


시계열 위성영상과 머신러닝 기법을 이용한 산림 바이오매스 및 배출기준선 추정

이용규 · 이정수 *

강원대학교 산림경영학과

Machine-learning Approaches with Multi-temporal Remotely Sensed Data for Estimation of Forest Biomass and Forest Reference Emission Levels

Yong-Kyu Lee and Jung-Soo Lee *

Department of Forest Management, Kangwon National University, Chuncheon 24341, Korea

요약: 본 연구는 다중시기 위성영상과 머신러닝 알고리즘을 이용하여 준국가수준의 시계열 산림바이오매스량을 추정하였으며, 이를 바탕으로 산림배출기준선 설정하여 비교·분석하였다. 머신러닝기반의 산림바이오매스 추정 모델을 구축하기 위하여 Landsat TM 위성영상과 유럽항공우주국에서 제공하는 Biomass Climate Change Initiative 정보를 이용하였으며, 머신러닝 알고리즘은 비모수 학습모델인 k-Nearest Neighbor(kNN)과 의사결정나무 기반의 Random Forest(RF)를 적용하였다. 또한, 추정된 산림바이오매스량은 Forest reference emission levels(FREL) 자료와 비교하였다. 머신러닝 알고리즘 별 산림바이오매스 추정 모델을 비교해보면, 최적의 kNN 모델과 RF 모델의 Root Mean Square Error (RMSE)는 각각 35.9와 34.41였으며, RF모델이 kNN모델보다 상대적으로 우수하였다. 또한, FREL, kNN, RF 모델 별 산림배출기준선의 기울기는 각각 약 -33천ton, -253천ton, -92천ton으로 설정되었다.

Abstract: The study aims were to evaluate a machine-learning, algorithm-based, forest biomass-estimation model to estimate subnational forest biomass and to comparatively analyze REDD+ forest reference emission levels. Time-series Landsat satellite imagery and ESA Biomass Climate Change Initiative information were used to build a machine-learning-based biomass estimation model. The k-nearest neighbors algorithm (kNN), which is a non-parametric learning model, and the tree-based random forest (RF) model were applied to the machine-learning algorithm, and the estimated biomasses were compared with the forest reference emission levels (FREL) data, which was provided by the Paraguayan government. The root mean square error (RMSE), which was the optimum parameter of the kNN model, was 35.9, and the RMSE of the RF model was lower at 34.41, showing that the RF model was superior. As a result of separately using the FREL, kNN, and RF methods to set the reference emission levels, the gradient was set to approximately -33,000 tons, -253,000 tons, and -92,000 tons, respectively. These results showed that the machine learning-based estimation model was more suitable than the existing methods for setting reference emission levels.

Key words: REDD+, biomass, forest reference emission level, reference emission level, machine learning

서론

기후변화는 인위적인 온실가스 배출에 따른 지구온난화에 의해 가속화되고 있으며, 이를 완화하기 위해 1992년 기후변화협약(United Nations Framework Convention on

Climate Change; UNFCCC)을 시작으로 1997년 교토의정서, 2015년 파리 협약 체결 등 국제적인 노력이 이어졌다 (Kuyper et al., 2018). 특히, UNFCCC 11차 당사국 총회 (Conference of the Parties; COP)는 개발도상국의 산림 전용 및 황폐화 방지(Reducing Emissions from Deforestation and forest Degradation; REDD) 의제를 채택하였으며, 15차 COP에서는 REDD의 산림전용과 산림황폐화 이외에 산림 보존, 지속가능한 산림관리, 탄소량 증대 활동 (Reducing Emissions from Deforestation and forest Degra-

* Corresponding author

E-mail: jslee72@kangwon.ac.kr

ORCID

Jungsoo Lee  https://orcid.org/0000-0003-2770-0528

dation in developing countries, and the role of conservation, sustainable management of forests, and enhancement of forest carbon stocks in developing countries; REDD+)을 채택하였다. 또한, 파리 협약은 협약국이 자발적으로 감축 목표(Nationally Determined Contributions; NDCs)를 설정하여 전세계 모든 국가에게 지구온난화 완화 의무를 부여하였다(Yoo et al., 2021).

최근, 우리나라는 2030 국가 온실가스 감축목표 상향안을 통해 국외 감축 부분을 16.2백만톤에서 33.5백만톤으로 상향 조정하였으며, REDD+ 사업의 확대를 필요성을 제시하였다(2050 Carbon Neutrality and Green Growth Commission, 2021). 또한, 제6차 산림기본계획에 따르면 산림청은 국가온실가스 감축 목표 달성과 신기후체제에 대응하기 위해 사업모델 개발과 2030년까지 10개국에서 REDD+ 사업을 진행하는 것으로 목표로 하고 있다(Korea Forest Service, 2018).

REDD+ 사업의 추진은 크게 준비단계, 이행단계, 확대 단계로 구성된다. 준비단계에서는 사업대상지 선정 등 사업계획과 실적에 대한 객관적이고 과학적인 측정, 보고, 검증의 시스템(MRV) 구축이 필요하며, 이에 따라 대상지의 산림현황에 대한 파악이 필요하다. 뿐만 아니라, 이행 단계에서 타당성 평가와 MRV 시행, 확대 단계에서 크레딧 확보, 탄소시장 거래 등 REDD+ 사업 전반에 있어 대상지 산림에 대한 모니터링은 필수적이다.

산림배출기준선(Forest reference emission levels)은 일정 기간의 산림 전용으로 인한 온실가스 배출량을 의미하며 사업대상지 내 전용되는 산림의 단위 면적 당 탄소저장량 및 전용면적을 이용하여 설정한다. FREL은 REDD+ 사업을 통해 발생한 경제적 인센티브를 입증하는데 활용되며 현재는 각 국가별로 산림배출기준선을 설정하여 제시하고 있다(Lee et al., 2013a; Lee et al., 2013b).

산림청의 REDD+ 실행가이드라인은 산림의 전용 면적과 산림유형 등의 정보를 얻기 위해 위성영상자료의 활용을 제시하고 있다(Lee et al., 2013a). REDD+ 에 관한 연구 동향을 보면, Goussanou et al.(2018)은 탄소량을 정량화하는 산림배출기준선을 설정하기 위해서는 산림지역의 변화 정보와 바이오매스 추정모델 구축이 필요하다고 강조하였다. 또한, 현재 위성영상자료를 이용하여 바이오매스량 추정과 관련된 연구 사례를 보면, 국내에서는 2010년대 초반 kNN 알고리즘을 이용한 연구가 주로 진행되었다(Seo et al., 2012; Yim et al., 2009). 하지만, 머신러닝 분야의 발전으로 최근에는 kNN 뿐만 아니라 높은 성능의 다양한 모델들이 개발되고 있으며, 이에 따라 외국에서는 Artificial neural network(ANN), Support vector regression

(SVR), Random forest(RF) 등 다양한 머신러닝 알고리즘을 적용한 바이오매스량 추정 연구가 진행되고 있다(Mutanga et al., 2012; Vahedi, 2016; Meng et al., 2016).

한편, 현재까지 진행된 국내 및 국외 바이오매스 추정 연구는 현장조사자료를 바탕으로 구축된 데이터가 필요하며, 현장조사자료를 수집하기 위해서는 큰 비용과 시간이 소요되는 실정이다. 따라서, 본 연구는 효율적으로 REDD+ 사업을 계획할 수 있도록 현장조사자료 대신 구축된 바이오매스 공간주제도와 위성영상을 활용하여 머신러닝 기반의 추정 모델을 구축하고, 이를 기반으로 한 산림배출기준선 설정을 목적으로 하였다.

재료 및 방법

1. 연구대상지 및 사용자료

연구대상지는 남아메리카 대륙의 내륙에 위치한 파라과이의 C주로서, 총 면적은 약 1,290천ha로 경기도 면적의 약 1.3배에 해당하며, 파라과이의 동부지역에 위치하고 있다. C주는 한국임업진흥원에서 2013년 조림을 시작으로 지속적인 산림투자가 이루어지고 있어 연구대상지로 선정하였다(Figure 1).

GIS 공간주제도 자료는 유럽우주국(European Space Agency; ESA)에서 제공하는 ESA biomass climate change initiative(Biomass_cci)와 CCI land cover map(CCI LCM)을 이용하였다. Biomass_cci는 공간해상도 100m×100m의 래스터형태로, 2017년도의 전세계의 산림의 지상바이오매스 추정치 분포를 제공하며, CCI LCM은 300m×300m의 래스터 형태로, 1992년부터 2020년까지의 전세계의 토지 피복분포도를 제공한다. Biomass_cci와 CCI LCM은 저해상도의 이미지로 세부적인 모니터링에는 한계가 있지만, 전세계에 대한 정보를 제공하는 데이터로 활용성이 높은 장점이 있다. 본 연구는 국가수준, 준국가수준의 REDD 관련 연구에서 300m급의 공간정보를 이용한 사례를 참고하여 두 개의 데이터를 활용하였다(Nyamari and Cabral, 2021). 대상지 추출을 위해 파라과이의 통계청 (instituto nacional de estadística)에서 제공받은 행정경계정보를 활용하였다.

위성영상자료는 USGS에서 제공하고 있는 2013년, 2017년, 2021년에 촬영된 Landsat 8 OLI의 Collection 2 Level 2 자료를 이용하였다. Collection 2 Level 2 자료는 천정각등을 고려하여 보정이 완료된 표면반사율과 지표 온도 자료를 포함하고 있으며, 본 연구에서는 표면반사율 자료 중 6개 밴드값(Band2, Band3, Band4, Band5, Band6, Band7)을 이용하였다. 또한, Landsat영상 내에 구름지역은

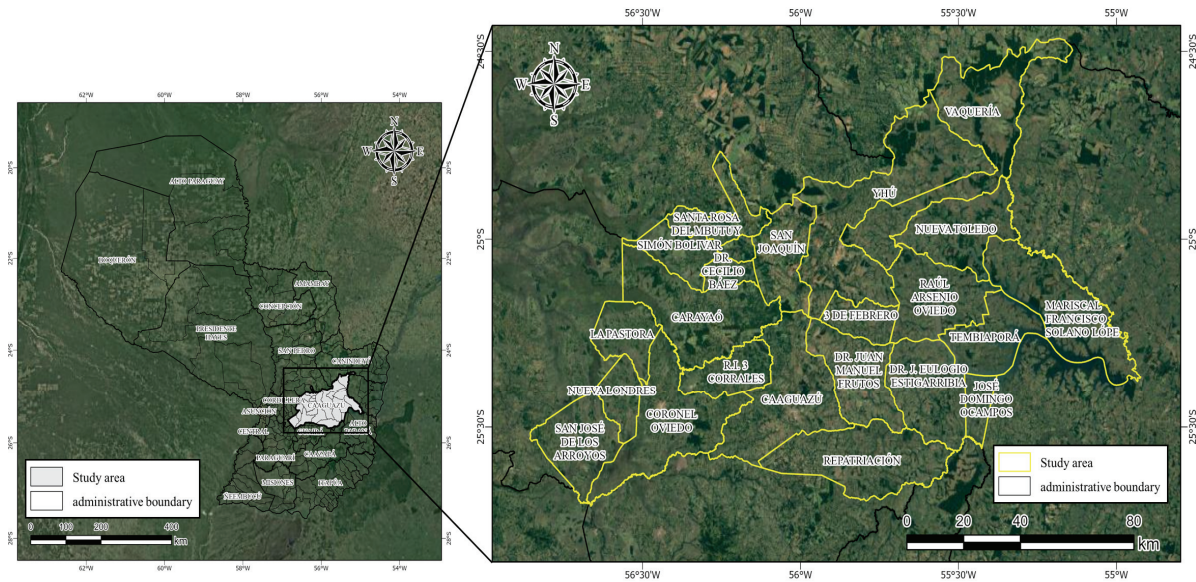


Figure 1. Location of the selected study area.

마스킹 처리하여 노이즈값을 제거하였으며, 위성영상 자료와 Biomass_cci 자료를 해상도가 가장 낮은 CCI LCM 자료를 기준으로 리샘플링하였다.

본 연구를 통해 구축된 추정 모델은 파라과이의 환경과 지속 가능한 개발부(Ministry of Environment and Sustainable Development)에서 UNFCCC에 제출한 FREL 자료와 비교하였다. FREL은 산림배출기준선을 설정하기 위한 산림 유형 별 ha당 바이오매스량 정보를 포함하고 있다. 파라과이는 2016년 FREL을 작성하기 위해 전 국토에 94개의 표본점을 설치하였으며, FREL은 전 국토의 산림 유형을 4개로 구분하여 ha당 바이오매스량 정보를 제공하고 있지만, 2016년 이후에 대한 FREL이나 NFI에 대한 정보는 제공하지 않고 있다(Ministry of Environment and Sustainable Development, 2015)(Table 1).

대상지 분석 및 위성영상 자료 분석은 ESRI 사의 Arcgis 10.1을 이용하였으며, 머신러닝 모델 구축 및 적용은 Python 3.7을 이용하였다.

2. 연구방법

시계열에 따른 토지피복변화량을 산출하기 위해 CCI LCM의 속성정보를 IPCC 기준에 따라 6개의 속성정보로 재분류하였다. 바이오매스 추정모델은 2017년 Landsat 자료와 Biomass_cci 자료를 이용하였으며, k-Nearest Neighbor 과 Random forest 알고리즘을 이용하여 구축하였다. 구축된 추정모델을 2013년, 2021년 Landsat 자료에 적용하여 시계열 바이오매스맵을 구축하였다. 구축된 모델별 바이오매스맵과 FREL을 이용하여 3개의 산림배출기준선을 설정하고 비교하였다(Figure 2).

Table 1. Used data in analysis.

Dataset	Date	Data source
FREL	2015	Ministry of Environment and Sustainable Development
Administrative boundary	2012	Instituto nacional de estadística
CCI land cover map	1992~2020	ESA climate change initiative (CCI)
Biomass_cci	2017	
Landsat OLI	(224/77)	2013.12.25. / 2018.01.05. / 2020.12.12
	(224/78)	2013.12.25. / 2018.01.05. / 2020.12.12
	(225/78)	2013.12.16. / 2017.12.11. / 2021.02.21
	(225/77)	2013.11.14. / 2017.12.11. / 2021.02.21

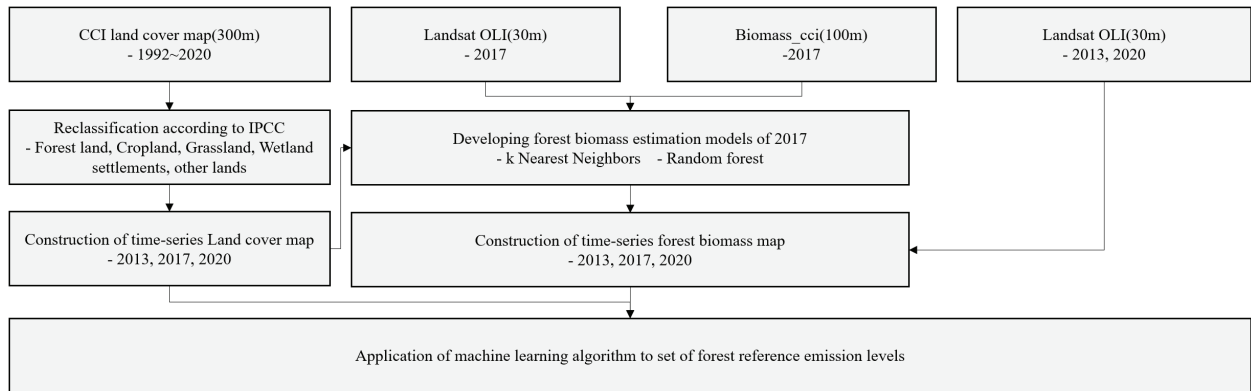


Figure 2. Flowchart of the methodology.

1) 머신러닝 알고리즘을 이용한 산림바이오매스 추정 모델 구축

(1) 데이터셋 구축

CCI LCM는 IPCC 토지이용범주(산림지, 농경지, 초지, 습지, 정주지, 기타토지) 기준으로 재분류하였으며, 2013년부터 2020년까지의 토지피복변화 및 산림면적을 추출하였다(Table 2). 또한, 데이터셋은 산림지를 대상으로 Biomass_cci 값과 Landsat 영상 픽셀값을 추출하여 56,952 개를 구축하였다. 학습 및 검증 데이터는 무작위추출을 통해 10%를 선정하였으며, 선정된 데이터를 각각 학습데이터(70%)와 검증데이터(30%)로 무작위로 구분하였다. 무작위로 선정된 학습데이터와 검증데이터는 머신러닝 모델의 튜닝과정에서 최초 1회에 선정된 데이터를 고정하여 동일하게 사용되었다.

(2) 머신러닝 알고리즘 적용

① k-Nearest Neighbor

k-Nearest Neighbor(kNN)은 데이터의 분류, 회귀 문제를 해결하기 위해 일반적으로 사용되는 알고리즘이다. kNN은 비모수적 학습 모델로 데이터분포에 영향을 받지 않아 다양한 환경의 데이터에서 활용할 수 있는 장점을 가지고 있다. 특히, 산림분야에서의 kNN알고리즘은 회귀모형에 의한 추정치보다 현실성이 있는 추정이 가능한 장점이 있다. kNN에서 선택한 이웃의 수 (k)에 따라 예측에 활용되는 값의 수가 변하게 됨에 따라 모델의 예측성능이 변하게 된다. 따라서 본 연구에서는 k의 값을 1부터 50까지 1씩 증가시키면서 모델을 구축 및 평가하였다. 모델의 평가는 모델의 예측값(\hat{y}_i)과 Biomass_cci 값을 이용하여 평균 제곱근 오차 (Root Mean Square Error; RMSE)를 산출하였으며, RMSE가 가장 낮은 최적의 모델을 선정하였다[Figure (3a)] (식 1).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

② Random forest

Random forest(RF)는 여러 개의 의사결정나무의 조합으로 구성된다. RF는 노이즈가 많은 데이터에 빠르고, 높은 성능 달성할 수 있는 알고리즘으로 평가되고 있다. RF를 구성하고 있는 의사결정나무는 랜덤적인 요소를 이용하여 서로 독립적으로 구축되며, n_estimator, max_depth와 min_samples_split의 하이퍼파라미터에 따라 구축된다. n_estimator는 독립적으로 구축되는 의사결정나무의 개수를 의미한다. 또한, max_depth는 의사결정트리의 최대 깊이를 의미하며, min_samples_split는 노드 분할 시 노드에 남아 있어야 하는 최소의 샘플 개수를 의미한다. 일반적으로 max_depth는 깊어질수록, min_samples_split의 개수는 작을수록 더 세밀한 회귀모델을 구축할 수 있지만, 학습데이터에 과적합 될 수 있는 위험이 있다(Pan et al., 2019). 따라서, 본 연구에서는 max_depth와 min_samples_split을 5부터 40까지 5씩 증가하며 모델을 구축하고 평가하였다. 또한, 최적의 의사결정나무 개수를 탐색하기 위해 n_estimator가 100, 150, 200개인 3종류의 RF 모델을 구축하여, 총 192개의 모델을 구축하고 평가하여 RMSE가 가장 낮은 최적의 모델을 구축하였다[Figure (3b)].

2) 시계열 산림 바이오매스량 추정 및 검증

시계열 산림 바이오매스량은 구축된 바이오매스 추정 모델을 2013년, 2017년, 2021년 Landsat 자료에 적용하였으며, 해당 시기의 토지피복도에서 산림지역의 바이오매스량을 합산하여 추정하였다. 또한, 머신러닝을 이용한 바이오매스 추정모델과 FREL을 비교하기 위해 해당 시기의 토지피복도에서 산림면적과 FREL의 ha당 바이오매스량

Table 2. Reclassification of categories in CCI land cover map.

Reclassified categories	Code	Definition
Forest land	50	Tree cover, broadleaved, evergreen, closed to open (>15%)
	60	Tree cover, broadleaved, deciduous, closed to open (>15%)
	61	Tree cover, broadleaved, deciduous, closed (>40%)
	62	Tree cover, broadleaved, deciduous, open (15-40%)
	70	Tree cover, needleleaved, evergreen, closed to open (>15%)
	71	Tree cover, needleleaved, evergreen, closed (>40%)
	72	Tree cover, needleleaved, evergreen, open (15-40%)
	80	Tree cover, needleleaved, deciduous, closed to open (>15%)
	81	Tree cover, needleleaved, deciduous, closed (>40%)
	82	Tree cover, needleleaved, deciduous, open (15-40%)
	90	Tree cover, mixed leaf type (broadleaved and needleleaved)
	120	Shrubland
	121	Evergreen shrubland
122	Deciduous shrubland	
Cropland	10	Cropland, rainfed
	11	Herbaceous cover
	12	Tree or shrub cover
	20	Cropland, irrigated or post-flooding
	30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)
Grassland	40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)
	100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)
	110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
	130	Grassland
	140	Lichens and mosses
	150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)
	151	Sparse tree (<15%)
	152	Sparse shrub (<15%)
153	Sparse herbaceous cover (<15%)	
Wetlands	160	Tree cover, flooded, fresh or brakish water
	170	Tree cover, flooded, saline water
	180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water
	210	Water bodies
Settlements	190	Urban areas
Other lands	200	Bare areas
	201	Consolidated bare areas
	202	Unconsolidated bare areas
	220	Permanent snow and ice

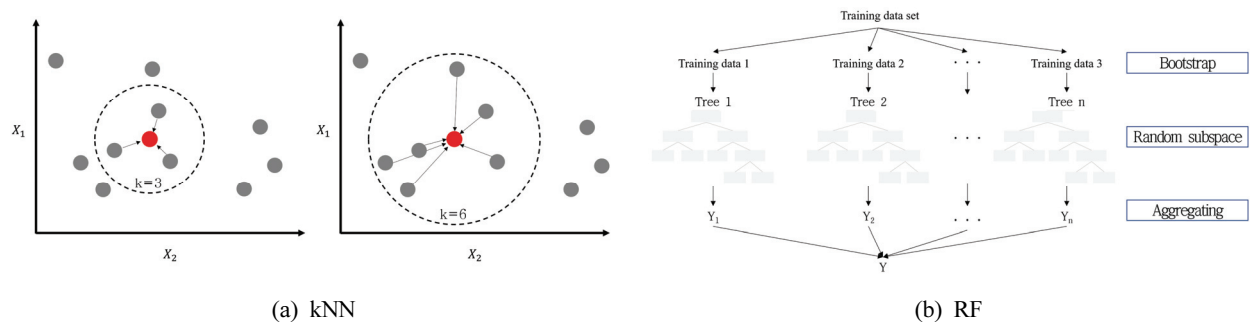


Figure 3. Conceptual map of machine learning algorithms.

Table 3. Distribution of land cover changes using CCI land cover map.

unit : %

Years	2013	2014	2015	2016	2017	2018	2019	2020
Forest land	33.9	33.8	33.8	33.5	33.5	33.7	33.8	33.7
cropland	29.6	29.6	29.6	29.6	29.6	29.5	29.5	29.5
Grassland	32.5	32.5	32.5	32.7	32.7	32.6	32.5	32.5
Wetlands	3.8	3.8	3.8	3.9	3.9	3.9	3.9	3.9
Settlements	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Other lands	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

을 곱하여 산림 바이오매스량을 산출하였다. 추정된 산림 바이오매스량의 검증은 2013년부터 2020년까지의 바이오매스의 변화량을 산출하고 감소한 지역을 Google earth의 시계열 고해상도 위성영상을 이용하여 비교하였다.

3) RF, kNN, FREL을 이용한 산림배출기준선 설정

산림배출기준선 설정은 kNN, RF, FREL 방법 3가지를 통해 산출된 산림 바이오매스량을 기반으로 선형회귀를 이용하였다. 산림배출기준선 설정 방법은 VCS 방법론에 따라 계획되지 않은 산림 벌채로 인한 산림배출기준선 추정 방법을 참고하였다(VCS, 2012). VCS 방법론에 따르면 산림배출기준선 설정 방법은 평균 연간 산림 벌채면적을 사용하는 방법, 선형회귀 추정 방법, 비선형회귀(지수모델, 로그모델) 세 가지 방법이 있으며, 관측 시기가 5개 이하인 경우에는 평균 연간 산림 벌채면적 또는 선형회귀 추정 방법만 활용할 수 있다. 따라서, 본 연구에서는 2013년, 2017년, 2020년의 총 산림 바이오매스량을 기반으로 선형회귀를 통해 산림배출기준선을 설정하고 비교하였다.

결과 및 고찰

1. 시계열에 따른 토지피복변화

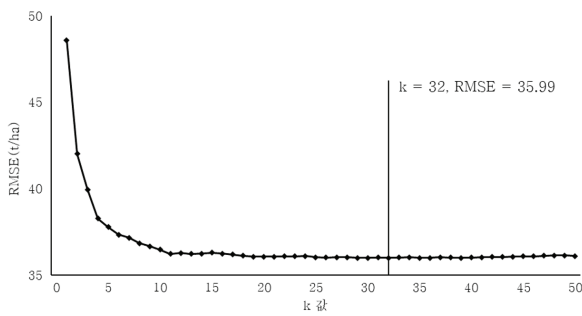
2013년부터 2020년까지의 토지피복변화를 보면, 산림지는 2017년까지 감소하다가 이후에 다시 증가하는 경향을 보였으며, 최근 2020년에는 감소하여 7년간 산림면적이 약 0.2% 감소하였다. 2013년에는 산림지가 전체면적의 33.9%를 차지하였으며, 농경지 29.6%, 초지 32.5%, 습지 3.8%, 정주지 0.2% 순으로 분포하였다. 2020년에는 산림지가 전체면적의 33.7%를 차지하였으며, 농경지 29.5%, 초지 32.5%, 습지 3.9%, 정주지 0.3% 순으로 분포하였다. 따라서, 2013년 이후 연구대상지의 토지피복변화는 산림지와 농경지 면적은 감소하는 반면 정주지, 습지의 면적은 증가하였다(Table 3).

2. kNN과 RF 기반 바이오매스 추정 모델 구축

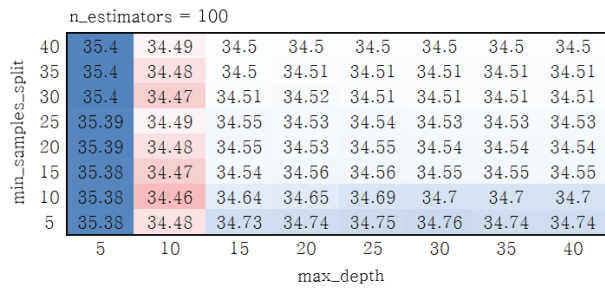
kNN알고리즘 이용하여 바이오매스 추정모델을 구축한 결과는 ha당 바이오매스량의 RMSE는 k=1일 때 48.59로 가장 높았으며, k=32일 때 35.9로 가장 낮게 분포하였다. 구축된 모델은 k의 값이 8이 될 때까지 RMSE 값이 약 12가 감소하여 모델이 크게 개선되었으며, 이후 RMSE 값이 약 36으로 감소 또는 증가가 반복되었다[Figure 4(a)].

RF알고리즘을 이용하여 바이오매스 추정모델을 구축한 결과는 ha당 바이오매스량의 RMSE는 n_estimators = 100, max_depth = 5, min_samples_split = 40일 때 35.4으로 가장 높았으며, n_estimators = 150, max_depth = 10, min_samples_split = 15일 때 RMSE가 34.41로 가장 낮게 분포하였다. RF를 이용한 바이오매스 추정 모델은 하이퍼파라미터에 따라 RMSE의 차이가 약 1 산출되어 kNN 보다 모델에 따른 정확도 차이가 크지 않았다. 하이퍼파라미터에 따른 모델의 정확도를 보면, max_depth는 5일 때 정확도가 낮았으며, 10 이상일 경우 큰 차이를 보이지 않았다. 반면, min_samples_split은 값이 증가할수록 정확도가 개선되는 추세를 보였으나, 최적의 모델은 15일 때 도출되었다[Figure 4(b)-4(d)]. 이는 max_depth와 min_samples_split의 값이 높아지면 학습정확도가 높아지지만, 학습자료의 과적합으로 인해 검증정확도가 낮아질 수 있다는 선행연구 결과와 유사하였다(Ma et al., 2019).

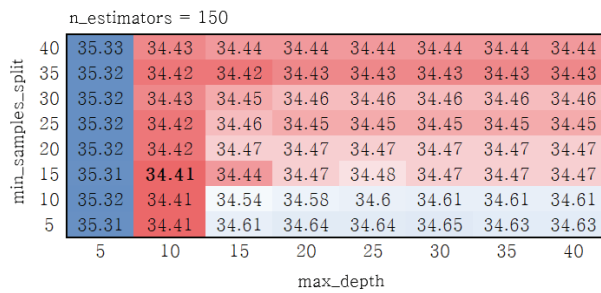
또한, 본 연구에서 추정된 모델의 RMSE는 선행연구의 RMSE 보다 유사하거나 높게 도출되었다. 이는, 기존 선행연구 활용된 바이오매스 추정 모델은 실제 현장조사 데이터를 이용하여 모델을 학습한 반면, 본 연구에서 구축된 바이오매스 추정 모델은 실용적이고 비용효율적인 산림배출기준선 설정을 위해 기 구축된 바이오매스 공간주제도인 Biomass_cci를 활용하여 모델의 학습한 것이 원인으로 생각된다(Wu et al., 2016; Zhao et al., 2016; Li et al., 2020; Jung et al., 2010; Seo et al., 2012; Yim et al., 2009).



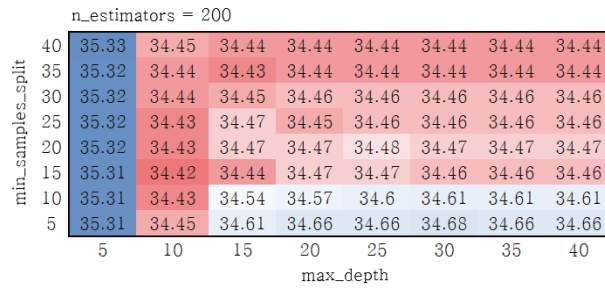
(a) kNN model with RMSE in relation to k value



(b) Heatmap of RF(n_estimators = 100)



(c) Heatmap of RF(n_estimators = 150)



(d) Heatmap of RF(n_estimators = 200)

Figure 4. Evaluation of the performance of RMSE in machine learning models by hyperparameter tuning.

3. 머신러닝 기반 시계열에 따른 산림 바이오매스량 추정 및 검증

구축된 머신러닝 기반 시계열 산림 바이오매스맵을 구축하고 바이오매스량을 산출한 결과, 연구대상지의 총 산림 바이오매스량은 kNN 모델을 이용하였을 때, 2013년 약 3,424만ton, 2017년 약 3,408만ton, 2020년 약 3,238만ton으로 산출되었으며, RF 모델을 이용하였을 때, 2013년 약 3,389만ton, 2017년 약 3,384만ton, 2020년 약 3,322만ton으로 산출되었다. 한편, FREL에 제시된 ha당 바이오매스량을 적용할 경우, 3개 시기에 같은 ha당 바이오매스량이 적용되어 2013년 약 5,746만ton, 2017년 약 5,684만ton,

2020년 약 5,727만ton으로 산출되었다. 각 산출 방법에 따른 산림 바이오매스량은 머신러닝 기반의 추정모델을 이용하였을 때는 시간이 지남에 따라 점차 감소한 반면, FREL을 이용하였을 때는 감소하였다가 다시 증가하는 경향을 보였다. 이는 대상지의 모든 지역에 동일한 ha당 바이오매스량이 적용되기 때문에, 산림지 면적이 감소하였다가 다시 증가한 점이 반영된 것으로 판단된다. 반면, 머신러닝 기반의 바이오매스량의 추정 모델은 지역에 따라 상이한 식생 현황 및 산림 활력도 등이 반영될 수 있기 때문에 산림지 면적이 증가함에도 불구하고 총 바이오매스량이 감소한 것으로 판단된다(Table 4, Figure 5).

Table 4. Comparison of forest biomass using machine learning models and FREL.

	Year	Forest area (ha)	Average of biomass per ha (ton)	Total biomass (thousand tons)
kNN	2013	517,617	66.1	34,239
	2017	512,136	66.5	34,078
	2020	515,979	62.8	32,382
RF	2013	517,617	65.5	33,892
	2017	512,136	66.1	33,836
	2020	515,979	64.4	33,219
FREL	2013	517,617	111	57,455
	2017	512,136	111	56,847
	2020	515,979	111	57,273

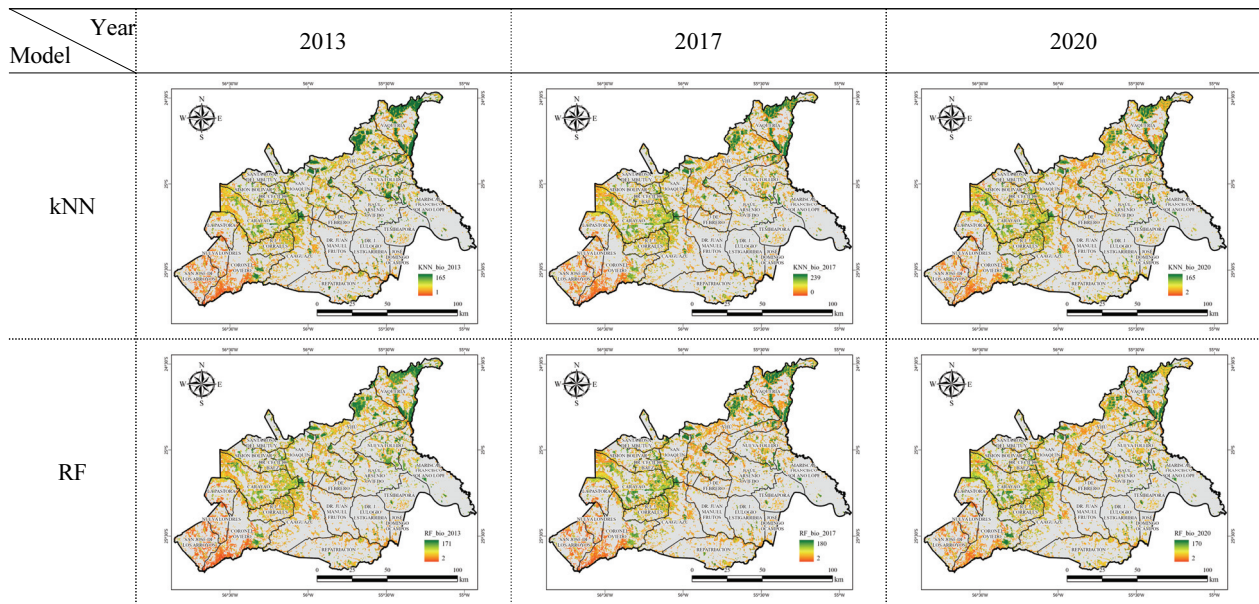


Figure 5. Time-series forest biomass map by machine learning models.

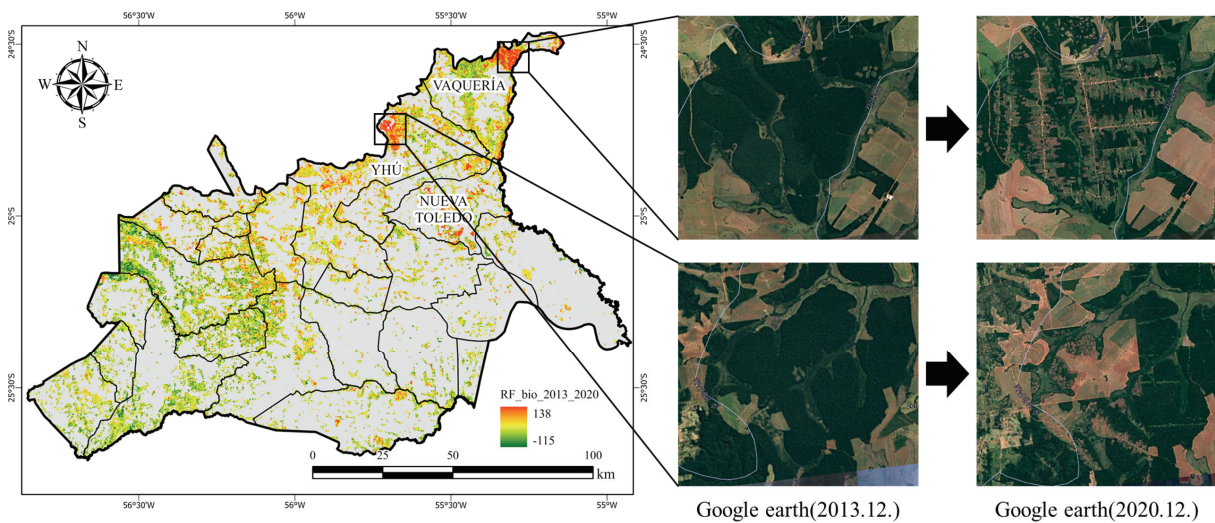


Figure 6. The distribution of forest biomass change area from 2013 to 2020.

시계열 바이오매스맵을 이용하여 바이오매스 변화량을 산출한 결과, ha당 산림 바이오매스량의 평균값이 가장 많이 감소한 지역은 N지역 14.7ton, Y지역 13.0ton, V지역 11.2ton 순이었으며, 위 지역은 모두 연구대상지의 북동쪽에 위치하였다. Google earth의 시계열 고해상도 위성영상을 이용하여 바이오매스가 집중적으로 감소한 지역에 대해 검토한 결과, 해당 지역에서 산림전용 및 황폐화가 진행되었다. 이러한 정보는 향후 산림배출기준선 설정 뿐만 아니라, REDD+ 사업 시 준국가수준의 대상지를 설정하거나, 세부적인 프로젝트 계획을 수립하는 것에 있어 기초 자료로 활용할 수 있을 것으로 판단된다(Figure 6).

4. RF, kNN, FREL을 이용한 산림배출기준선 설정 결과 비교

산림 바이오매스량에 대한 산림배출기준선 설정 결과, 기울기는 kNN, RF, FREL 순으로 높았다. 또한, 머신러닝을 이용하여 설정한 산림배출기준선의 결정계수는 kNN 0.7465, RF 0.7435으로 큰 차이를 보이지 않았으나, FREL을 이용하여 설정한 산림배출기준선의 결정계수는 0.1360으로 머신러닝을 이용한 산림배출기준선의 결정계수보다 약 0.6 낮았다. 따라서, 산림배출기준선 설정 시, FREL 보다 위성영상과 머신러닝을 이용한 방법이 적합하다고 생각된다(Figure 7). 하지만, VCS의 산림배출기준선 설정 방법론에 따르면, 산림배출기준선의 결정계수는 0.75를 기

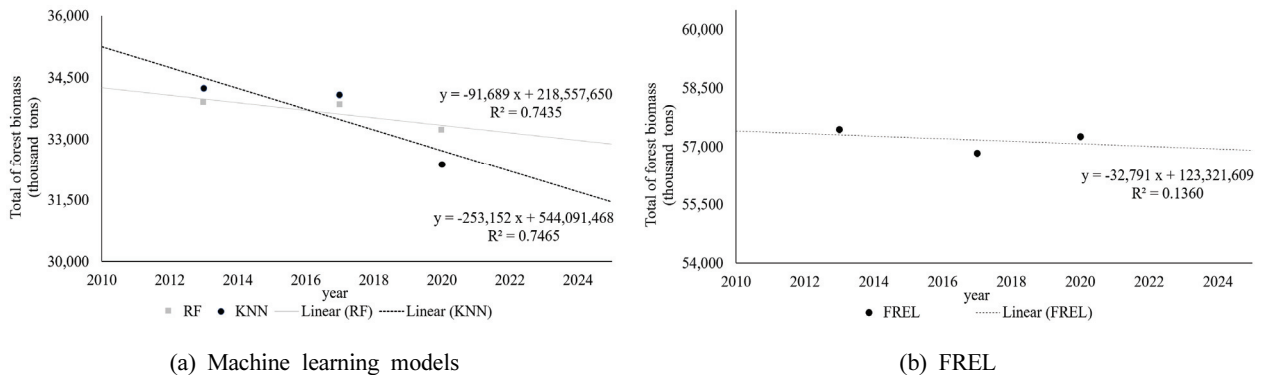


Figure 7. Setting of reference emission levels by machine learning models and FREL.

준으로 하고 있어 산림배출기준선 설정 시 참조 기간을 늘리거나, 산림면적 및 바이오매스 추정의 정확도 향상에 대한 보완이 필요할 것으로 판단된다.

축에 있어 하이퍼파라미터에 따라 모델의 성능이 달라짐에 따라, 하이퍼파라미터 선정에 대한 구체적인 방법론이 필요할 것으로 사료 된다.

결론

감사의 글

본 연구는 REDD+ 사업이 진행된 파라과이의 C주를 대상으로 원격탐사 자료를 활용하여 토지피복도 및 바이오매스 추정모델을 도출하고, 최적의 모델을 이용하여 시계열 바이오매스맵을 작성하여 산림배출기준선을 설정하고자 하였다. 위 방법은 현장조사가 어려운 지역과 대면적에 대한 시계열 바이오매스 변화량을 추정할 때 효과적으로 활용할 수 있으며, 미측정 기간에 대한 바이오매스량을 추정하여 현장조사가 불가능한 과거의 바이오매스량을 추정할 수 있다. 또한, 산림바이오매스량을 산출할 때, FREL 자료를 이용할 경우에는 생육불량지, 산림황폐지 등을 구분하지 않고 획일적인 ha당 바이오매스량이 적용되는 반면, 머신러닝 기반 추정 모델을 이용할 경우에는 산림 현황 및 분광특성 등이 반영된 ha당 바이오매스량이 적용되는 장점이 있다. 따라서, REDD+ 사업 실행 시 머신러닝을 이용한 바이오매스 추정모델을 산림배출기준선 설정에 활용 가능하며, 바이오매스 감소지역을 이용하여 준국가수준의 대상지를 설정하거나 프로젝트 계획 수립에 있어 용이하게 활용될 수 있다. 또한, REDD+ 사업에서 직접적인 성과를 확인할 수 있는 기초적인 자료로 활용이 가능할 것으로 판단되며, 향후 프로젝트 수준에서는 정확도 향상을 위해 현장조사 데이터와 고해상도의 원격탐사자료가 필요하다. 바이오매스 추정모델의 정확도 향상을 위해서는 다양한 질감정보, 지형정보 등을 입력자료로 활용하거나, 이미지 분석에 유용하게 활용할 수 있는 딥러닝 알고리즘 적용이 필요하다. 또한, 모델 구

본 연구는 산림청(한국임업진흥원) 산림과학기술 연구개발사업(2021367A00-2123-BD01)의 지원에 의하여 이루어진 것입니다.

References

2050 Carbon Neutrality and Green Growth Commission. 2021. 2030 Nationally Determined Contribution (NDC) Upgrade Plan. 2050 Carbon Neutrality and Green Growth Commission. Sejong, Korea pp. 23.

Goussanou, C.A., Guendehou, S., Assogbadjo, A. E. and Sinsin, B. 2017. Application of site-specific biomass models to quantify spatial distribution of stocks and historical emissions from deforestation in a tropical forest ecosystem. *Journal of Forestry Research* volume 29(1): 205-213.

Jung, J.H., Heo, J., Yoo, S.H., Kim, K.M. and Lee, J.B. 2010. Estimation of Aboveground Biomass Carbon Stock in Danyang Area using kNN Algorithm and Landsat TM Seasonal Satellite Images. *journal of Korean Society for Geospatial Information Science* 18(4): 119-129.

Korea Forest Service. 2018. Forest Master Plans 6th. Korea Forest Service. Daejeon, Korea. pp. 55.

Kuyper, J., Schroeder, H. and Linner, B. 2018. The Evolution of the UNFCCC. *Annual Review of Environment and Resources* 43: 343-368.

Lee, W.K., Son, Y.H., Yang, S.R., Kim, O.S., Lim, S.T., Yoo, S.M., Kim, J.Y., Kwak, H.B. and Kim, M.I. 2013a.

- REDD+ Implementation Guidelines. Korea Forest Service. Daejeon, Korea. pp. 1.
- Lee, W.K., Son, Y.H., Yang, S.R., Kim, O.S., Lim, S.T., Yoo, S.M., Kim, J.Y., Kwak, H.B. and Kim, M.I. 2013b. REDD+ Introduction. Korea Forest Service. Daejeon, Korea. pp. 33.
- Li, Y., Li, M., Li, C. and Liu, Z. 2020. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports* 10: 1-12.
- Ma, J., Ding, Y., Cheng, J.C., Tan, Y., Gan, V.J. and Zhang, J. 2019. Analyzing the leading causes of traffic fatalities using XGBoost and grid-based analysis: a city management perspective. *IEEE* 7: 148059-148072.
- Meng, S., Pang, Y., Zhang, Z., Jia, W. and Li, Z. 2016. Mapping aboveground biomass using texture indices from aerial photos in a temperate forest of Northeastern China. *Remote Sensing* 8(3): 230-243.
- Ministry of Environment and Sustainable Development. 2015. Nivel de Referencia de las Emisiones Forestales por Deforestación en la República del Paraguay para pago por resultados de REDD+ bajo la CMNUCC. Ministry of Environment and Sustainable Development. Asunción, Paraguay. pp. 10.
- Mutanga, O., Adam, E. and Cho, M.A. 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation* 18: 399-406.
- Nyamari, N. and Cabral, P. 2021. Impact of land cover changes on carbon stock trends in Kenya for spatial implementation of REDD+ policy. *Applied Geography* 133: 102479.
- Pan, Y., Chen, S., Qiao, F., Ukkusuri, S.V. and Tang, K. 2019. Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees. *Science of The Total Environment* 660: 741-750.
- Seo, H.S., Park, D.H., Yim, J.S. and LEE, J.S. 2012. Assessment of Forest Biomass using k-Neighbor Techniques -A Case Study in the Research Forest at Kangwon National University-. *Journal of Korean Society of Forest Science* 101(4): 547-557.
- Vahedi, A.A. 2016. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. *Biomass and Bioenergy* 88: 66-76.
- VCS. 2012. VCS MODULE VMD0007 Redd Methodological Module: Estimation of Baseline Carbon Stock Changes and Greenhouse Gas Emissions from Unplanned Deforestation (BL-UP). VERRA. Washington DC, U.S.A. pp. 16.
- Wu, C., Shen, H., Shen A., Deng, J., Gan, M., Zhu, J., Xu, H. and Wang, K. 2016. Comparison of machine-learning methods for above-ground biomass estimation based on Landsat imagery. *Journal of Applied Remote Sensing* 10(3): 035010.
- Yim, J.S., Han, W.S., Hwang, J.H., Chung, S.Y., Cho, H.K. and Shin, M.Y. 2009. Estimation of Forest Biomass based upon Satellite Data and National Forest Inventory Data. *Korean Journal of Remote Sensing* 25(4): 311-320.
- Yoo, S.M., Song, C.H., Hong, M.A., Kim, W.J., Kim, J.W., KO, Y.J. and LEE, W.K. 2021. Analysis and evaluation of A/R CDM projects in India for abroad afforestation project. *Journal of Climate Change Research* 12(5): 443-460.
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y. and Yu, S. 2016. Examining spectral reflectance saturation in Landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing* 8(6): 469-495.

Manuscript Received : June 15, 2022

First Revision : September 28, 2022

Second Revision : October 31, 2022

Accepted : November 1, 2022