

산림조사 자료의 표본점 할당 모델링을 통한 고해상도 산림공간정보 구축

김승욱¹ · 하의린¹ · 이성재² · 임종수³ · 한 희^{1,4} · 최혜영^{1,4*}

¹서울대학교 농림생물자원학부, ²서울대학교 학술림,

³국립산림과학원 산림탄소연구센터, ⁴서울대학교 농업생명과학연구원

Mapping High-resolution Forest Geospatial Information Using Imputation Modeling of Forest Inventory Plot Data

Seunguk Kim¹, Uirin Ha¹, Sung-Jae Lee², Jongsu Yim³, Hee Han^{1,4} and Hyeyeong Choe^{1,4*}

¹Department of Agriculture, Forestry and Bioresources, Seoul National University, Seoul 08826, Korea

²Seoul National University Forest, Seoul 08826, Korea

³Forest Carbon Center on Climate Change, National Institute of Forest Science, Seoul 02455, Korea

⁴Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

요약: 기후위기 속에서 산림의 가치는 더욱 중요해지고 있으며, 이를 극대화하기 위해서는 수종 구성, 영급 분포, 구조적 특성, 탄소저장량 등 종합적인 산림공간정보가 필수적이다. 최근 머신러닝 기반 공간분석을 응용하여 표본점 중심의 산림조사 자료를 넓은 지역으로 확장하는 연구가 활발하며, 이를 국내 산림 환경에 적용하기 위한 검토가 필요하다. 이 연구는 지리산 지역을 대상으로 랜덤 포레스트(random forests) 기반의 표본점 할당 모델링을 수행하여 임목 수준의 고해상도 산림공간정보를 구축하였다. 이를 위해, 국립공원연구원, 서울대학교 남부학술림의 산림조사 자료를 일관된 기준으로 통합하였으며, 세 자료를 선택하는 조합에 따라 총 7개의 모형을 구성하여 성능을 다각적으로 검증한 후 최적의 모형을 선정하였다. 표본점 할당 모델링은 연구지역 각 지점에서 임상, 임분고, 경급이 가장 유사한 표본점을 식별하여 산림조사 자료를 할당하도록 설계하였다. 모형 학습에서는 산림조사 지점에서 분류된 임상, 임분고, 경급의 식생변수를 다중반응변수로 설정하였고, 각 지점의 환경변수를 결합하여 참조자료로 이용하였다. 학습된 모형은 임상도, 기후, 지형, 위치 등의 환경정보를 포함하는 목표자료에 따라 연구지역 전반에 표본점을 할당하였으며, 최종적으로 100 m × 100 m 해상도의 래스터(raster) 기반 공간자료를 생성하였다. 모형 성능 평가 결과, 단일 기관의 자료만으로 구성한 모형은 해당 기관의 조사 특성이 과도하게 반영되어 특정 산림 유형을 과대 혹은 과소 추정하는 경향을 나타냈다. 그에 반해, 세 기관의 자료를 모두 통합하여 활용한 모형은 산림조사 자료 및 임상도를 재현하는 성능이 가장 우수하였으며, 전체적인 산림 분포 패턴을 안정적으로 예측하였다. 이 연구의 표본점 할당 모델링은 표본점의 고유번호를 가지는 래스터 지도를 제공하여, 사용자가 고유번호와 연결된 상세 산림정보에 접근하고 다양한 속성지도를 제작할 수 있도록 한다. 따라서 이 프레임워크는 수종의 분포 예측, 산림 구조 분석, 탄소저장량 평가 등 다양한 분야에 적용될 수 있으며, 산림조사 자료의 활용성을 크게 향상할 수 있다. 다만, 모형의 특성상 넓은 지역을 대상으로 할 때는 산림조사 지점과의 국지적 일치율이 낮아질 수 가능성이 있으며, 좁은 지역에서는 균형 잡힌 산림조사 자료를 이용해 최소한의 생태계의 정보를 충분히 확보해야 한다. 향후 산림조사 자료의 활용도를 더욱 높여려면 기관 간 조사 프로토콜을 표준화하고, 모델링 과정에서 각 기관 조사방식의 차이를 면밀히 분석하여 반영할 필요가 있다.

Abstract: Forests are increasingly valuable as climate change intensifies, and their conservation management requires comprehensive spatial information on the species composition, stand age distribution, structural characteristics, and carbon storage. Recent studies have employed machine learning-based spatial analyses for plot-based forest surveys across broad regions, but the effectiveness of this approach in South Korea remains understudied. We applied a plot imputation approach using random forests to create a high-resolution tree-level forest map of the Jiri Mountain. We standardized and integrated forest survey data from the Korea Forest Service, Korea National Park Research Institute, and Seoul National University Nambu Forest and then built seven models for evaluation. Each model identified sample points with similar vegetation attributes (class, height, and diameter) across the study area. We trained the

* Corresponding author
E-mail: hy.choe@snu.ac.kr

ORCID

Hyeyeong Choe  https://orcid.org/0000-0003-2130-1622

models using a reference dataset of vegetation and environmental variables and then assigned sample points throughout the study area based on target datasets including a forest-type map and environmental data on climate, topography, and location to generate a raster map at a 100 m $\times$ 100 m resolution. When applied to a single-institution dataset, the models tended to over- or under-estimate certain vegetation attributes likely due to survey biases. Integrating all three datasets resulted in a more balanced view of overall forest patterns and yielded the highest reproduction accuracy. The developed framework generates raster maps with unique plot identifiers to enable users to access plot details and create thematic maps for the species distribution, forest structure, and carbon storage. However, uneven distribution of the survey points may cause local losses of accuracy. Balanced forest surveys are crucial for capturing details about rare or specialized ecosystems. Standardizing survey protocols across institutions and carefully examining the differences in each dataset will be crucial to enhancing both the data quality and modeling applications of the framework.

Key words: forest mapping, forest type map, machine learning, national forest inventory, random forests

서론

기후위기 대응을 위해서 전 세계적으로 산림의 탄소흡수량에 대한 관심이 높아지는 동시에 산림 재해와 생물다양성 감소 문제도 심화되어(Cooper and MacFarlane, 2023), 산림의 수종 구성, 영급 분포, 구조적 특성, 탄소저장량 등의 종합적인 산림공간정보에 대한 사회적 요구가 증가하고 있다(Nitoslawski et al., 2021). 산림청은 1972년 이래 제7차 국가산림자원조사(2016~2020)까지 전국의 산림을 지속적으로 조사해왔으며, 그 결과를 임상도와 연계하여 산림기본통계로 제공하고 있다(KFS, 2024). 국가산림자원조사는 전국에 4 km $\times$ 4 km 간격으로 배치된 약 4,500개의 격자 표본점을 대상으로 임목, 임분, 지형 등에 관한 70개 항목을 조사하여 방대한 정보를 수집한다. 그러나 연속적인(wall-to-wall) 식생정보를 제공하지 못하기 때문에, 대개 면적 단위의 통계를 요약되어 활용된다(Chen et al., 2023). 현재 산림청의 통계는 주로 산림면적과 임목축적에 국한되어 있으며, 보다 정밀한 산림 현황 및 변화 분석을 위해서는 추가적인 공간분석이 필수적이다(Yim et al., 2020).

산림공간정보를 활용하는 다수의 연구는 산림조사 자료와 환경 요인의 관계를 분석함으로써 더 넓은 지역의 정보를 예측한다(Wilson et al., 2013). 최근 국내에서는 특정 수종이나 임분 유형을 대상으로 산림 바이오매스 혹은 탄소저장량을 추정하는 연구가 활발히 이루어졌었다(Seo et al., 2017; Moon and Yim, 2021; Kwon et al., 2022). 이때, 선형회귀모형, 크리깅, k-최근접 이웃(kNN) 기법 등을 이용하였으나, 이러한 통계적 접근법으로는 산림조사 자료를 충분히 활용하기 어렵다. 선형회귀모형은 임목 분포와 환경 조건의 비선형 관계를 설명하기 어렵고, 크리깅은 수종, 임상과 같은 범주형 자료 처리에 적합하지 않다(Riley et al., 2016). kNN 기법은 고차원 자료를 다룰 때 성능이 저하되고 이상치에 민감하여 임분 구조와 지형이 복잡한 산림에서 적용하기에 한계가 있다(Kim et al., 2014). 산림

조사 자료의 다양한 양적·질적변수로부터 식생과 환경 간의 복잡한 공간적 상관관계를 파악하기 위해서는, 기존 방법론의 한계를 극복하는 새로운 접근이 필요하다.

머신러닝 기법의 발전은 산림과학의 고도화에 광범위하게 기여하고 있다. 특히, 랜덤 포레스트(random forests)는 변수 간의 복잡한 비선형 관계를 효과적으로 모델링할 수 있고, 범주형 및 연속형 변수를 모두 처리할 수 있어 활용도가 높다(Breiman, 2001). 국내에서는 소나무림 또는 임상 변화 분석에 랜덤 포레스트가 적용된 바 있다(Kim et al., 2022; Kim et al., 2024). 국외에서는 한층 정교한 접근법이 개발되어, 미국은 전국 산림조사 자료에 지형, 생물물리적 변수, 교란 변수를 결합하고 ‘변형 랜덤 포레스트’ 기반 프레임워크를 적용하여 임목 수준의 산림공간정보를 제공하고 있다(Riley et al., 2016; Riley et al., 2021; Riley et al., 2022). 이러한 결과물은 기후변화에 따른 산림 변화를 예측하는 동적식생모형 연구나, 산림 관리에 따른 생태계서비스 변화를 평가하는 다양한 모델링 연구에서 활용되고 있다(Maxwell et al., 2022; Povak et al., 2022).

하지만 국외 연구 사례의 머신러닝 기반 프레임워크 방법론을 국내 산림 환경에 적용하기 위해서는 현지의 원자료(raw data)에 대한 철저한 검토가 필요하다. 우리나라 산림은 험난한 산악 지형으로 인해 기후와 토양 환경의 공간적 변이가 크고(Kim et al., 2014), 식생이 복잡하게 분포한다. 이러한 환경에서 산림조사는 희소한 식생을 놓치거나 과소 추정하기 쉬워, 조사 결과가 편향될 수 있다(Portela et al., 2017). 예를 들어, 아고산 침엽수림이 험준한 고지대에 제한적으로 분포하여 대규모 조사 설계에서 누락되었다면, 이 결과를 활용한 모델링 연구는 보전 정책 수립에 중요한 관리 지역을 간과할 수 있다(Park et al., 2024).

이러한 한계를 극복하기 위해서는 여러 연구기관에서 수집된 산림조사 자료를 통합하여 풍부한 산림정보를 확보하려는 노력이 중요하지만, 조사 시점과 프로토콜의 차이로 인해 드물게 이루어진다(Miller et al., 2019; Zweifel

et al., 2023). 국가 차원의 산림조사만으로는 지형이나 기후 조건이 특이한 지역, 그리고 조사 설계에서 누락된 지역의 자료 공백을 해소하기 어렵다(Tinkham et al., 2018). 여러 연구기관의 조사 자료를 포함한다면 희소 식생을 보완적으로 파악하고, 높은 표본 밀도로 인해 모형에 투입되는 자료의 다양성이 증가하여 예측 성능 향상을 기대할 수 있다(Chojnacki et al., 2014; Frost et al., 2023).

이 연구의 목적은 지리산을 대상으로 임목 수준의 정보를 포함한 연속적인 고해상도 산림공간정보를 구축하는 것이다. 지리산은 수종이 다양하고 지형이 복잡하며, 국가 산림자원조사 외에도 국립공원연구원, 서울대학교 남부학술림의 산림조사 자료가 축적되어 있어 자료 통합과 모형 검증에 적합한 지역이다(Park et al., 2020). 이 연구에서는 산림조사 자료의 표본점을 지리산 일대에 할당하는 랜덤 포레스트 기반 모델링을 수행하였다. 이때, 다기관에서 수집한 여러 입력자료의 조합을 이용해 자료 특성에 따른 모형 성능을 검증하고, 모형의 보편적 적용 가능성을 검토하였다. 이를 통해 국내 산림에 적용 가능한 모델링 프레임워크를 마련하고, 산림 보전 및 관리의 기반이 되는 산림공간정보를 구축하였다.

재료 및 방법

이 연구는 랜덤 포레스트를 활용하여 지리산 일대의 산림조사 자료를 연구지역 전체로 확장하는 표본점 할당 모델링을 수행하였다. 모델링의 최종 목표는 연구지역에서 모든 100 m × 100 m 격자점에 대해 임상, 임분고, 경급 등 식생변수가 가장 유사한 표본점을 찾아내어 표본점의 고유번호를 부여하는 것이다. 모델링에 이용한 래스터 자료의 공간해상도가 서로 달랐기 때문에 중간 수준인 100 m 크기로 맞추어 분석하였다. 모델링에 앞서 산림청, 국립공원연구원, 서울대학교 남부학술림의 산림조사 자료를 통합하였다. 이후, 여러 기관의 조사 자료를 선택하는 조합에 따라 모형 성능을 검증하고 최적의 모형을 선정하였다.

1. 연구지역

지리산국립공원을 둘러싼 연구지역은 면적이 893 km²이며, 최고 고도가 1,915 m에 이르는 고지대 주위로 봉우리가 많아 지형이 복잡하다(Figure 1). 기상청의 MK-PRISM 기법에 기반한 2000~2019년의 기후 자료에 따르면 지점별 연평균 최고기온은 9.6~20.0 °C, 최저기온은 2.9~8.1 °C, 연강수량은 1,403~2,687 mm로 지역 내 기후 변동성이 크다(Kim et al., 2022). 온대림이 주를 이루고 난대수종과 아고산 침엽수림이 공존한다.

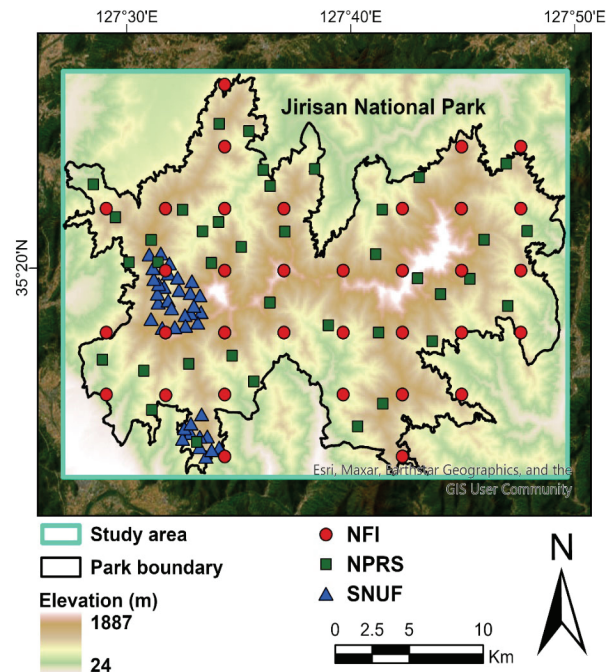


Figure 1. Elevation in the study area, Jiri Mountain, and plot locations of three forest surveys.

2. 랜덤 포레스트를 이용한 표본점 할당 모델링

랜덤 포레스트 모델링에 요구되는 참조자료(reference data)와 목표자료(target data) 데이터셋을 구축하였다. 참조자료란 연구지역 내 일부 지점에서 수집된 상세 측정자료를 의미한다. 이 연구에서는 지리산에서 수행된 3개 기관의 산림조사 자료에 환경 정보를 결합하여 참조자료로 활용하였다(Table 1). 목표자료는 연구지역 전반을 포괄하는 일반 자료로서, 주로 원격탐사 기반 자료가 활용된다. 여기서는 지리산의 식생, 기후, 지형, 식생 특성에 대한 100 m 해상도의 래스터 기반 공간자료가 이에 해당한다(Table 2). 목표자료를 기준으로 유사한 참조자료를 할당함으로써 연구지역의 모든 위치에서 산림조사 자료의 세부정보를 활용할 수 있다.

랜덤 포레스트는 참조자료와 목표자료 모두에 동일한 예측변수 집합을 요구한다(Table 2). 모델링에 이용한 예측변수는 총 12개로, 식생변수(임상, 임분고, 경급), 기후변수(강수량, 최고기온, 최저기온, 광합성유효광량), 지형변수(고도, 경사, 사면향), 위치변수(위도, 경도)로 구성하였다. 식생변수인 임상, 임분고, 경급을 예측변수이자 동시에 다중 반응변수로 이용하였고, 이는 해당 변수의 예측 정확도를 높일 수 있는 방식이다(Riley et al., 2021). 식생변수는 수종분포와 탄소저장량에 대한 영향을 고루 반영할 수 있도록 선정하였으며, 동시에 연구의 향후 활용도를 높이기 위해서 전국 범위에서의 데이터 가용성을 고려하였다.

1) 참조자료(reference data) 처리

참조자료로 제7차 국가산림자원조사(NFI, national forest inventory), 2019년 지리산국립공원 공원자원조사(NPRS, national park resource survey) 및 서울대학교 남부학술림(SNUF, Seoul National University Nambu Forest) 조사 자료를 이용하였다(Table 1, Table 2). 총 111개 조사구에서 101종, 12,900개체에 관한 자료가 수집되었다. 조사 위치, 면적과 임목의 수종, 수고, 흉고직경을 공통적으로 활용하였고, 이를 바탕으로 임상, 임분고, 경급을 정의하였다.

임상은 5개 유형 즉, 신갈나무림, 참나무림, 활엽수림, 침엽수림, 침활혼효림으로 나누었으며, 지리산의 최우점 종인 신갈나무(*Quercus mongolica*)를 비롯한 참나무류의 분포를 그 외의 임상과 구분하였다. 각 분류는 순서대로 신갈나무, 참나무류, 기타 활엽수, 침엽수의 수종별 흉고 단면적 비율이 75% 이상인 경우에 해당하며, 나머지는 침

활혼효림으로 분류하였다. 임분고는 각 표본점의 상위 3개 임목의 평균 수고를 기준으로 0~15 m, 15~21 m, 21 m 이상의 3개 범주로 분류하였다. 경급은 임목의 흉고직경을 소경목(6~18 cm), 중경목(18~30 cm), 대경목(30 cm 이상)으로 구분한 다음, 흉고단면적 비율이 가장 높은 등급을 할당하였다. 각 조사구의 좌표를 기준으로 기후, 지형, 위치 등 환경 정보를 추가하여 최종 참조자료를 구축하였다(Table 2).

식생변수를 정의하기 위해서는 조사방법, 조사항목, 조사면적 및 자료구조 등 여러 측면에서 상이한 세 기관의 자료를 일관성 있게 통합하는 자료 처리 과정이 중요하였다. NFI 자료는 전국에 분포한 표본점 중 지리산국립공원 경계로부터 500 m 이내에 있는 34개를 선택하였다. NFI의 4개의 부표본점으로 구성된 집락표본점을 중점에 위치한 하나의 표본점으로 취급하였으며, 일부 부표본점이 산림

Table 1. Three sources of forest survey reference data.

Survey Name	Source	Surveyed Periods	Number of Plots	Plot Size	Sampling Method
<i>The Seventh National Forest Inventory</i>	Korea Forest Service	2016~2020	34	0.16 ha (Circular subplots, 0.04 ha × 4)	Systematic sampling (4 km × 4 km)
<i>The Fourth National Park Resource Survey</i>	Korea National Park Research Institute	2019	40	0.04 ha (20 × 20 m)	Spatial balanced sampling
<i>The Third Permanent Plots Survey</i>	Seoul National University Nambu Forest	2018~2019	37	0.1 ha (20 × 50 m)	Cluster sampling

Table 2. Predictor variables for reference (forest survey plots) and target (raster) data.

Category	Variable	Source for reference data	Source for target data
Vegetation	Vegetation class	Forest surveys (NFI, NPRS, SNUF)	National Institute of Forest Science, Forest type map (1:5,000)
	Height class	"	"
	Diameter class	"	"
Climate	Maximum temperature	Overlay of plot location with MK-PRISM	Korea Meteorological Administration, MK-PRISM v.2.1
	Minimum temperature	"	"
	Precipitation	"	"
	Photosynthetic active radiation (PAR)	Overlay of plot location with Solar resource map	Korea Meteorological Administration, Solar resource map
Topographic	Elevation (DEM)	NASA Shuttle Radar Topography Mission (SRTM)	NASA Shuttle Radar Topography Mission (SRTM)
	Slope	Derived from DEM	Derived from DEM
	Aspect	"	"
Location	Latitude	Forest surveys (NFI, NPRS, SNUF)	Center of 100 m grid cell for raster files
	Longitude	"	"

이외의 토지에 위한 경우 나머지의 평균적인 정보를 이용하였다. 각 표본점에서 수종별로 단위면적당 흉고단면적을 산출하기 위하여 NFI의 기본조사원(0.04 ha)과 흉고직경 30 cm 이상의 대경목을 조사하는 대경목조사원(0.08 ha)을 구분하여 계산하였다. 또한, 임분 내 대표성이 높은 수종만 포함하기 위해서 수관급이 우세목, 준우세목, 중층목에 해당하는 임목만 선택하고, 피압목, 폭목은 분석에서 제외하였다. NPRS 자료는 지리산 국립공원 내에 40개의 표본점에서 교목층(수고 >8 m), 아교목층(2~8 m), 관목층(<2 m), 초본층으로 구분된 정보를 제공하며, 이 중 교목층과 아교목층만 이용하였다. 20 m × 20 m의 교목층 조사구와 10 m × 10 m 사분면의 아교목층 조사구를 구분하여 단위면적당 흉고단면적을 계산하였다. SNUF의 자료는 지리산 반야봉 이하 남서쪽 영역에 집중적으로 분포한 37개의 연구표본구에서 층위 구분 없이 조사되었다. NPRS와 SNUF 자료는 흉고직경 6 cm 이하의 임목을 포함하는데, NFI의 기준에 맞추어 흉고직경 6 cm 이상의 임목만 선택하였다.

2) 목표자료(target data) 처리

목표자료를 구축하고자 다양한 출처의 래스터 자료를 수집하였다. 임상도는 벡터 기반 자료이므로 격자 형태의 래스터 자료로 변환하였다. 식생변수 항목이 참조자료와 일치하도록 대축척 임상도(1:5,000, 이하 임상도)의 임상, 임분고, 경급 속성을 재분류하였다(Table 2). 연구지역의 26개 임상을 신갈나무림, 참나무림, 활엽수림, 침엽수림, 침활혼효림의 5개 유형으로 단순화하였다. 임분고는 2 m 단위로 구분된 기존 자료를 참조자료와 동일한 3개 범주로 재분류하였다. 경급은 소경목, 중경목, 대경목으로 구분된 기존자료의 속성을 그대로 이용하였다.

지형변수는 NASA SRTM 30 m 수치표고모형 자료(Farr et al., 2007)를 100 m 해상도로 리샘플링하여 고도, 경사, 사면향을 계산하였다. 위경도는 격자 중심점의 좌표를 추

출하였다. 최고기온, 최저기온, 강수량은 기상청의 2000~2019년 기후 자료를(Kim et al., 2022), 광합성유효광량(PAR)은 기상청 태양광자원지도(Kim et al., 2020)의 2016년 7월~2021년 6월의 평균 경사면 일사량을 각각 격자별로 평균 내어 이용하였다.

3) 랜덤 포레스트 모델링

랜덤 포레스트는 무작위로 구성된 의사결정나무들을 결합해 예측 성능을 높이는 앙상블 기법이며(Breiman, 2001), 이 연구에서는 식생, 기후, 지형, 위치에 관한 총 12개 변수를 모형 학습에 활용하였다(Table 2). 이 연구는 다중 반응변수 예측 기능이 추가된 R yaImpute 패키지(Crookston and Finley, 2008)의 ‘변형 랜덤 포레스트’를 활용했으나, 간략히 ‘랜덤 포레스트’로 지칭하고 있다. 참조자료의 조합에 따라 총 7개의 랜덤 포레스트 모형을 구성하고, 표본점 할당을 통해 7개의 표본점 고유번호의 래스터 자료를 획득하였다(Table 3). 모형 1(NFI), 모형 2(NPRS), 모형 3(SNUF)은 각각 34, 40, 37개의 표본점을 가지는 단일 자료를 기반으로 하였으며, 다기관 자료를 통합한 모형인 모형 4(NFI+NPRS), 모형 5(NFI+SNUF), 모형 6(NPRS+SNUF), 모형 7(NFI+NPRS+SNUF)은 각각 73, 71, 77, 109개의 표본점을 활용하였다. 계산 부담과 정확도를 최적화하는 사전 분석을 통해 각 모형이 식생변수마다 100개씩, 총 300개의 의사결정나무를 학습하도록 구성하였다(Rigatti, 2017).

이후, 학습된 참조자료의 식생변수를 반응변수로, 목표자료의 12개 변수를 예측변수로 설정해 모든 격자점에 가장 유사한 표본점을 예측하였다. 개별 의사결정나무가 산출한 표본점 후보 중 가장 여러 번 선택된 고유번호가 최종적으로 할당되며, 이를 통해 생성된 고유번호 래스터는 산림조사 자료와 연계되어 각 격자점에서 수종, 흉고직경, 수고, 밀도 등 구체적인 임목정보를 활용할 수 있다(Riley et al., 2016).

3. 모형 성능 검증 및 최적 모형 선정

모형 성능 검증에 앞서, 주요 입력자료인 산림조사 자료와 임상도 간의 일치율을 평가하였다. 모형의 현장 재현 능력은 임상도의 품질에 좌우되므로(Riley et al., 2021), 산림조사와 해당 지점의 임상도의 값을 서로 대조해보며 해당 지역의 모델링 난이도를 평가하고자 하였다. 평가는 임상, 임분고, 경급의 세 식생변수를 대상으로 점대점(point-to-point)과 점대면(point-to-area) 방식으로 진행하였다. 점대점 비교에서는 표본점과 해당 위치의 격자점 간의 일치율을 산출했고, 점대면 비교에서는 표본점의 정보

Table 3. Seven models and their reference data sources.

Model	Reference data source
Model 1	NFI
Model 2	NPRS
Model 3	SNUF
Model 4	NFI+NPRS
Model 5	NFI+SNUF
Model 6	NPRS+SNUF
Model 7	NFI+NPRS+SNUF

가 이웃한 3 × 3 범위 격자 영역에서 발견되는지를 기준으로 일치율을 구하였다. NFI 부표본점 등의 면 단위 조사 자료를 점 자료로 취급하거나, 벡터 기반자료인 임상도의 경계를 격자자료로 변환하는 과정 등은 공간적 오차를 유발할 수 있다. 따라서, 점대면 비교를 통해 모형 불확실성에 대한 완화된 평가 관점을 추가적으로 제공하였다.

입력자료에 따른 7개 모형의 성능을 두가지 방식으로 검증하였다. 첫째로, 모형 예측 결과를 임상도와 격자 단위로 비교하여 연구지역 전체에 걸친 분류 능력을 평가했으며 식생변수별 F1 점수를 주요 지표로 활용하였다. F1 점수는 식 1과 같이 정밀도(precision)와 재현율(recall)의 조화평균으로 계산되며 모형의 분류 성능이 우수하면 값이 1에 가깝고, 저조하면 0에 가깝다. 범주 간 자료 수가 불균형할 경우, 정확도(accuracy)만으로는 모형의 성능을 왜곡할 수 있으나 F1 점수는 비교적 보수적으로 성능을 평가하기에 식생 분류, 토지피복 분류 연구에서 널리 활용된다(Baek et al., 2021; Lee et al., 2022). 이에 더하여, 모형 예측 결과와 임상도의 공간적 분포 패턴과 분포 비율을 비교하였다.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1)$$

둘째, 모형 예측 결과를 산림조사 자료와 비교해 기관별 자료로 학습한 모형의 재현 능력을 살폈고, 점대점 및 점대면 비교 방식을 적용해 일치율을 계산하였다. 이때 모든 111개 표본점에서의 일치율과 모형 학습에 사용된 일부 표본점에 대한 일치율을 구분하여 평가하였다.

위의 평가 결과를 종합하여 모형 간 성능 차이를 분석한 뒤, 전체 예측 정확도와 재현 능력이 가장 뛰어난 모형을 최적 모형으로 선정하였다. 모든 공간분석에는 R(v.4.4.2, R Core Team, 2024)과 ArcGIS Pro(v.3.3.2)를 사용하였다.

결과 및 고찰

1. 지리산 전역에 표본점 할당

입력자료 조합에 따른 모형 7개로부터 표본점 정보가 할당된 7개의 지리산 식생 지도가 생성되었다(Table 4). 모형 4와 모형 7은 일부 표본점을 할당 과정에서 제외하였으며, 나머지 모형은 학습에 이용된 모든 표본점을 최소 한번 할당하였다. 학습에 사용된 표본점의 종류가 많을수록 개별 표본점의 할당 횟수는 낮아졌다. 단일 자료만 사용한 경우에는 특정 표본점이 과도하게 대표될 가능성이 높지만, 다기관 자료 통합 모형에서는 할당된 횟수가 상대적으로 균일하여 개별 표본점의 영향력이 상대적으로 낮아졌다.

2. 모형 성능 검증 및 최적의 지리산 식생 모형 선정

1) 산림조사 자료와 임상도의 일치율

산림조사 자료와 임상도의 일치율은 전반적으로 낮았으며(Table 5), 이는 해당 지역의 높은 모델링 난이도를 나타낸다. 점대점 비교에서는 임상, 임분고, 경급의 일치율이 각각 34.2%, 42.3%, 46.8%이었으며, 점대면 비교에서는 각각 61.8%, 57.7%, 55.9%로 더 높았다. 임상도는 천연림 구획의 최소기준을 0.5 ha 이상으로 설정하지만, 산림조사는 0.04~0.16 ha에서 수행되었으므로 공간적 규모의 차이

Table 5. Reference data (Forest surveys) vs. target data (Forest type map). Point-to-point compares each pixel on a one-to-one basis, whereas point-to-area applies a 3×3 neighborhood tolerance.

Variable	Point-to-point (%)	Point-to-area (%)
<i>Vegetation class</i>	34.2	61.8
<i>Height class</i>	42.3	57.7
<i>Diameter class</i>	46.8	55.9

Table 4. Summary of the number of plots used and the imputed cell counts per plot for models 1 to 7. The percentages are relative to the total vegetated areas (75,858 cells).

Model	Number of plots used	Imputed cell counts per plot		
		Mean±SD	Minimum	Maximum
<i>Model 1</i>	34	2,231.7 ± 2,513.1	36 (0.0%)	8,529 (11.2%)
<i>Model 2</i>	40	1,897.0 ± 2,166.3	90 (0.1%)	8,700 (11.5%)
<i>Model 3</i>	37	2,050.8 ± 2,315.8	62 (0.1%)	7,371 (9.7%)
<i>Model 4</i>	73	1,025.4 ± 1,528.3	0 (0.0%)	6,469 (8.5%)
<i>Model 5</i>	71	1,084.0 ± 1,595.3	10 (0.0%)	8,133 (10.7%)
<i>Model 6</i>	77	985.4 ± 1,528.7	1 (0.0%)	7,666 (10.1%)
<i>Model 7</i>	109	709.1 ± 1,117.4	0 (0.0%)	5,725 (7.5%)

가 낮은 일치율의 원인 중 하나일 것이다(Choi et al., 2024). 이때, 두 평가방식의 큰 일치율 차이는 지리산의 산림 환경이 복잡하고 임상도의 공간적 패턴의 이질성이 높다는 것을 보여준다.

2) 임상도에 대한 모형 성능 검증

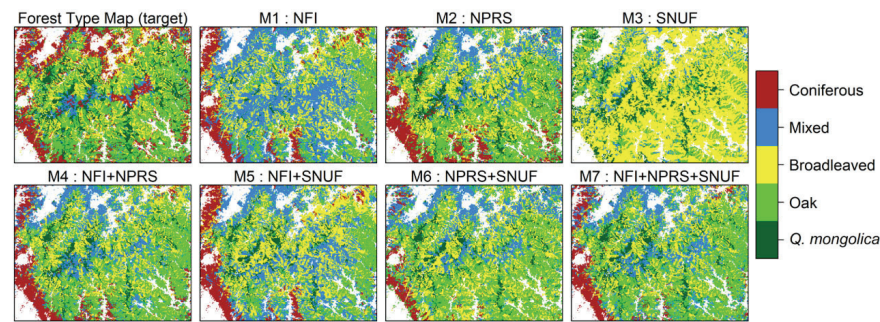
모형의 임상도에 대한 예측 성능을 F1 점수로 평가한 결과, 세 기관의 자료를 모두 결합한 모형 7이 임상(0.66), 임분고(0.81), 경급(0.65) 모두에서 가장 우수한 성능을 보였다(Table 6). 그러나 최저점의 경우 임상은 SNUF 모형의 0.34, 임분고는 NFI 모형의 0.55, 경급은 NPRS 모형의 0.41로 나타나, 단일 자료 모형 간에 평가항목마다 예측 성능의 차이가 있음을 보여준다. 전반적으로, 다기관 자료를 통합한 모형이 단일 자료 모형에 비해 각 변수의 예측 성능이 우수했다.

이러한 경향은 모형 예측 결과의 공간적 분포 패턴에서도 확인된다(Figure 2). 단일 자료 모형에서는 기관별 조사 자료의 특성이 크게 반영되었으나 여러 자료를 활용한 모

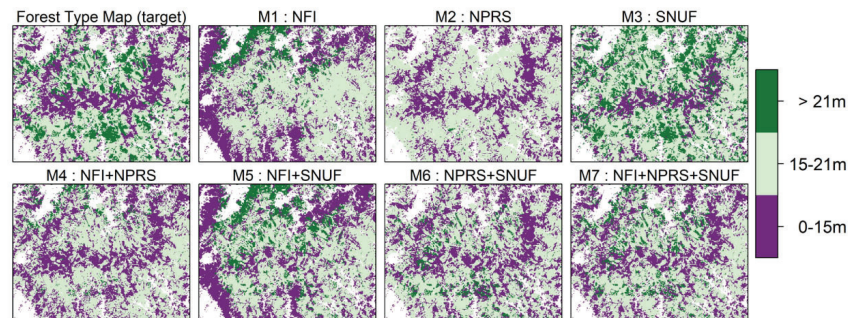
Table 6. Model predictions vs. target data (Forest type map).

Model	F1 score		
	Vegetation class	Height class	Diameter class
Model 1	0.45	0.55	0.53
Model 2	0.57	0.58	0.41
Model 3	0.34	0.75	0.61
Model 4	0.64	0.68	0.59
Model 5	0.48	0.64	0.62
Model 6	0.56	0.77	0.63
Model 7	0.66	0.81	0.65

(a) Vegetation class



(b) Height class



(c) Diameter class

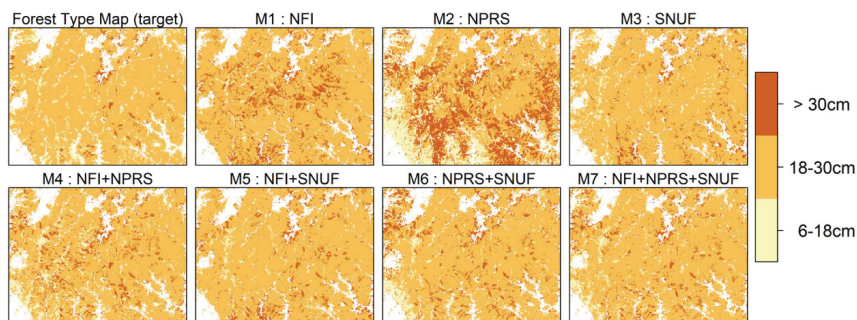


Figure 2. Forest attribute maps derived from the target image (Forest type map) and outputs from models 1 to 7: (a) Vegetation class, (b) height class, and (c) diameter class.

형에서는 조사 자료의 특성이 완화되어 임상도와 유사한 패턴이 나타났다(Yoccoz et al., 2001; Zipkin and Saunders, 2018). 단일자료가 예측한 임상의 공간적 패턴을 보면, NFI는 침활혼효림, NPRS는 참나무림, SNUF는 활엽수림을 지리산의 주요 임상으로 예측하였다. NFI와 NPRS 모형은 지리산국립공원 경계 주위에 분포하는 소나무림을 나타내는 저지대 침엽수림 분포를 잘 예측하였으나, SNUF 모형은 저지대 식생을 잘 예측하지 못하였다. 임상도는 지리산 천왕봉과 반야봉 부근의 아고산침엽수림 분포를 잘 드러내고 있으나, 표본점 할당 모델링은 고지대 침엽수를 예측하지 못하였다. 이는 모든 자료에 고지대 조사가 부족하기 때문에 나타난 현상이다.

모형 예측 결과와 임상도의 임상, 임분고, 경급의 비율을 비교한 결과, 입력자료에 따른 차이가 정량적으로 드러났다(Figure 3). 단일 자료 모형을 비교할 때, NFI 모형은 침활혼효림을, SNUF 모형은 활엽수림을 과대 추정하였으며, NPRS 모형이 임상도와 가장 유사한 분포를 도출하였다. 이러한 차이는 NPRS 조사가 398~1,680 m의 가장 넓은 고도 범위에서 수행된 반면, SNUF는 622~1,350 m의 제한된 고도 범위의 산림을 집중적으로 조사한 결과로 해석된다. NFI는 481~1,477 m의 고도 범위에서 수행되었으나, 연구에 활용된 34 개 조사지점 중 6 개가 지리산국립공원 경계 외부에 위치하여 저지대의 소나무 및 낙엽송 군락지를 다수 포함한 점이 영향을 미친 것으로 파악된다(Figure 1).

모든 모형은 임상도에 비해서 신갈나무, 참나무림, 침엽수림의 비율을 과소 추정하고, 활엽수림과 침활혼효림은 과대 추정하였다(Figure 3). 이 연구에서는 임상을 정의할 때 참조자료의 흉고단면적 비율을 기준으로 삼았으나, 임상도는 주로 항공영상의 수관 면적 비율을 이용한다. 수관 상층을 차지하지 못하는 비우점 활엽수종의 비율은 흉고 단면적을 기준으로 할 때 더 높게 계산된다. 임상을 정의하는 기준의 차이로 인한 오차를 개선하기 위해서는, 향후 산림조사에서 수관 면적에 대한 정보를 파악하거나 기존 자료로부터 수관 면적을 추정할 수 있어야 한다.

임분고의 경우, NPRS 모형은 15~21 m 범위, SNUF 모형은 21 m 이상 범위의 식생 비율을 과대 추정하였으며, NFI 모형은 임상도와 비교적 유사한 비율을 도출했다. 다른 모형이 고도가 높을수록 임분고를 낮게 예측하는 것과 달리, NFI 모형은 저지대에서 더 높은 임분고를 예측한 것이 예외적이었다(Figure 2). 경급에서는 NFI 모형이 소경목 비율은 낮고 대경목 비율은 높은 것에 반해, NPRS 모형은 소경목과 대경목 비율이 모두 높고 중경목 비율은 낮게 나타났다. 임분고의 오차는 각 기관의 수고 측정 방식의 차이에서, 경급의 오차는 조사지점 선정의 차이에서 주로 비롯되었을 수 있다.

3) 산림조사 자료에 대한 모형 성능 검증

모든 111개 표본점에서 모형 예측 결과와 산림조사 자료 간의 일치율을 평가한 결과, 세 기관의 자료를 모두

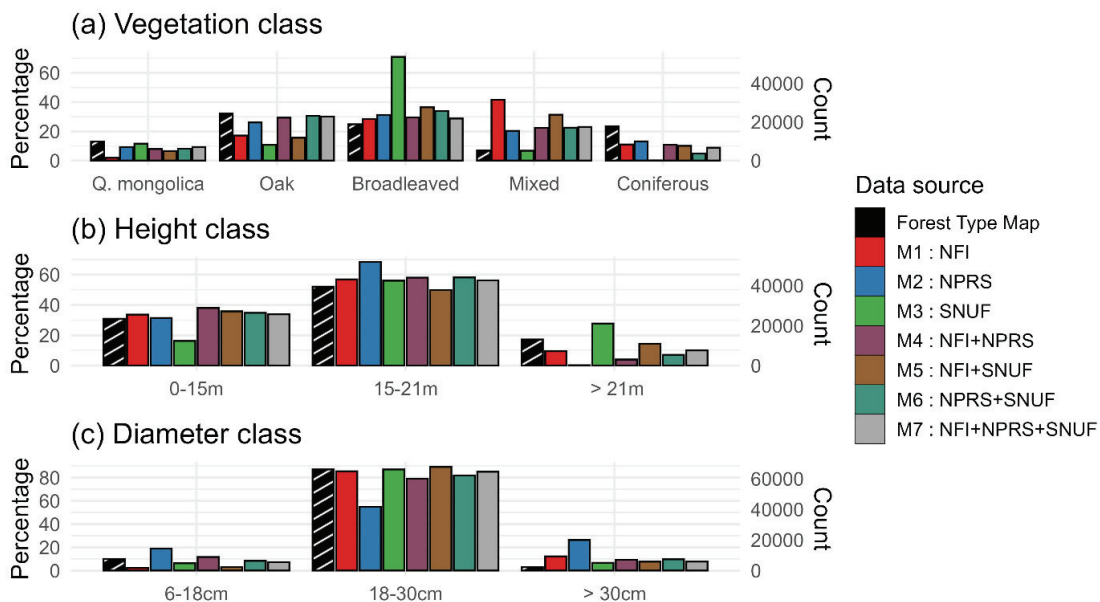


Figure 3. (a) Vegetation, (b) height, and (c) diameter distributions: Forest type map (target) vs. predictions for models 1 to 7 with percentages (left axis) and counts (right axis). Missing classes were assigned a count of 1.

결합한 모형 7이 전반적인 일치율이 가장 높았다(Table 7). 임상, 임분고, 경급 각각에서 점대점 61.3%, 62.2%, 63.1% 및 점대면 82.9%, 79.3%, 75.7%의 일치율을 기록하였다. 단일 자료 모형 중에서는 NPRS 모형이 모든 변수에서 가장 우수한 성능을 보였다. 그러나 입력자료의 수가 많을수록 모형에 미치는 영향이 커지므로, 표본점 수가 많은 NPRS 자료가 평가에서 유리함을 인지해야 한다. SNUF 모형은 일부 지역에 조사 자료가 한정되어 다른 기관 조사 자료의 임상을 상호 예측하는데 어려움이 있었다.

학습에 사용된 표본점에서의 일치율을 별도로 평가한 결과, 전반적으로 61.3% 이상의 수준을 보였으나, 경향은 전체 표본점에 대한 재현 능력과 상반되었다(Table 8). 단일 자료 모형은 일치율이 76.5% 이상으로 대부분 높았으나, 여러 자료를 결합할수록 특정 지점에 대한 국지적 예측 정확도가 낮아졌다. 즉, 다기관 자료를 활용하면 전체 지역에 대한 예측 능력은 향상되지만, 실제 산림조사 지점에 정확한 값을 할당하는 능력은 떨어질 수 있음을 의미한다. 이러한 오차가 발생할 가능성은 임상도와 산림조사 자료가 불일치하거나, 자료가 일부 지역에 편중되어 있을 경우, 기관

별 측정 방식에 차이가 있을 경우에 증가할 수 있다(Roberts et al., 2017). 따라서, 산림조사 단계에서부터 공간적 표본 분포를 균형 있게 확보하는 전략이 중요하며, 모델링 과정에서도 단순한 자료 결합에만 의존하기보다 기관별 측정 편차를 보정하는 절차를 거쳐야 한다(Morrison, 2016). 표본점 할당 모델링에서는 전체 예측 성능뿐 아니라 입력자료의 다양성과 지점별 공간 분포가 국지적 오차에 미치는 영향을 종합적으로 고려할 필요가 있다.

4) 최적의 지리산 식생 모형 선정

최종적으로 선정된 모형은 세 기관의 조사 자료를 모두 결합한 모형 7로서, 임상도와 산림조사 자료 전체 지점에 대해 모든 식생변수에서 상당한 예측 성능을 보였다(Table 6, Table 7). 다기관 자료 통합은 기관별 자료의 특성을 상쇄하여 임상도의 식생 분포와 산림 구조를 안정적으로 예측하는 데 기여하였다. 단, 학습에 이용된 표본점에서의 국지적 예측에서 오차가 증가할 수 있다는 점에서 모형 해석에 유의해야 한다.

Table 7. Model prediction vs. reference data (forest surveys) at 111 entire plots. Point-to-point compares each pixel on a one-to-one basis, whereas point-to-area applies a 3×3 neighborhood tolerance.

Model	Point-to-point (%)			Point-to-area (%)		
	Vegetation class	Height class	Diameter class	Vegetation class	Height class	Diameter class
Model 1	36.0	36.9	37.8	61.3	60.4	55.0
Model 2	55.0	55.9	49.5	77.5	67.6	72.1
Model 3	27.0	42.3	39.6	54.1	67.6	67.6
Model 4	55.0	62.2	61.3	79.3	73.9	78.4
Model 5	33.3	40.5	33.3	55.9	71.2	68.5
Model 6	55.9	55.0	50.5	72.1	77.5	74.8
Model 7	61.3	62.2	63.1	82.9	79.3	75.7

Table 8. Model prediction vs. reference data (forest surveys) at training plots. Point-to-point compares each pixel on a one-to-one basis, whereas point-to-area applies a 3×3 neighborhood tolerance.

Model	N	Point-to-point (%)			Point-to-area (%)		
		Vegetation class	Height class	Diameter class	Vegetation class	Height class	Diameter class
Model 1	34	76.5	91.2	85.3	94.1	100.0	94.1
Model 2	40	87.5	80.0	82.5	92.5	87.5	92.5
Model 3	37	86.5	78.4	78.4	91.9	86.5	83.8
Model 4	74	70.3	70.3	70.3	87.8	83.8	82.4
Model 5	71	80.3	78.9	71.8	93.0	88.7	81.7
Model 6	77	66.2	67.5	62.3	88.3	79.2	76.6
Model 7	111	61.3	62.2	63.1	82.9	79.3	75.7

3. 표본점 할당 모델링 프레임워크의 활용을 위한 시사점

최종 채택된 모형 7은 임상도와 산림조사 자료에 대해 우수한 예측 성능을 보여, 지리산 식생정보를 높은 정확도로 제공할 수 있다(Table 6, Table 7). 이 모형은 각 지점과 표본점을 직접적으로 연결하기 때문에 산림조사가 제공하는 모든 정보를 필요에 따라 다양한 방식으로 활용할 수 있다는 장점이 있다. 단순하게는 할당된 표본점의 수종 구성을 바탕으로 지리산의 우점 수종 분포도를 작성할 수 있다(Figure 4). 나아가서 밀도, 영급, 구조적 특성 등 다양한 산림 속성을 종합하여 산림관리 전략 수립, 기후변화 대응, 재해 예방, 서식지 평가 등에 관련된 다양한 의사결정을 지원할 수 있다(Riley et al., 2022).

이 연구의 모델링 프레임워크는 향후 국가산림자원조사와 임상도를 활용하여 전국 범위의 정밀 산림공간정보를 구축하기 위한 핵심 기반이 될 수 있다. 전국의 조사 자료를 모두 활용하면 입력자료의 다양성이 증가하여 모형 성능에 긍정적인 영향을 미칠 것이며, 조사 방식과 자료 구조가 표준화되어 있기에 분석이 더욱 용이할 것이다. 다만, 지리산은 1967년 이후 국립공원으로 보호되어 왔으나, 전국 산림은 국가주도 녹화사업의 영향으로 식생과 환경 간 관계에 인위적인 간섭이 작용했을 가능성이 있다는 점을 신중히 고려해야 한다(Bae et al., 2012).

상대적으로 좁은 지역의 자세한 식생정보를 얻기 위해 이 프레임워크를 적용할 경우, 해당 지역의 산림 특성을 균질하게 반영하는 충분한 조사 자료 확보가 모형의 예측

정확도 향상을 위해 필수적이다(Table 6). 예를 들어, 고도 범위가 좁아 유사한 생태 조건을 보이는 지역에서는 조사 자료 불균형의 영향이 상대적으로 작아, 소수의 입력자료로도 안정적인 결과를 도출할 수 있다. 그와 달리, 이질적이고 복잡한 산림 생태계에서는 희소한 식생유형이 누락되지 않도록 고도에 따른 층화표본추출법 등의 보완적 전략이 뒷받침되어야 높은 모형 성능을 확보할 수 있다. 이러한 조건이 충족된다면, 이 연구에서 제시한 프레임워크는 전국 및 지역 규모 모두에서 효과적인 산림공간정보 제공에 활용될 수 있다.

다기관 자료의 통합 활용은 모형 성능 향상에 유효한 전략이었으며, 일부 지역에 집중된 SNUF 자료 또한 모형의 성능을 높이는 데 도움이 되었다(Table 6, Table 7). 그러나 자료 통합 과정은 복잡한 전처리를 거쳐야 했으며, 각 기관의 조사 의도와 방식의 차이가 모형 결과에 영향을 미쳤다(Figure 2, Figure 3). 예를 들어, 형태학적으로 유사한 들메나무(*Fraxinus mandshurica*)와 물들메나무(*Fraxinus chiisanensis*)의 경우에 연구지역에서 각각 NFI가 5.3본/ha와 11.7본/ha, 그리고 NPRS가 21.9본/ha와 0.0본/ha로 다르게 보고하고 있다. 이는 잠재적으로 두 기관의 수종 동정에 차이가 있을 가능성을 제기한다. 이렇게 종 분류 기준, 조사자의 숙련도 등의 차이가 모델링에 반영될 수 있기에 주의가 필요하다(Paillet et al., 2015; Traub and Wüest, 2020). 국내 산림 유관 기관이 조사 과정의 투명성을 최대한 확보하고, 기관 간 조사 프로토콜의 표준화를

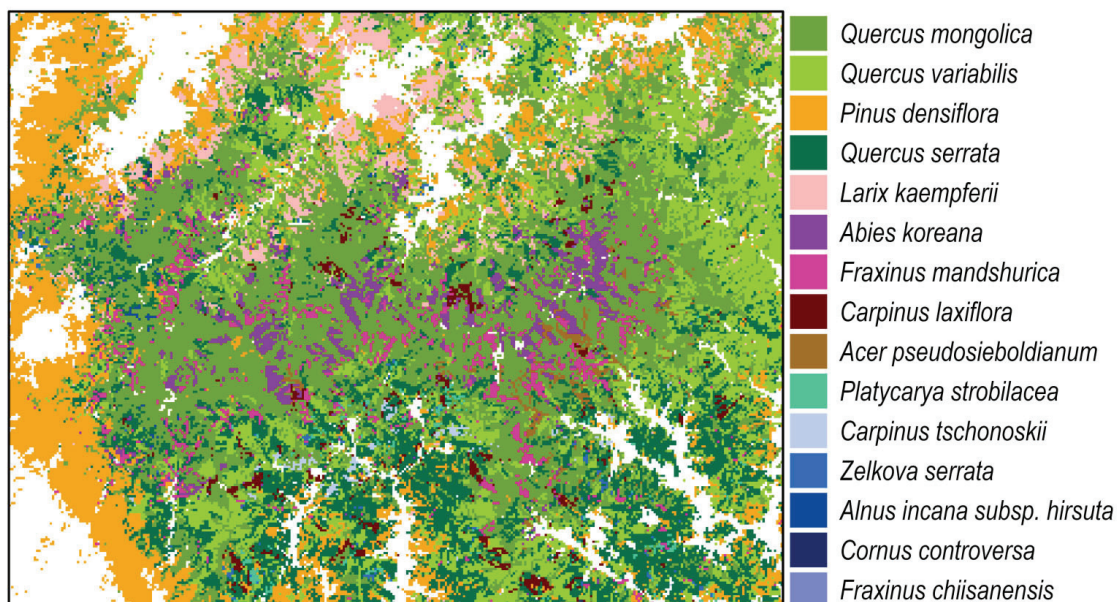


Figure 4. Map of the dominant species in Jiri Mountain National Park derived from the best model, showing 15 species out of 101, with the legend sorted by dominant area.

적극 추진한다면, 보다 유의미한 자료 활용이 이뤄질 것으로 기대된다. 또한, 모델링 연구자는 각 조사 방식에 대한 면밀한 이해를 바탕으로 모형 개발에 접근할 필요가 있다.

더불어, 모형 예측이 임상도의 성능에 크게 의존하는 만큼, 임상도의 지속적인 갱신과 개선이 중요하다. 국지적 재현율 감소 문제를 완화하기 위해 보완적 접근법 도입과 추가 현장 검증을 통한 모형 보정이 고려되어야 하며, 이를 통해 신뢰성 있는 산림공간정보 제공이 가능할 것이다.

결론

이 연구는 산림청, 국립공원연구원, 서울대학교 남부학술림에서 수행한 산림조사 자료를 통합하여, 지리산 일대의 산림 현황을 세밀하게 반영하는 고해상도 산림공간정보를 구축하였다. 이를 위해 기관별 조사 자료 조합에 따른 총 7개의 랜덤 포레스트 표본점 할당 모형을 구성하였고, 입력자료 예측 능력을 다각도로 평가하였다. 그 결과, 세 기관의 자료를 모두 결합한 모형이 예측 정확도가 가장 높아, 산림 환경의 복잡성과 다양성을 광범위하게 포착할 수 있는 최적 모형으로 선정하였다. 이때, 대기관 자료 통합 모형은 학습자료의 국지적인 재현보다는 전체 지역을 균질하게 표현하는 데 유리하였다.

이 연구의 프레임워크는 생물다양성 관리, 산림 재해 대응, 산림 탄소저장량 평가 등에 필요한 종합적인 산림공간정보를 생산할 수 있다. 나아가, 장기적으로는 기후변화에 따른 산림관리 전략 수립과 같은 고차원적 의사결정에도 기여할 것으로 기대한다. 향후 체계적인 현장조사 설계, 기관 간 조사 프로토콜 표준화, 임상도의 주기적 갱신과 현장 검증 등을 통해 표본점 할당 모델링을 적용한 산림공간정보의 신뢰성을 높일 수 있을 것이다.

감사의 글

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(No. RS-2023-00277059), 이에 감사드립니다.

References

- Bae, J.S., Joo, R.W. and Kim, Y.S. 2012. Forest transition in South Korea: reality, path and drivers. *Land use Policy* 29(1): 198-207.
- Baek, W.K., Lee, Y.S., Park, S.H. and Jung, H.S. 2021. Classification of natural and artificial forests from KOMPSAT-3/3A/5 images using deep neural network. *Korean Journal of Remote Sensing* 37(6-3): 1965-1974.
- Breiman, L. 2001. Random forests. *Machine Learning*, 45, 5-32.
- Chen, F., Hou, Z., Saarela, S., McRoberts, R.E., Ståhl, G., Kangas, A., Packalen, P., Li, B. and Xu, Q. 2023. Leveraging remotely sensed non-wall-to-wall data for wall-to-wall upscaling in forest inventory. *International Journal of Applied Earth Observation and Geoinformation* 119, 103314.
- Choi, S.E., Song, C.H., Kim, H.J., Lee, S.J., Yim, J.S. and Lee, W.K. 2024. Spatial data overlap analysis for constructing a land-use and land-use-change matrix: A focus on Gyeonggi-do. *Journal of Climate Research* 15(1): 15-34.
- Chojnacky, D.C., Heath, L.S. and Jenkins, J.C. 2014. Updated generalized biomass equations for North American tree species. *Forestry* 87(1): 129-151.
- Cooper, L. and MacFarlane, D. 2023. Climate-Smart Forestry: Promise and risks for forests, society, and climate. *PLOS Climate* 2(6): e0000212.
- Crookston, N.L. and Finley, A.O. 2008. yalImpute: an R package for kNN imputation. *Journal of Statistical Software* 23: 1-16.
- Farr, T.G. et al. 2007. The shuttle radar topography mission. *Reviews of geophysics*, 45(2).
- Frost, F., McCrea, R., King, R., Gimenez, O. and Zipkin, E. 2023. Integrated population models: Achieving their potential. *Journal of Statistical Theory and Practice* 17(1): 6.
- KFS (Korea Forest Service), 2024. 2024 Statistical yearbook of forestry. Korea Forest Service. Daejeon, Korea. pp. 439.
- Kim, E.S., Jung, B.H., Bae, J.S. and Lim, J.H. 2022. Future prospects of forest type change determined from national forest inventory time-series data. *Journal of Korean Society of Forest Science* 111(4): 461-472.
- Kim, E.S., Jung, J.B. and Park, S. 2024. Analysis of changes in pine forests according to natural forest dynamics using time-series NFI data. *Journal of Korean Society of Forest Science* 113(1): 40-50.
- Kim, J., Sang, J., Kim, M., Byun, Y., Kim, D. and Kim, T. 2022. Future climate projection in South Korea using the high-resolution SSP scenarios based on statistical downscaling. *Journal of Climate Research* 17(2): 89-106.
- Kim, K.M., Roh, Y.H. and Kim, E.S. 2014. Comparison of three kinds of methods on estimation of forest carbon stocks distribution using national forest inventory DB and

- forest type map. *Journal of the Korean Association of Geographic Information Studies* 17(4): 69-85.
- Kim, M., Lee, W.K., Guishan, C., Nam, K., Yu, H., Choi, S.E., Kim, C.G. and Gwon, T.S. 2014. Estimating stand volume *Pinus densiflora* forest based on climate change scenario in Korea. *Journal of Korean Society of Forest Science* 103(1): 105-112.
- Kim, Y.H., Lee, S.S., Choi, H.W., Jung, H.S., Yoon, J.A., Lee, D.W., Kim, D.H. and Lee, S.W. 2020. Development of production techniques on user-customized high resolution weather information. National Institute of Meteorological Science. Seogwipo, Korea pp. 72.
- Kwon, K.C., Han, S.A., Lee, D.K., Jung, I.K., Seo, Y.J., Shin, K.T. and Jeon, C.S. 2022. Site Characteristics and stand structure of *quercus mongolica* forests in the Republic of Korea. *Journal of Korean Society of Forest Science* 111(1): 100-107.
- Lee, Y.K., Lee, J.S. and Park, J.W. 2022. A study on classification of crown classes and selection of thinned trees for major conifers using machine learning techniques. *Journal of Korean Society of Forest Science* 111(2): 302-310.
- Maxwell, C.J., Scheller, R.M., Wilson, K.N. and Manley, P.N. 2022. Assessing the effectiveness of landscape-scale forest adaptation actions to improve resilience under projected climate change. *Frontiers in Forests and Global Change* 5: 740869.
- Miller, D.A., Pacifici, K., Sanderlin, J.S. and Reich, B.J. 2019. The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution* 10(1): 22-37.
- Moon, G.H. and Yim, J.S. 2021. Changes in carbon stocks of coarse woody debris in national forest inventories: Focus on Gangwon Province. *Journal of Korean Society of Forest Science* 110(2): 233-243.
- Morrison, L.W. 2016. Observer error in vegetation surveys: A review. *Journal of Plant Ecology* 9(4): 367-379.
- Nitoslawski, S.A., Wong-Stevens, K., Steenberg, J.W., Witherspoon, K., Nesbitt, L. and Konijnendijk Van Den Bosch, C. 2021. The digital forest: Mapping a decade of knowledge on technological applications for forest ecosystems. *Earth's Future* 9(8): e2021EF002123.
- Paillet, Y., Coutadeur, P., Vuidot, A., Archaux, F. and Gosselin, F. 2015. Strong observer effect on tree microhabitats inventories: A case study in a French lowland forest. *Ecological Indicators* 49: 14-23.
- Park, H.C., Kim, E.O. and Kim, W.C. 2020. A Study on plant community structure based on the fourth national park resource survey plots in Mt. Jirisan National Park. *Korean Journal of Plant Resources* 33(5): 482-500.
- Park, K.E., Chun, J.H. and Choi, H.T. 2024. Suggestions for utilizing climate change impact assessment information for the in-situ conservation of subalpine coniferous forest in Republic of Korea. *Journal of Climate Research* 15(5-2): 901-908.
- Portela, R.D.C.Q., Pires, A.D.S., Braz, M.I.G. and de Mattos, E.A. 2017. Species richness and density evaluation for plants with aggregated distributions: fixed vs. variable area methods. *Journal of Plant Ecology* 10(5): 765-770.
- Povak, N.A., Furniss, T.J., Hessburg, P.F., Salter, R.B., Wigmosta, M., Duan, Z. and LeFevre, M. 2022. Evaluating basin-scale forest adaptation scenarios: wildfire, streamflow, biomass, and economic recovery synergies and trade-offs. *Frontiers in Forests and Global Change* 5: 805179.
- R Core Team 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Rigatti, S.J. 2017. Random forest. *Journal of Insurance Medicine* 47(1): 31-39.
- Riley, K.L., Grenfell, I.C. and Finney, M.A. 2016. Mapping forest vegetation for the western United States using modified random forests imputation of FIA forest plots. *Ecosphere* 7(10): e01472.
- Riley, K.L., Grenfell, I.C., Finney, M.A. and Wiener, J.M. 2021. TreeMap, a tree-level model of conterminous US forests circa 2014 produced by imputation of FIA plot data. *Scientific Data* 8(1): 11.
- Riley, K.L., Grenfell, I.C., Shaw, J.D. and Finney, M.A. 2022. TreeMap 2016 dataset generates CONUS-wide maps of forest characteristics including live basal area, aboveground carbon, and number of trees per acre. *Journal of Forestry* 120(6): 607-632.
- Roberts, D.R. et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8): 913-929.
- Seo, Y., Jung, S. and Lee, Y. 2017. Mapping of Spatial Distribution for Carbon Storage in *Pinus rigida* Stands Using the National Forest Inventory and Forest Type Map: Case Study for Muju Gun. *Journal of Korean Society of Forest Science* 106(2): 258-266.
- Tinkham, W.T., Mahoney, P.R., Hudak, A.T., Domke, G.M., Falkowski, M.J., Woodall, C.W. and Smith, A.M. 2018. Applications of the united states forest inventory and analysis dataset: A review and future directions. *Canadian Journal of Forest Research* 48(11): 1251-1268.

- Traub, B. and Wüest, R.O. 2020. Analysing the quality of swiss national forest inventory measurements of woody species richness. *Forest Ecosystems* 7: 1-11.
- Wilson, B.T., Woodall, C.W. and Griffith, D.M. 2013. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage. *Carbon Balance and Management*, 8: 1-15.
- Yim, J.S., Kang, J.T., Kim, K.M., Lee, S.J., Moon, G.H., Im, J.B. and Ko, C.W. 2020. Integrated application of national forest inventory data and remote sensing technology. National Institute of Forest Science. Seoul, Korea pp. 102.
- Yoccoz, N.G., Nichols, J.D. and Boulmier, T. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16(8): 446-453.
- Zipkin, E.F. and Saunders, S.P. 2018. Synthesizing multiple data types for biological conservation using integrated population models. *Biological Conservation* 217: 240-250.
- Zweifel, R. et al. 2023. Networking the forest infrastructure towards near real-time monitoring—A white paper. *Science of the Total Environment* 872: 162167.

Manuscript Received : March 17, 2025

First Revision : April 4, 2025

Second Revision : April 30, 2025

Accepted : May 7, 2025