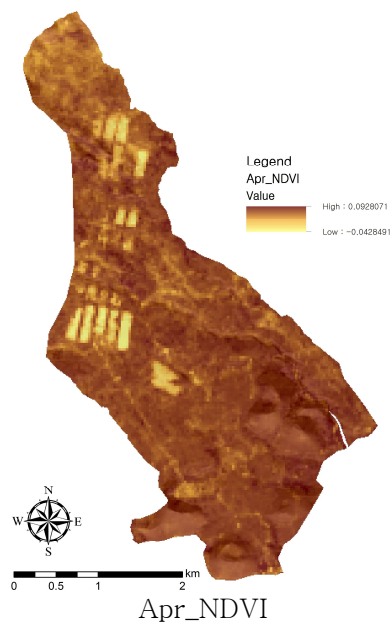


Fig. 16. Environmental variables of Apr_NDVI in HEF



2. MaxEnt 모델링 결과

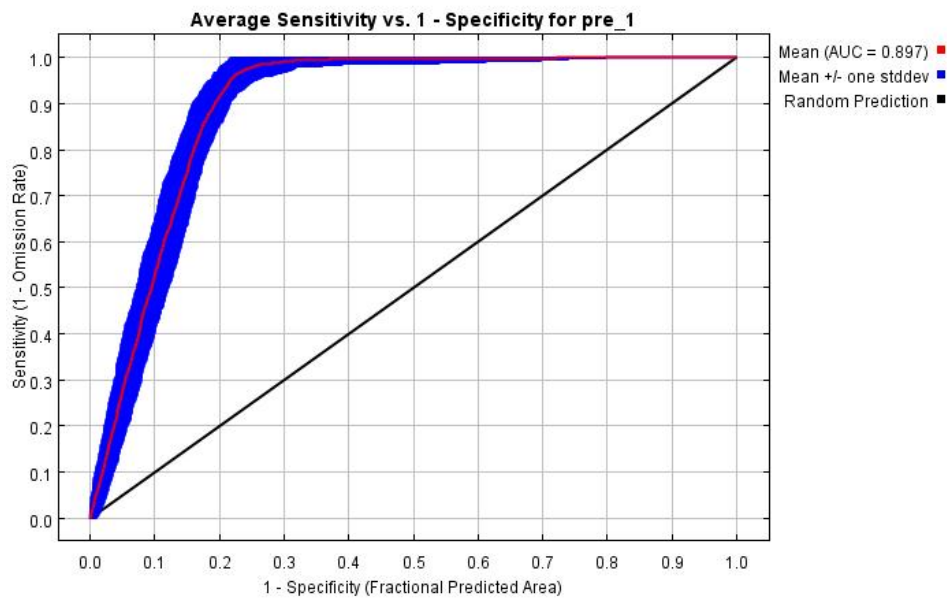
(1) ROC curve & AUC

ROC(Receiver Operating Characteristic) 분석은 모델 수행 능력 평가에 주로 사용되고 있다(Mas *et al.*, 2013). 최근에는 기계학습이나 데이터 마이닝(data mining)에서 수행 능력 평가를 위해 사용이 증가되고 있는 추세이다(Fawcett, 2006).

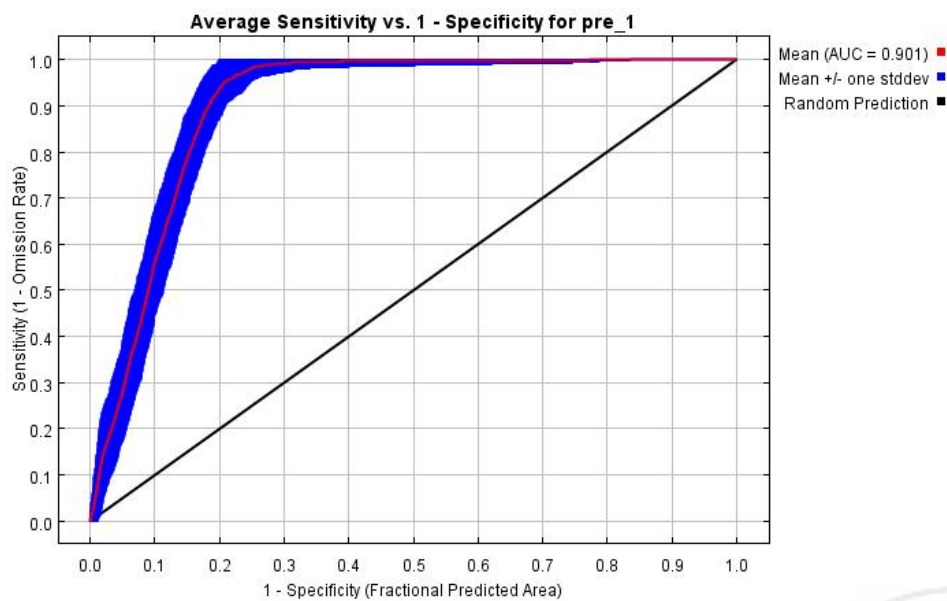
AUC는 무작위 선택된 출현지점과 무작위로 선택된 비출현지점을 평가하였을 때, 출현지점이 높게 평가될 확률을 말하고 있는데, MaxEnt의 경우 출현좌표만 사용하지만 가상의 비 출현좌표(pseudo-absence data)를 이용하여 AUC값을 측정하고 있다(Phillips *et al.* 2008). (0,0)과 (1,1) 사이의 면적을 가져 0.5 이상의 값을 가지는데, 사각형의 면적이 1인 것을 감안하면, AUC값은 0.5~1.0 사이의 값을 가지게 되고, 1에 가까울수록 모형의 설명력 혹은 적합성이 높다고 평가한다(Fawcett, 2006).

본 연구에서는 ROC curve와 AUC값으로 모델의 적합성을 평가하였는데, 보통 0.8 이상의 값을 가질 때, 모델의 적합성과 예측력은 우수하다고 평가하고 있는 점을 고려해서 평가하였다(박현철 등, 2014, 재인용). 모든 변수를 넣어서 분석을 하였을 때 AUC값은 0.897로 적합성과 예측력이 우수하다고 평가되는 0.8을 넘는 값을 보이고 있다. 이는 사용된 변수들로 설명된 모델이 잘 설명되고 있음을 보여주고 있다. 뿐만 아니라 선행 연구(김아름, 2016)와 비교하기 위해 percent contribution의 값을 기준으로 1 이하인 변수를 삭제 후, 분석을 돌린 결과의 AUC 값은 0.901로 나왔다. 이 또한 적합성과 예측력이 우수하다고 평가되는 0.8을 넘는 값을 보여주고 있다.





a. AUC curve using all variable



b. AUC curve using summarize variable

Fig. 17. ROC curve in HEF



(2) Percent contribution 과 Permutation importance

MaxEnt를 이용한 분석에서 percent contribution과 permutation importance 값이 도출되는데 이 값을 이용하여 변수들의 중요성을 추정할 수 있다(Songer *et al.*, 2012). percent contribution의 경우 해당 변수가 모델을 구축하는데 있어 얼마만큼 기여도를 가지는 지를 설명해주며(songer *et al.*, 2012), Permutation importance는 변수의 기여도를 나타내는 다른 지표로서 최종 모형의 결과에만 의존하며, permutation은 각 환경변수의 차례에서 그 환경변수의 값에 training presence와 background data의 값이 무작위로 변경되는 것을 말하고 이 자료를 통해 모형을 재평가하였을 때 AUC값을 얼마나 떨어뜨리는지를 percentage로 나타낸 값이다(Quillfeldt *et al.*, 2013; 김아름, 2016, 재인용).

한남시험림에 대한 변수의 기여도는 (Table 5.)에서 확인할 수 있다. 선행연구(Table 6.)과 비교하였을 때 차이가 보이는 변수가 존재하는데, 선행연구에서 높은 값을 보인 D_forest(-)의 경우 56.5에서 7.4로 떨어졌으며, D_stream의 경우 35.1에서 5.6으로 떨어진 것을 확인할 수 있다. 그 외 다른 변수들의 경우 금회 연구에서는 기여도가 낮아 삭제된 것을 확인할 수 있다. 본 연구에서는 Basin_map(57.9), D_oreum(14), DEM(10.7), somildo(3.5), D_agriculture(0.8), D_roead(0.1)로 추가된 것을 확인할 수 있다. 선행연구에서 사용되지 않았던 Basin_map과 somildo의 경우 추가적으로 percent contribution의 값을 보이고 있다.

Permutation importance의 경우도 Percent contribution의 경우와 마찬가지로 변화가 있었다. 기존에 높은 값을 보였던 D_forset(-)의 경우 71.6에서 26.4로 떨어졌고, D_stream의 경우 25.0에서 5.1로 떨어졌다. 금회 연구에서 나온 permutation importance를 통한 변수의 기여도를 살펴보면, Basin_map, D_oreum, DEM, D_forest(-), D_stream, somildo, D_agriculture, D_road의 순으로 기여도가 높게 나왔다.

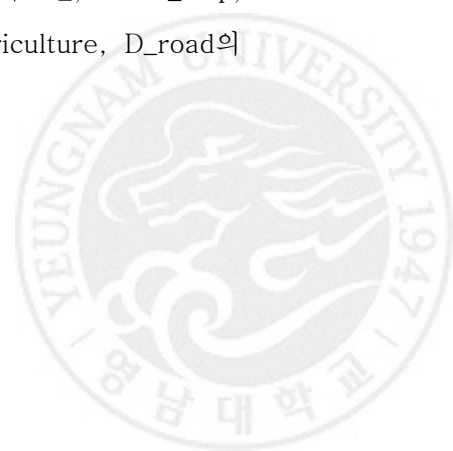


Table 6. Result of Percent contribution and Permutation importance previous study

Variable	Percent contribution	Permutation importance
d_forest(-)	56.5	71.6
d_stream	35.1	25.0
aspect	3.6	1.6
slope	2.7	0.8
hillshade	1.3	0.8
yung	0.7	0.2

Table 7. Result of Percent contribution and Permutation importance in this study

variable	Percent contribution	Permutation importance
Basin_map	57.9	26.4
D_oreum	14	21
DEM	10.7	18.8
D_forest(-)	7.4	20.1
D_stream	5.6	5.1
somildo	3.5	1.9
D_agriculture	0.8	4.9
D_road	0.1	1.8



(3) Jackknife test

MaxEnt 결과에 대한 기여도 측정에 대한 또 다른 방법은 Jackknife test가 있는데, 이 또한 결과로 확인할 수 있다. 각 변수들이 그 변수 1개만 모델을 만들었을 때, 사용된 변수가 개별적으로 모델을 만드는데 있어 기여도를 얼마나 가지는 지 확인할 수 있다(파랑). 또한 그 변수만을 제거하고 모형을 만들었을 때, 제거된 변수가 모델에 영향을 얼마나 가지는지도 확인할 수 있으며(청록), 전체 변수를 사용하였을 때의 기여도를 확인할 수 있다(빨강).

Fig. 18.의 결과를 보면 Basinmap의 경우가 가장 파란 막대를 보여주고 있으며, 그 다음으로 D_forest(-), DEM이 높은 것을 보여주고 있다. D_oreum, D_stream, somildo, D_road, D_agriculture가 그 다음 순으로 모델의 기여도를 보여주고 있다.

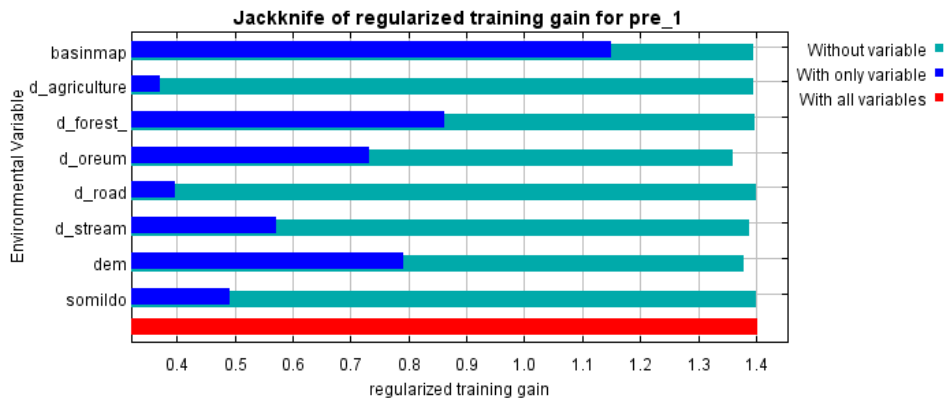


Fig. 18. Jackknife of regularized training gain



(4) 예측지도

Fig. 19.은 한남시험림을 대상으로 실시한 MaxEnt 분석의 예측지도 이다. 출현확률은 0~1사이의 값으로 표현하고 있으며, 그 값이 높을수록 출현확률이 높음을 나타낸다.

출현확률은 0~0.643을 보이고 있으며, 붉은색에 가까울수록 출현확률이 높으며, 파란색에 가까울수록 출현확이 낮음을 나타내고 있다. 출현예측지도를 보면 한남시험림 전체에 비해 남쪽 일부지역에 집중되어 출현확률이 높은 것을 확인할 수 있다. 이는 선행에 시행된 결과(Fig. 20.)과 유사한 모습을 보이고 있다. 세부적으로는 출현지역에 대해 그 값과 지역에 대해서 다른 일부 구간이 보이거나, 전체적인 결과로 보았을 때는 비슷한 경향을 보이고 있다. 비록 분석에 사용된 변수가 추가되고, 각 변수들이 분석에 기여한 정도가 각기 다름에도 불구하고, 그 결과는 비슷함을 알 수 있다. 이는 선형연구와 마찬가지로 presence-only data로 사용된 좌표들을 사용하고, 그 좌표들이 위치하는 것이 편향된 결과를 보여준 것으로 판단된다. 일부 지역에 편향된 자료를 사용하는 것은 다른 변수를 사용하여 그 변수들이 다른 기여도를 보이지만, 결과에는 크게 영향을 주지 않는 것을 확인하였다.



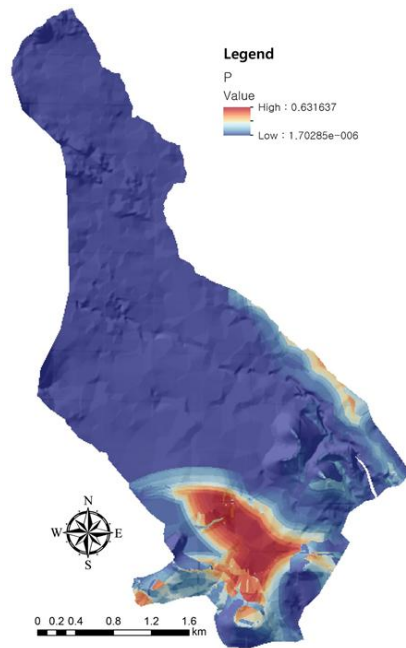


Fig. 19. Previous study prediction map of roe deers in HEF

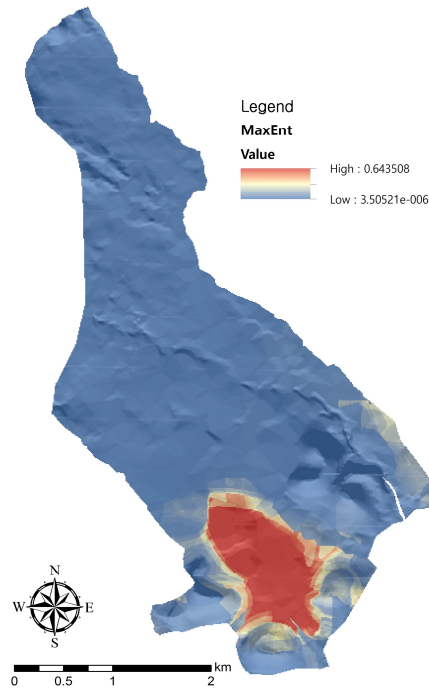


Fig. 20. In this study prediction map of roe deers in HEF



3. GMM 결과

(1) 변수별 결과

GMM 분석 전 변수들의 특징에 따라서 분류를 나누었다. 그 특징은 Table 8.와 같다.

Table 8. A bundle of variables according to a characteristic

Group	contents	description
Topographical variable(Tv)	DEM, Hillshade, Aspect, Slope	DEM에서 파생되어 만들어지는 물리 및 지형적 요소
Distance variable(Dv)	D_agriculture, D_forest(-), D_road, D_stream, D_oreum	특정 조건에 대하여 거리적으로 영향을 받는 것을 확인하기 위한 거리적 요소
Forest characteristics variable(Fv)	Kung, Yung, Imsang, Deciduous, Evergreen, Conifer, Hect_each, Count_under, Basin_map	경급, 영급, 임상 등 임목의 특성 및 축적 등 숲의 특성과 관련된 요소
Satellite image variable(Sv)	TM_Jan, TM_Apr, Jan_NDVI, Apr_NDVI	Landsat 8 위성에서 추출한 위성 영상 요소

MaxEnt의 경우 각 변수에 따라서 percent contribution이 분석되어 변수에 대한 가감이 가능하지만, GMM의 경우 precent contribution이 분석되지 않는다. 이런 단점으로 인해 변수들의 기여도를 확인할 수 없으며, 이에 따른 변수 음 가감을 할 수 없다. 변수의 가감에 대한 선택을 시행할 수 없어, 임의적으로 변수별 특징으로 구분지어 분석을 실시하였으며, PCA(principal component analysis)를 통해 변수를 감소시켜 분석을 실시하였다.

각 결과는 Fig. 21. 과 같다. 각 결과들의 출현확률은 0~1로 표시하였으며, 그 값이 1에 가까울수록 출현확률이 높게 나타내는데, 붉은색일수록 0에 가까우며, 초록색일수록 1에 가깝다. Tv의 결과의 경우 한남시험림 북쪽지역과 동쪽지역에 있는 오름 지역을 제외하고 많은 지역에서 높은 출현확률을 보여주고

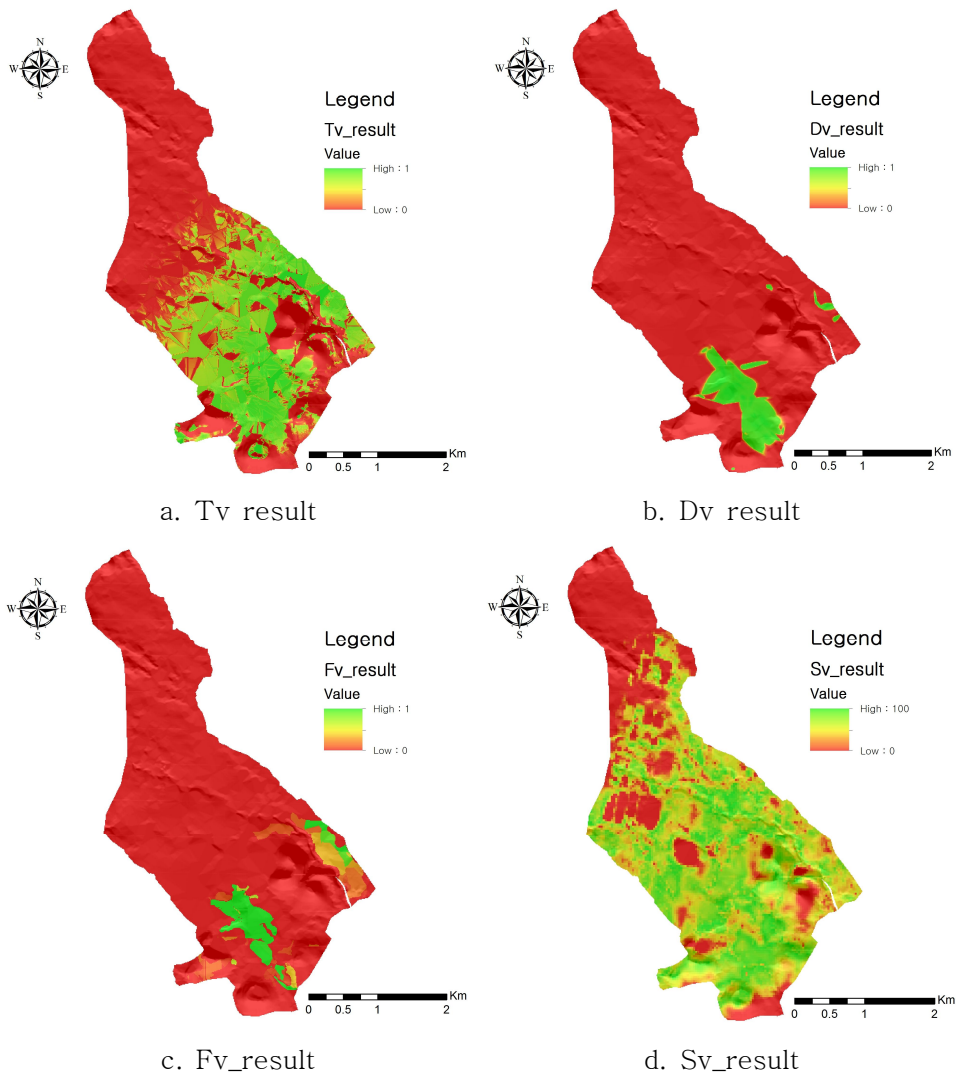


Fig. 21. result of GGM using parameter feature

있다. Dv의 경우 남쪽 일부 지역과 동쪽 일부 지역에서 높은 출현 확률을 보이고 있으며, 대부분의 지역에서 낮은 출현확률을 보이고 있다. Fv의 경우 남쪽 일부지역과 동쪽 일부지역에서 높은 출현 확률을 보이고 있으며, 대부분의 지역에서는 낮은 출현확률을 보이고 있다. Sv의 경우 다른 3개의 결과 보다 많은 지역에서 높은 출현확률을 보여주고 있다.

각 변수별 결과들은 사용된 변수들에 영향을 받고 있는데, 출현 데이터의 경

우 일부 지역에 편향되어 있는 점과 주어진 데이터의 집합 분산 분포 밀도를 단 하나의 확률밀도 함수를 추정해주는 GMM의 특성을 고려해 볼 때, 출현 데이터에 의해 추정된 확률밀도의 추정 값이 편향되어 영향을 받은 것으로 보인다. Tv와 Sv에 사용된 변수들의 경우, 각 격자 값들의 분포가 고루 퍼져있어 출현 데이터에서 추출된 격자 값들이 일부 지역에 편향된 값이라도 한남시험림 전체에 고루 편재되어 있는 경향이 있다. 이와 반면에, Dv와 Fv의 경우 Tv와 Sv와 다르게 격자 값들이 일부 지역에 편향되어 있는 모습을 보이고 있으며, 사용된 변수들에서 도출된 확률 밀도함수의 교집합 되는 지역이 일부 지역에 한정된 것으로 판단된다. 그 결과 일부지역에서 한정된 결과를 보이고 있다.

(2) 전체 변수를 사용한 결과

분석에 사용된 23개의 변수 전체를 사용해서 GMM 분석 결과는 다음 Fig. 22.와 같다.

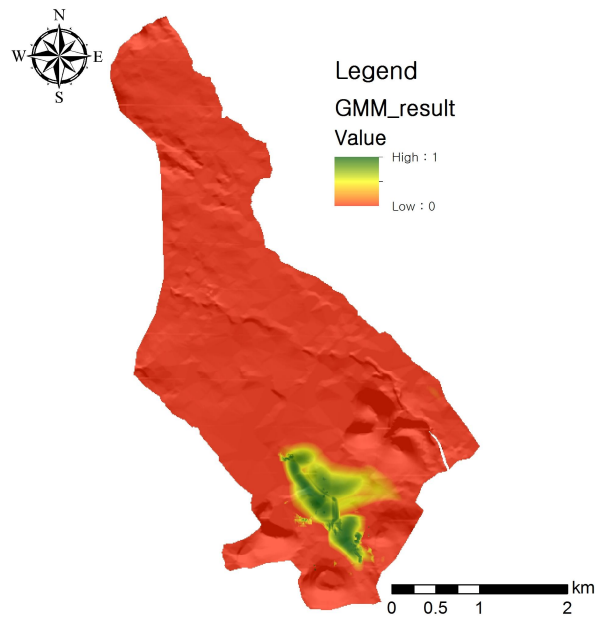


Fig. 22. Prediction map of roe deers in HEF

결과의 출현확률은 0~1로 표시하였으며, 그 값이 1에 가까울수록 출현확률이 높게 나타내는데, 붉은색일수록 0에 가까우며, 초록색일수록 1에 가깝다. GMM

결과의 경우 한남시험림 남쪽 일부지역에서 출현이 높게 보이고 있으며, 대부분의 지역에서는 낮은 출현 확률을 보이고 있다. 뿐만 아니라 출현을 보이고 있는 지역이 Dv와 Fv와 유사한 것을 알 수 있다.

사용된 환경변수들의 특징들을 클러스터링 하고, 그 정보를 이용하여 확률밀도 함수를 추정하는 GMM의 특성을 보았을 때, 사용된 변수와 출현데이터에 따라 출현예측이 높은 지역이 분석되는 것으로 판단된다. 출현데이터의 경우 출현 예측지도와 유사한 지역에서 높은 출현확률을 보이고 있으며, 출현데이터의 기반으로 클러스터링한 환경변수들 또한 일부지역에 한정적인 값을 가지고 있다. 한정적인 값을 통한 확률밀도함수를 추정하고, 그 추정 값으로 확률을 도출하는 GMM의 경우, 그 값이 유사하지 않거나 다른 경우 출현확률에 대해 매우 낮게 평가하는 것을 알 수 있다.

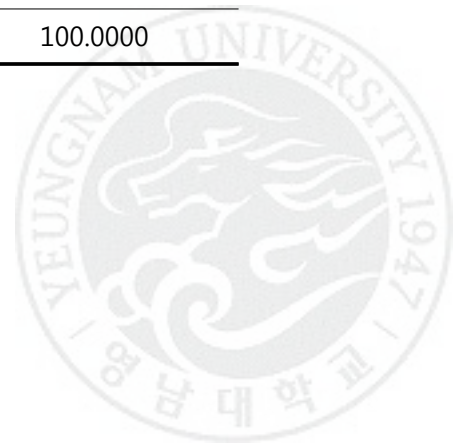
(3) PCA(Principal component analysis)를 사용한 후의 GMM 결과

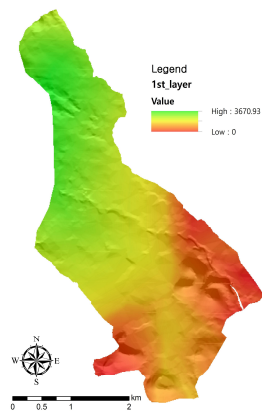
PCA는 다차원의 변수를 저차원으로 낮추어 주는데, 각 변수들 간에 상관관계나 유사성 등을 파악하여 분석하여 준다. PCA를 실시한 결과는 다음 Table 9.과 같다. PCA 결과에서 Percent of EigenValues값이 높고, Accumulative of EigenValues 누적값이 높은 상위 5개를 선택하였다(Fig. 23). EigenValues 값이 높을 수록 Layer가 가지는 설명력과 기여도가 높은 것을 의미하는데 이를 바탕으로 선택한 5개의 Layer를 이용하여 GMM 분석을 실시하였다. 이 과정은 23개의 변수를 사용함에 따라 변수의 특성, 각 변수 내 격자정보의 분포 등으로 발생하는 예측 모델의 편향성을 감소시켜 보고자 시행하였다.

Table 9. Percent and Accumulative eigenvalues

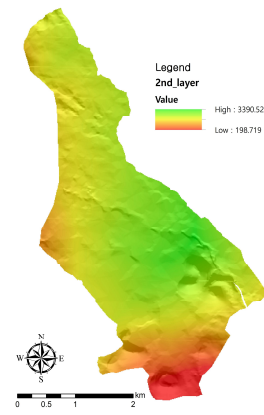
Layer	EigenValue	Percent of EigenValues	Accumulative of
-------	------------	------------------------	-----------------

			EigenValues
1	5.472411e+005	65.8925	65.8925
2	1.560484e+005	18.7896	84.6821
3	7.686887e+004	9.2557	93.9377
4	3.063831e+004	3.6891	97.6268
5	1.112879e+004	1.3400	98.9669
6	5.325833e+003	0.6413	99.6081
7	1.573172e+003	0.1894	99.7975
8	6.316842e+002	0.0761	99.8736
9	4.299961e+002	0.0518	99.9254
10	1.809560e+002	0.0218	99.9472
11	1.529742e+002	0.0184	99.9656
12	1.325697e+002	0.0160	99.9816
13	1.181463e+002	0.0142	99.9958
14	1.741875e+001	0.0024	99.9979
15	1.065699e+001	0.0013	99.9992
16	4.538701e+000	0.0005	99.9997
17	1.921170e+000	0.0002	99.9999
18	2.807115e-001	0.0000	100.0000
19	1.868611e-001	0.0000	100.0000
20	2.461927e-002	0.0000	100.0000
21	1.692626e-002	0.000	100.0000
22	4.629855e-005	0.000	100.0000
23	7.989300e-006	0.000	100.0000

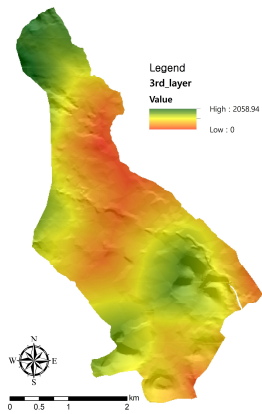




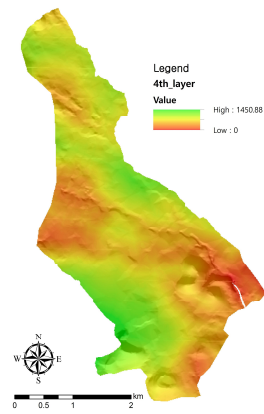
a. Layer 1



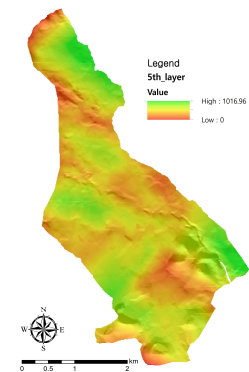
b. Layer 2



c. Layer 3



d. Layer 4



e. Layer 5

Fig. 23. result of PCA



PCA를 실시한 후 나온 결과는 Fig. 24와 같다. 결과의 출현확률은 0~1로 표시하였으며, 그 값이 1에 가까울수록 출현확률이 높게 나타내는데, 붉은색일수록 0에 가까우며, 초록색일수록 1에 가깝다. PCA 과정을 거치지 않은 Fig. 22와 비교하였을 때, 세부적으로는 출현 확률이 높은 지역이 차이가 있으나, 한남시험림 전체적으로 보았을 때는 유사한 지역에서 출현확률이 높게 나타남을 확인할 수 있다. 다차원의 변수를 저차원으로 낮추어주는 PCA를 사용하였음에도 불구하고 비슷한 결과가 도출되었다. 출현데이터에 편향된 결과가 나왔는데, 이는 사용되는 변수를 감소시켜도 편향된 출현데이터에 영향을 더 받는 것을 의미한다.

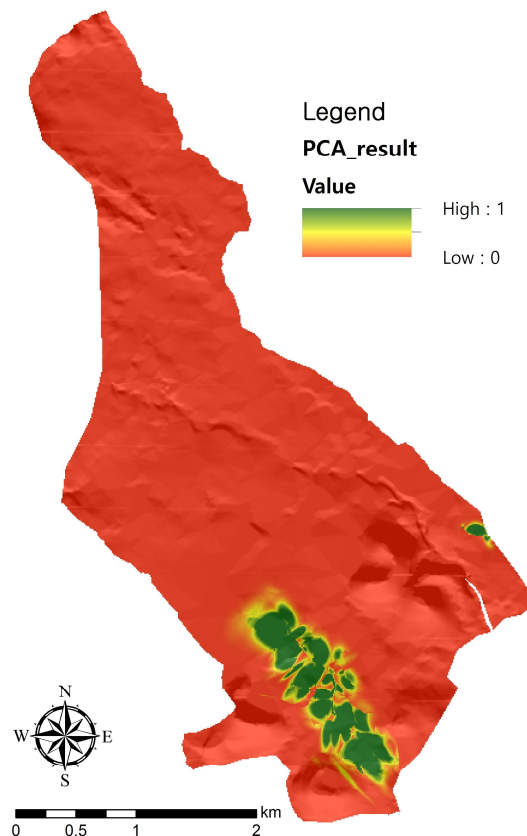


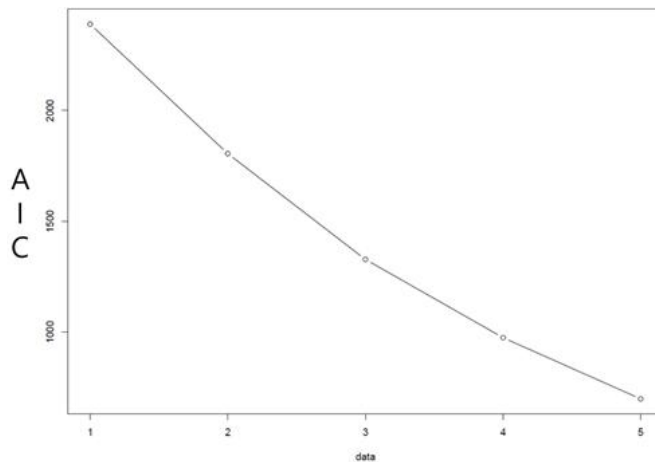
Fig. 24. result of GMM using PCA



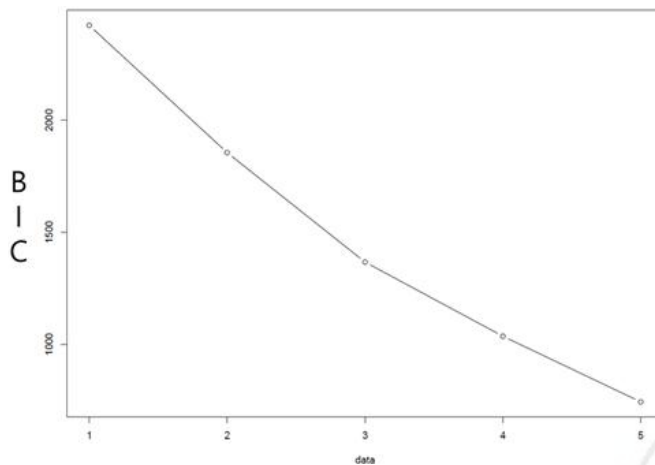
4. Logistic regression 결과

(1) AIC와 BIC 결과

실제적인 서식지 이용자료를 적절히 선택하여 사용하기 위해서는 귀납적, 경험적 모형인 다변량 통계분석이 사용된다(서창완과 박종화, 2000). Logistic regression의 결과에 대한 적합성 평가로 AIC와 BIC를 이용하였는데, 각 사용한 데이터의 AIC와 BIC 값은 다음 표와 같다.

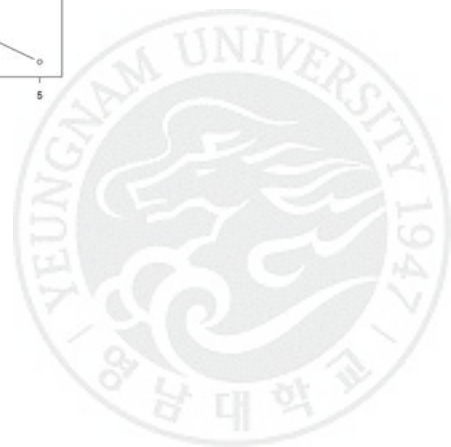


a. AIC result



b. BIC result

Fig. 25. Result of AIC and BIC of logistic regression



사용한 데이터 5개의 분류 중 data 5에서 가장 낮은 AIC와 BIC 값을 보이고 있다. AIC와 BIC의 값이 낮으면 낮을수록 모델의 적합성이 뛰어난 것을 바탕으로 data 5의 Logistic regression 결과가 가장 뛰어난 것을 확인할 수 있다. 이 결과를 통해 data 5의 logistic regression 공식을 통해 지도화 하였으며, 다른 결과들도 지도화하여 비교하였다. 다음은 노루의 출현확률을 구할 때 사용된 회귀식이다.

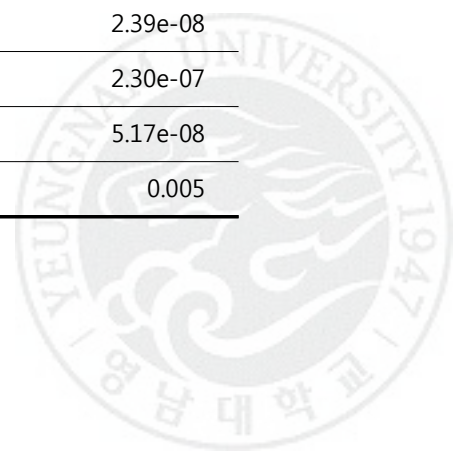
$$P = \frac{e^A}{1 + e^A}$$

회귀공식에서 A는 각 data 분석에서 나온 측정값을 통해 그 A의 값을 구하게 되는데, 각 data 따라 나온 측정값은 다음 Table 10.와 같으며, 이 값들을 이용해 A의 값을 구하게 된다.

Table 10. Result of logistic regression

data	variable	Estimate	Std. Error	Pr(> z)
1	(intercept)	0.8005812	0.2892152	0.005638
	D_oreum	0.0058166	0.0007862	1.38e-13
	D_stream	-0.0019984	0.0004767	2.76e-05
	D_forest(-)	0.0045604	0.0008415	5.99e-08
	D_agriculture	0.0022098	0.0005646	9.07e-05
	Slope	-0.0529512	0.0141668	0.000186
2	(intercept)	6.8963199	3.4555464	0.04596
	Jan_NDVI	38.4484216	13.1615570	0.00349
	TM_Apr	-0.3684941	0.1753403	0.03559
	Hect_each	0.0013886	0.0005874	0.01808
	Deciduous	0.0067084	0.0032841	0.04109
	D_oreum	0.0077923	0.0010530	1.86e-13
	D_stream	-0.0033714	0.0006628	3.64e-07
	D_forest(-)	0.0073520	0.0012252	1.96e-09
	D_agriculture	0.0041999	0.0008226	3.29e-07

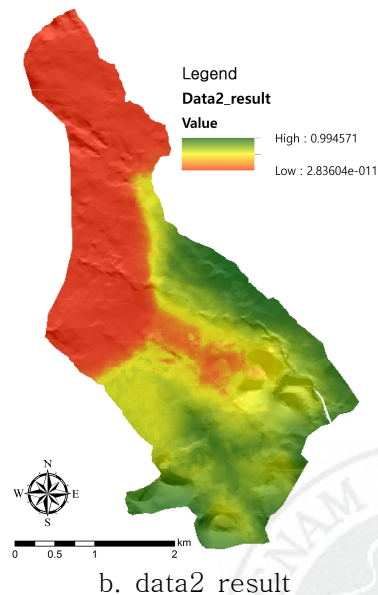
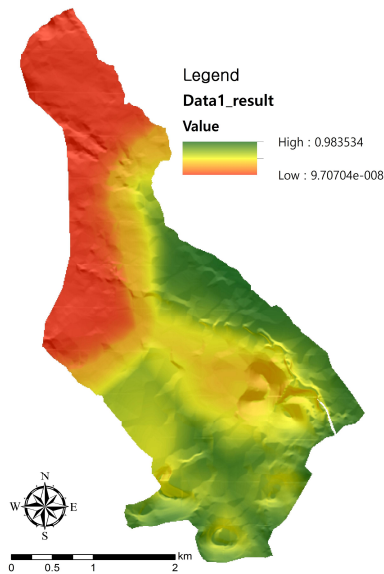
3	(intercept)	10.2689501	3.2842384	0.00177
	Jan_NDVI	61.1941659	14.5670888	2.66e-05
	TM_Jan	-0.8982902	0.2756422	0.00112
	D_oreum	0.0083561	0.0013099	1.78e-10
	D_stream	-0.0034858	0.0008123	1.77e-05
	D_forest(-)	0.0076417	0.0015047	3.80e-07
	D_agriculture	0.0042037	0.0009992	2.59e-05
4	(intercept)	14.7663280	4.3798331	0.000748
	Apr_NDVI	54.7410395	16.3542898	0.000816
	TM_Jan	-0.7737210	0.3307481	0.019320
	Hect_each	0.0033159	0.0009626	0.000572
	Decodipis	0.0149483	0.0050427	0.003033
	D_oreum	0.0066578	0.0016200	3.96e-05
	D_stream	-0.0055007	0.0011390	1.37e-06
	D_road	0.0009741	0.0004871	0.045526
	D_forest(-)	0.0089116	0.0019174	3.36e-06
	D_agriculture	0.0054629	0.0013988	9.41e-05
	hillshdae	-0.0316949	0.0128890	0.013930
5	(Intercept)	23.324758	4.952674	2.18e-06
	Apr_NDVI	85.191259	21.316380	6.43e-05
	TM_apr	-1.024279	0.230016	8.46e-06
	D_oreum	0.011853	0.002281	2.03e-07
	D_stream	-0.007376	0.001322	2.39e-08
	D_forest(-)	0.013557	0.002621	2.30e-07
	D_agriculture	0.009054	0.001663	5.17e-08
	slope	-0.107986	0.038467	0.005

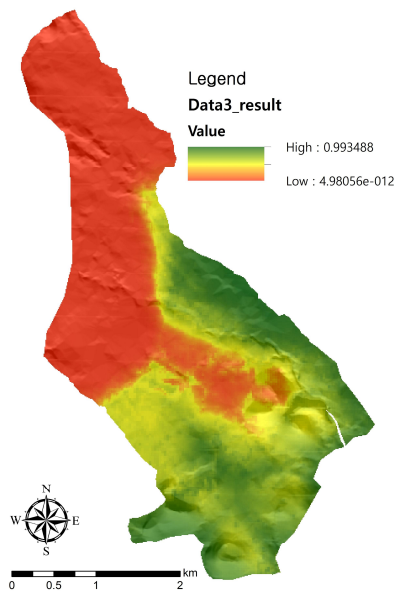


A의 값은 각 data들의 분석에서 나온 Estimate 값을 각 변수들에 곱하고, 그 값들을 더하여 A를 구하여 준다. 모델의 가장 적합하다고 판단된 data 5의 결과를 공식으로 표현하면 다음과 같다.

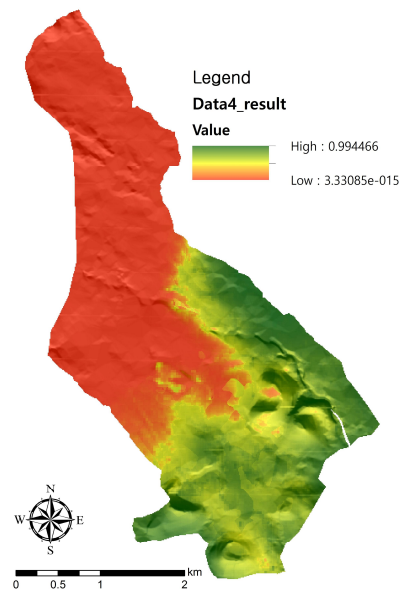
$$A = 23.324758 + \text{Apr_NDVI} * 85.191259 + \text{TM_apr} * (-1.024279) + \text{D_oreum} * 0.011853 + \text{D_stream} * (-0.007376) + \text{D_forest} * 0.013557 + \text{D_agriculture} * 0.009054 + \text{Slope} * (-0.107986)$$

A 값을 구하고 이를 회귀공식에 대입하여 P의 값을 구하게 되며, 이 값을 ArcMap을 이용하여 지도화하여 표현한다(Fig. 26.). 결과의 출현확률은 0~1로 표시하였으며, 출현 확률이 높을수록 초록색으로 나타나며, 출현확률이 낮을수록 붉은색으로 나타난다. 결과를 보면 Data 1에서 data 5로 갈수록 출현확률이 높은 지역이 점차 줄어들고 있음을 확인할 수 있다. 가장 출현확률이 넓게 표현해주는 결과는 Data 1이고, 가장 좁은 지역을 표현해주는 것은 Data 5이다. AIC 값과 BIC 값을 고려한 Data 5의 결과가 logistic regression 결과 중에서 가장 적합한 예측지도임을 확인할 수 있다.

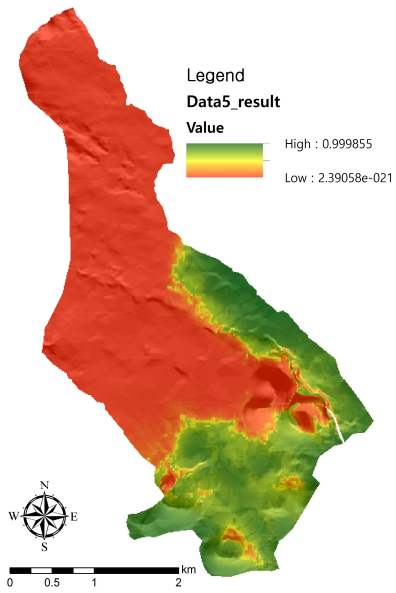




c. data3 result



d. data4 result



e. data5 result

Fig. 26. result of logistic regression



5. 현장검증

한남시험림을 대상으로 종 분포 모형을 구축한 것에 대한 평가를 위해 현장검증을 실시하였다. 한남시험림에서 현장조사를 한 지역은 Fig. 27.과 같다. 실제 현장조사에서 ②와 ⑧의 지역에서 실제 노루의 출현을 확인하였다. 모델링의 결과에서는 노루의 출현이 낮은 지역으로 분석이 되었지만 실제 검증에서는 확인하였다. 또한 ⑥와 ⑦의 지역에서는 노루의 사체와 두개골을 발견하였는데, 이를 통해 노루의 간접적인 출현을 확인하였다. 모델링에서 출현확률이 매우 낮은 지역으로 분석된 ④와 ⑤지역에서는 노루의 출현을 확인하지 못하였으나, 노루가 먹이원으로 사용할 수 있는 하층식생이 고루 분포함을 확인하였고, 국립산림과학원 난대·아열대산림연구소의 전문가의 증언을 통해 노루의 출현을 간접적으로 확인하였다. 다른 지역에 대해서도 노루가 서식할 수 있는 하층식생과 지형적 요소들을 확인하였다.

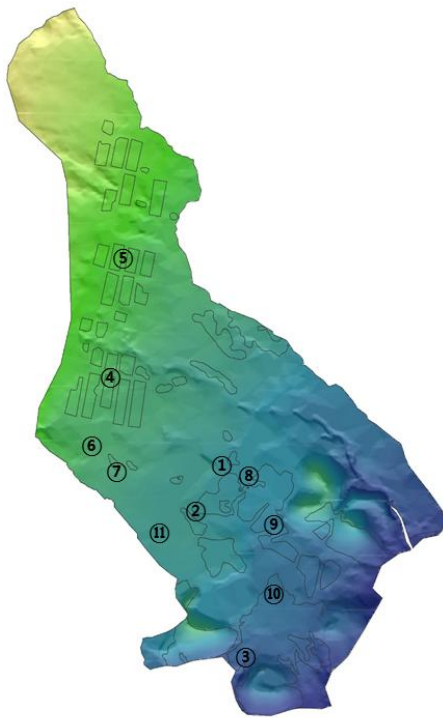


Fig. 27. Filed test area





① 지점



② 지점



③ 지점



④ 지점



⑤ 지점



⑥ 지점



⑦ 지점



⑧ 지점



⑨ 지점



⑩ 지점



⑪ 지점

Fig. 28. Field test result



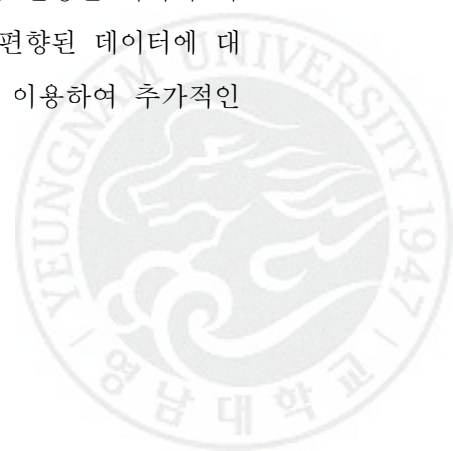
V. 결 론

본 연구는 제주도에 있는 한남시험림을 대상으로 제주노루의 종 분포 모형을 구축하고 그 결과에 따른 해석을 실시하였다.

1. MaxEnt를 활용하여 선행연구와 비교하였을 때, 예측지도의 결과는 선행연구와 본 연구와 비슷하다. 하지만 사용된 변수와 그 변수들의 percent contribution과 Permutation importance 값이 달라지는 것을 확인하였다. 추가된 변수에 따라 percent contribution과 permutation importance의 값이 변화하였는데, 이는 본 연구에서 추가로 사용된 변수들이 노루의 종 분포 모형 구축에 유의미하게 적용되었다는 것을 확인할 수 있다. 또한, 선행연구에서 사용된 변수들의 경우, 본 연구에서는 낮은 기여도를 보였는데, 이는 사용되는 변수에 따라 종 분포 모형은 구축되지만 그 변수가 기여하는 정도가 달라진다는 것을 알 수 있다. 종 분포모형을 구축할 때는 다양한 변수들의 고려하여 사용하여야 한다는 점을 알 수 있는데, 소수의 변수를 사용할 경우 종 분포 모형은 구축되지만 분포에 대한 설명이 오류가 발생할 수 있다.
2. GMM은 주어진 데이터의 집합 분산 분포 밀도를 단 하나의 확률밀도 함수로 모델링 하는 방법을 개선한 밀도 추정방법으로 평균과 분산의 파라미터를 가지는 여러 개의 가우시안 함수로 데이터의 분포를 모델링하는 방법으로 새롭게 종 분포 모형에서 시도해 본 모델이다. 사용된 변수들에 따라 나오는 결과가 다른 부분들이 도출되며, 각 변수들의 특징을 클러스터링 한다는 점에서 출현 지점에 영향을 많이 받는 것을 결과를 통해 확인하였다. 각 변수들의 특징에 따라 도출되는 결과들 또한 달라짐을 확인하였다. 종 분포 모형으로 시도된 본 연구에서는 출현에 대한 예측결과를 제시하였다. 차후 GMM을 사용한 종 분포 모형에 대한 심도깊은 연구가 필요할 것으로 보인다.

다.

3. Logistic regression을 이용하여 한남시험림에 대해 분석을 실시하였는데, 다른 모델에 비해 비편향적인 결과를 제시하였다. AIC와 BIC의 검증을 통해 모델의 적합성 또한 평가하였는데, 출현빈도가 높은 지역을 실제 출현지역으로 설정하였을 때 그 모델의 적합성이 높게 평가되었다. 다른 모델과 같이 일부 지역에 편향된 출현자료를 사용하였지만, 출현지역에 상대적으로 편향되지 않고 한남시험림에서 광범위하게 노루가 출현할 수 있는 환경을 제시해 주었다. 이는 출현 데이터와 변수들의 상관관계, 변수들의 유의미한 기여도 등을 분석한 결과로 편향된 출현데이터를 사용할 때는 logistic regression이 다른 두 모델보다 중 분포 모형에 적합한 것으로 보인다.
4. 현장검증과 중 분포 모형들을 비교하였을 때, 모든 중 분포 모형이 노루의 출현을 완벽하게 설명하지는 못한다. 실제 검증에서는 한남시험림 내 노루의 분포는 고루 분포한 편이지만, logistic regression을 제외한 모델링의 결과는 대부분 출현지역에 의존되어 편향된 결과를 보이고 있다. 뿐만 아니라 Logistic regression도 완벽하게 설명을 하진 못하고 있다. 만약 실제 노루의 출현이 출현데이터와 동일하다면, 중 분포 모형은 노루의 출현확률 및 출현이 가능한 지역에 대한 평가가 올바르게 이루어 졌다고 할 수 있다. 하지만 노루의 출현은 일부 지역에 대해 편향되어 있지 않고 한남시험림 일대에 고루 분포하고 있기에, 중 분포 모형으로 완벽한 설명이 되지 않는다. 데이터가 일부 지역에 편향되어 있는 출현 데이터로 분석을 실시하여, 그 편향된 출현 데이터가 결과에 영향을 준 것으로 판단된다. 완벽한 노루의 중 분포 모형을 위해서는 한남시험림 내에 노루의 출현 데이터를 편향된 지역이 아닌 다른 지역에서 확보하여 추가적인 분석을 이루거나, 편향된 데이터에 대해 합리적으로 편향된 결과를 상쇄시킬 수 있는 방법을 이용하여 추가적인 분석이 이루어 져야 할 것으로 보인다.



VI. 참 고 문 헌

- Anderson, R. P., Lew, D., Peterson, A. T.. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, 2162, 211-232.
- Abhishek Kumar Chauhan, Prashant Krishan. (2013) Moving object tracking using gaussian mixture model and optical flow. *International journal of advanced research in computer science and software engineering*, 4(3)
- David J. Abson, Andrew J, Dougill, Lindsay C. Stringer. (2012) Using principal component analysis for information-rich socio-ecological vulnerability mapping in souther africa. *applied geography*, 35, 515-524
- Fangliang He, (2010) Maximum entropy, logistic regression, and species abundance. *Oikos*, 119, 578-582
- Fawcett T.. (2006) An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.
- Guangshun Jiang, Minghai Zhang, Jianzhang Ma. (2008) Habitat use and separation between red deer *Cervus elaphus xanthopygus* and roe deer *Capreolus pygargus bedfordi* in relation to human

disturbance in the Wandashan Mountains, northeastern china.
Wildlife Biology, 14(1), 92-100

Guisan, A., Zimmermann, N. E.. (2000) Predictive habitat distribution models in ecology. Ecological Modelling, 135, 147-186.

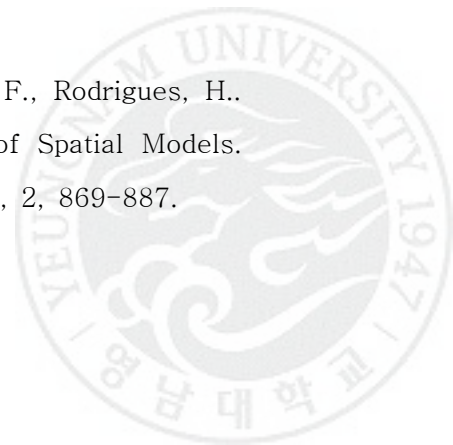
Hernandez, P. A., Graham, C. H., Master, L. L., Albert, D. L.. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29, 773-785.

Huaifei Hu, Haihua Lue, Zhiyong Gao, Lu Huang. (2013) Hybrid segmentation of left ventricle in cardiac MRI using gaussian-mixture model and region restricted dynamic programming. Magnetic resonance imaging, 31, 575-584

Irina N. Sheremetyeva, Ilya S. Sheremetyev. (2008) Skull variation in the Siverian roe deer *Capreolus pygargus* from the Far East: a revision of the distribution of the subspecies. Eur J wildl Res, 54, 557-569

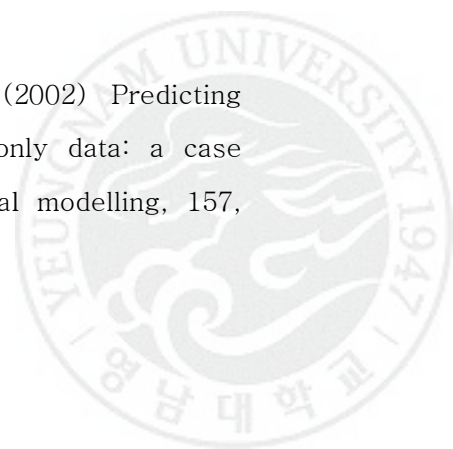
Kenneth P. burnham, David R. Anderson, (2004) Multimodel inference understanding AIC and BIC in model selection. Sociological methods & research, 33(2), 261-304

Mas, J., Filho, B. S., Pontius, R. G. Jr., Gutierrez, M. F., Rodrigues, H.. (2013) A Suite of Tools for ROC Analysis of Spatial Models. ISPRS International Journal of Geo-Information, 2, 869-887.



- Pearson, R. G., Raxworthy, C. J., Nakamura, M., Peterson, A. T.. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34, 102-117.
- Phillips, S. J. and Dudik, M.. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.
- Phillips, S. J., Anderson, R. P., Schapire, R. E.. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.
- Phillips, S. J., Dudik, M., Schapire, R. E.. (2004) A Maximum Entropy Approach to Species Distribution Modeling. *International Conference on Machine Learning*, 21, 665-662.
- Pradeep Adhikari, Seon-mi Park, Tae-Wook Kim, Jun-Won Lee, Ga-Ram Kim, Sang-Hyun Han, Hong-Shik Oh. (2016) Seasonal and altitudinal variation in roe deer (*Capreolus pygargus tianschanicus*) diet on Jeju Island, South Korea. *Journal of Asia-Pacific Biodiversity*, 9, 422-428
- Quillfeldt, P., Masello, J. F., Navarro, J., Phillips, R. A.. (2013) Year-round distribution suggests spatial segregation of two small petrel species in the South Atlantic. *Journal of Biogeography*, 40, 430-441.

- Ratko Grbic, Drazen Sliskovic, Petr Kadlec. (2013) Adaptive soft sensor for online prediction and process monitoring based on a mixture of gaussian process models. *Computers and chemical engineering*, 58, 84-97
- Songer, M., Delion, M., Biggs, A., Huang, Q.. (2012) Modeling Impacts of Climate Change on Giant Panda Habitat. *International Journal of Ecology*, 2012, 1-12.
- Stockwell, D. R. B. and Peterson, A. T.. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1-13.
- Timonin, V.; Bai, S. B.; Wang, J.; Kanevski, M.; and Pozdnukhov, A., "Landslide Data Analysis with Gaussian Mixture Model" (2008). *International Congress on Environmental Modelling and Software*. 54.
- Vladimir A Trifonov, Polina V Dementyeva, Denis M Larkin, Patricia CM O'Brien, Polina L Perelman, Fengtang Yang, Malcolm A Ferguson-Smith and Alexander S Graphodatsky. Transcription of a protein-coding gene on B chromosomes of the Siberian roe deer(*capreolus pygargus*). *BMC Biology*, 11(90)
- Zaniewski, A. E., Lehmann, A. Overton, J. McC.. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *ecological modelling*, 157,



261-280.

권혁수. (2014) 종분포모형의 불확실성 확인을 위한 앙상블모형 적용. 한국지형
공간정보학회지. 22(4), 47-52.

권혁수, 류지은, 서창완, 김지연, 임동옥, 서민환. (2012) 종분포모형을 이용한
히어리 서식지의 분포 특성 연구. 환경영향평가, 21(5), 735-743.

김도훈과 홍영교. (2006) 제주도 한라산 노루에 관한 시스템적 접근. 한국 시스
템다이내믹스 연구, 7(2), 191-213.

김병수, 오장근, 오홍식. (2007) 한라산에 서식하는 노루(*Capreolus pygargus*
tianschanicus Satunin)의 개체수 현황과 관리방안. 한국환경생태학회
지, 21(4), 366-373.

김아름, (2016) 제주노루(*Capreolus pygargus*)의 공간적 분포와 서식지 선호
도 분석. 영남대학교 일반대학원 석사학위논문.

김은미, 권진오, 강창완, 송국만, 민동원. (2013) GPS와 CDMA를 이용한 난대
림 출산 전후 암노루행동권 및 서식환경 조사. 한국지리정보학회지,
16(2), 65-74.

김지은, 김문홍. (2001) 노루(*Capreolus pygargus tianschanicus*)가 선호하는
한라산의 자생식물 조사. J. of Basic Sciences, 14(1), 63-72

김호걸, 이동근, 모용원, 길승호, 박찬, 이수재. MaxEnt 모형을 이용한 기후변
화에 따른 산사태 발생가능성 예측. Journal of Environmental
Impact Assessment, 22(1), 39-50



- 박용수, 이우신. (2014) 설악산국립공원 지역의 노루 서식지 이용 특성에 관한 연구. 한국환경복원기술학회지, 17(1), 91-109
- 서창완과 박종화. (2000) GIS와 로지스틱 회귀분석을 이용한 멧돼지 서식지 모형 개발. 한국GIS학회지, 8(1), 85-99.
- 송원경과 김은영. (2012) 고라니 서식지 분석을 위한 기계학습식 종분포모형 비교. 대한원격탐사학회지, 28(1), 171-180.
- 송원경, (2015) 종분포모형을 이용한 수원청개구리의 번식지 서식지 분석, 한국 환경복원기술학회지, 18(1), 71~82
- 신형식. (2017) 한국 신생아의 출생체중 데이터 보정. 한국정보통신학회논문지, 21(2), 259-264
- 오장근. (2004) 제주도 노루(*Capreolus pygargus tianschanicus*)의 생태학적 행동 특징. 한국교원대학교 대학원 박사학위논문.
- 오장근. (2010) 노루의 분포특징과 보호관리방안. 제주특별자치도 환경자원연구원, 제주노루의 가치와 효율적인 관리방안 심포지엄 자료집, 1-40.
- 오장근. (2015) 제주도 전역 노루 개체수 조사. 제주특별자치도 세계유산·한라산연구원 조사연구보고서, 14, 267-274.
- 이용준, 박근애, 김성준. (2006) 로지스틱 회귀분석 및 AHP 기법을 이용한 산사태 위험지역 분석. 대한토목학회논문집 D, 26(5D), 861-867.



전미진, 이준재. (2012) 가우시안 혼합 모델 기반의 영상 히스토그램 평활화.

Journal of Korea multimedia society, 15(6), 748-760



M.S. Thesis

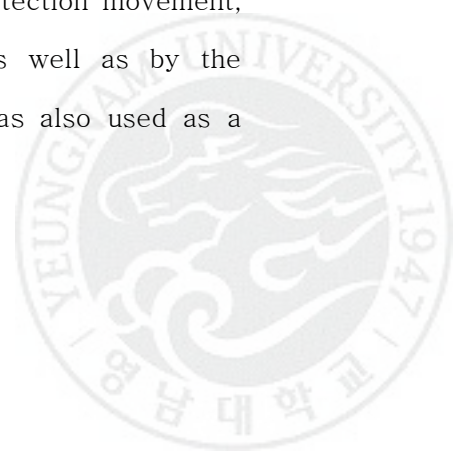
Prediction of roe deer(*Capreolus pygargus*) habitat
using GMM and Logistic regression model

Kim, Junhyuk

Department of Life Science
Graduate School
Yeungnam University
(Supervised by Professor Jang, Gab-Sue)

abstract

Roe deer in jeju island were sharply increased the number of objects by Jeju protection polices implemented in the 1980s. In the background, there is Jeju Island and Jeju City's will to make Jeju Island as a symbol of Jeju Island.t has been reported that the number of individuals has increased to several thousand by the efforts of the protection movement, such as providing feeders and collecting catches, as well as by the NGOs and the National Park Management Office. It was also used as a tourism resource for Jeju Island.



However, as the number of individuals increased, new problems occurred. Problems occurred in various fields such as agriculture, forestry, and environment. Especially, damage to farm households continued to occur. Due to these problems, the Jeju Special Self - governing City Council designated the gazelle as a harmful animal from July 2013 to June 2016 and captured about 5000 roe deer and re-designated as a harmful animal until June 30, 2019. On the other hand, opinions of environmental groups have started to come out that the protection of the deer should be protected due to the decrease of the number of the deer. The damage farmers are in favor of an additional capture of the deer, but environmental organizations claim that they should find a way to coexist with humans and roe deer.

Therefore, in this study, we approach from the perspective of space to understand the gaze. Species Distribution Model (SDM) was used to identify the distribution habitat of the roe. The GPS tracking data was used in Hannam Experimental Forest (HEF) and analyzed using three models. MaxEnt, which is known to show the most reasonable and efficient results in using presence-only data, was used, considering presence-only data only in the case of acquired data. In addition, the parameter values were clustered and analyzed using Gaussian mixture model (GMM), which shows the probability distribution after machine learning. Further analysis was performed using logistic regression. Based on these results, we constructed a species distribution model and presented the limitations.

