

人工衛星データを利用した我が国の森林炭素蓄積量を推定するための機械学習モデル開発

代表研究者 加藤 知道 北海道大学 大学院農学研究院 准教授

1 はじめに

森林が炭素 (=CO₂) を、いつ・どこで・どのくらい蓄積しているかを調べることは、地球の気候変動の正確な予測に貢献し (IPCC, 2013)、持続可能社会を形成する上で非常に重要である。特に国土の70%を森林で覆われている我が国の炭素蓄積量は、第二次世界大戦後の拡大造林と木材利用の減少により、森林バイオマスが増加しつづけているため、現在は大きな炭素吸収源であると考えられている (Fang, Kato et al., 2014, *PNAS* など)。そのため、京都議定書では、日本の排出量削減目標6%のうち3.8%を森林に割り当てられているなど、大きな生態系サービスを提供するものと非常に期待されているが、その詳細な時間・空間分布はわかっていない。

それらを克服するために、これまで、人工衛星データを利用した森林炭素蓄積量の推定が、世界の各地域で盛んに試みられている (Saatchi et al., 2008, *Remote Sens. Env.*)。特に、機械学習モデルを利用した方法が発達し始めているが、一方で学習用の地上検証データを入手することが非常に困難なために、研究が実現した地域に限られている。

そこで著者らは、森林地上部構造を強く反映する合成開口レーダー (PLASAR: JAXA) 等の衛星データを入力値とした機械学習モデルを構築し、我が国の森林地上部炭素蓄積量のマップを作成することを目指している。我々はその初めの一步として、高解像度の航空機光検出と測距 (Lidar) データに基づき計算された森林地上部バイオマス量を学習データとし、茨城県のスギ・ヒノキ林および大分県の落葉広葉樹林を対象として、地域スケールの森林地上部バイオマス量の推定を行うランダムフォレスト機械学習モデルの開発を目的とした。

2 森林地上部炭素蓄積量の推定

2-1 茨城県のスギ・ヒノキ林における森林地上部炭素蓄積量の推定

(1) はじめに

茨城県北部 (図1) において、日本国土の18%と10%を占める主要な植林品種であるスギ (Japanese cedar) とヒノキ (Japanese cypress) を対象として森林地上部炭素蓄積量 (地上部バイオマス) の推定を行った。

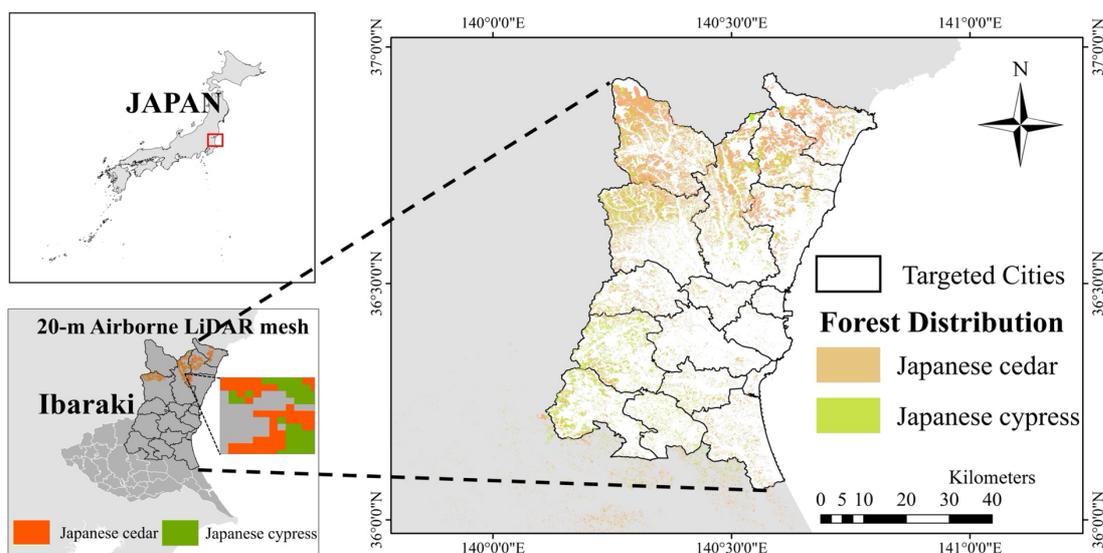


図1. 研究対象 (茨城県北部のスギ (Japanese cedar) とヒノキ (Japanese cypress))

(2) データの整備

データ解析の流れとしては、a) 衛星データ収集と加工（リサンプリング、統一した座標系の適用、フィルタリングと画像くり抜き）、b) PALSAR-2（JAXA）衛星データより景観テクスチャの抜き出し、c) 各種衛星画像からの複数指標の計算（HV & HH, NDVI, EVI）d) 機械学習モデルの構築（最適変数の選別、ハイパーパラメタの調整）、e) 茨城県の自治体における検証となる（図2）。

機械学習モデルは、航空機 Lidar（レーザー）観測に基づいた森林地上部バイオマス（=森林地上部炭素蓄積量）データを学習データとして利用する。本研究では茨城県庁が測量会社に発注した当該航空機観測（2020年7月実施）のデータを、各種手続きを経て入手し利用した。測量会社は、Lidar による波形解析で、地面高(DTM: Digital Terrain Model)と表面高(DSM: Digital Surface Model)の間の差分として、群落高(DCHM: Digital Canopy Height Model)を計算し、樹木個体の頂点での高さを樹高として、あらかじめ地上計測で求められたアロメトリー式（樹高や胸高直径 DBH からバイオマス量を求める経験式）を使用して、個体の材積量（材木生産の対象となる幹の容量）を計算している。対象個体は、スギで 201,854 本、ヒノキで 69,374 本になった（表1）。これら個体材積量に材積密度(WD: Wood density)と拡大係数(BEF: Biomass Expansion Factor)をかけることで地上部バイオマス量（=地上部炭素蓄積量）に変換することができる。これら処理を施し、20m x 20m メッシュサイズの地上部バイオマス量の学習データを得ることができた。

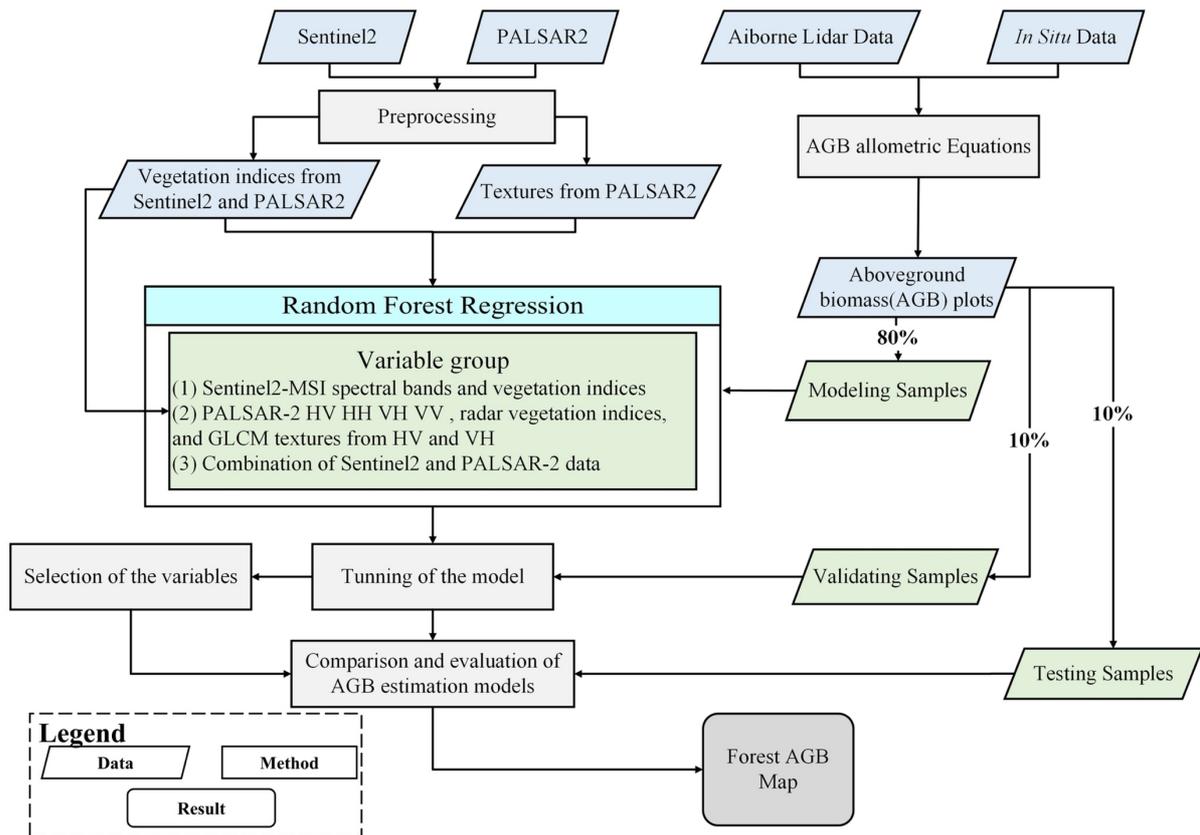


図2. データ解析のフロー. AGB, 地上部バイオマス(=地上部炭素蓄積量); GLCM, gray-level co-occurrence matrix; HH, horizontal transmit–horizontal channel; HV, horizontal transmit–vertical channel; MSI, multispectral instrument; VH, vertical transmit–horizontal channel; VV, vertical transmit–vertical channel.

表 1. 航空機 Lidar から得られた樹木計測に関する統計値

Species	Stand variable	Mean	Standard deviation	Min.	Max.	Sample size
Japanese cedar	Tree height (m)	24.1	5.2	2.1	46.7	201 854
	Diameter at breast height (cm)	24.1	5.3	9.9	78.0	
	Stem volume (m ³ ha ⁻¹)	403.6	170.7	0.3	1516.3	
	Biomass (Mg ha ⁻¹)	155.9	65.9	0.1	585.6	
Japanese cypress	Tree height (m)	19.2	4.3	2.3	39.6	69 374
	Diameter at breast height (cm)	27.0	5.8	10.1	72.0	
	Stem volume (m ³ ha ⁻¹)	585.9	246.2	0.3	1800.0	
	Biomass (Mg ha ⁻¹)	295.7	124.3	0.1	908.4	

機械学習モデルの入力には、森林地上部構造をよく反映すると考えられている日本の宇宙航空研究開発機構 JAXA による Advanced Land Observing Satellite-2 (ALOS-2) 衛星に搭載の合成開口レーダ Phased Array L-band Synthetic Aperture Radar-2 (PALSAR-2) を利用した。4 つの偏光モード (HH, HV, VH, VV) における後方散乱係数 γ^0 を以下のように求めた。

$$\gamma^0 = 10 \log_{10}(DN^2) - CF \quad (1)$$

ここで DN はデジタルナンバー、CF はキャリブレーション係数で -83 になっている。さらに 3x3 ピクセルを対象とした LEE スペックルフィルターを施し、林分上の不均一性によるスペックルノイズを減少させた。その後 3x3 ウィンドウサイズの Gray-level co-occurrence matrix (GLCM) (相対修正ベクター: $d=1, \theta=45^\circ$) における HV と VH のレーダー係数及びテクスチャ (各種統計値: 表 2) を計算した。

表 2. PALSAR データのリスト

Variables (Abbreviation)		Definition
Polarization	HV	Horizontal transmit-vertical channel
	HH	Horizontal transmit-horizontal channel
	VV	Vertical transmit-vertical channel
	VH	Vertical transmit-horizontal channel
Radar Indices	I1 [50]	HH-HV
	I2 [51]	HV+HH
	I3 [52]	(HH-HV)/(HV+HH)
	I4 [53]	HV/HH
	I5 [50]	VH-VV
	I6 [51]	VH+VV
	I7 [52]	(VH-VV)/(VH+VV)
	I8 [53]	VH/VV
	I9 [54]	8*HV/(HH+VV+2*HV)
Texture (HV, VH)	Mean (ME)	$\sum_h \sum_k h * m_{hk}$

Variance (VA)	$\sum_h \sum_k h * m_{hk} * (h - Mean)^2$
Homogeneity (HO)	$\sum_h \sum_k \frac{m_{hk}}{1 + (h - k)^2}$
Contrast (CON)	$\sum_h \sum_k (h - k)^2 m_{hk}$
Dissimilarity (DIS)	$\sum_h \sum_k h - k m_{hk}$
Entropy (ENT)	$-\sum_h \sum_k m_{hk} \lg(m_{hk})$
Second Moment (SM)	$\sum_h \sum_k (m_{hk})^2$
Correlation (COR)	$\frac{\sum_h \sum_k h * k * m_{hk} - \mu_x \mu_y}{\sigma_x \sigma_y}$

さらに欧州宇宙機関 (ESA) の衛星である Sentinel2 による光学センサーのデータも補助として利用した。我々は 2019 年 5 月 8 日の Level2-A の Band2-12 までのデータをダウンロードし、3 x 3 ピクセルの平均を行い、表 3 に示す植生指標 (Vegetation Indices) を計算した。NDVI, EVI 等の植生指標は、地上植生の葉の総量や光合成活性を表すものとして、これまでたくさんの研究によって利用されてきており、森林構造を反映する PALSAR データを補完するものとして利用した。

表 3. 入力に利用した Sentine2-MSI データおよび計算した植生指標

Variables		Definition (Central Wavelength)
Bands, indices (Abbreviation)		
Multispectral Bands	Band2(B2)	Blue, 490nm
	Band3(B3)	Green, 560nm
	Band4(B4)	Red, 665nm
	Band5(B5)	Red edge, 705nm
	Band6(B6)	Red edge, 749nm
	Band7(B7)	Red edge, 783nm
	Band8(B8)	Near Infrared (NIR), 842nm
	Band8A(B8a)	Near Infrared (NIR), 865nm
	Band11(B11)	SWIR-1, 1610nm
	Band12(B12)	SWIR-2, 2190nm
Vegetation Indices	NDVI [55]	$\frac{(NIR - Red)}{(NIR + Red)}$
	EVI [56]	$\frac{2.5 * (NIR - Red)}{(NIR + 6 * Red - 7.5 * Blue + 1)}$
	DVI [57]	$NIR - Red$
	GARI [58]	$\frac{NIR - [Green - 1.7 * (Blue - Red)]}{NIR + [Green - 1.7 * (Blue - Red)]}$
	SAVI [59]	$\frac{1.5 * (NIR - Red)}{(NIR + Red + 0.5)}$

GNDVI [60]

GDVI [61]

SR [62]

GRVI [63]

$$\begin{array}{c} \frac{(NIR - Green)}{(NIR + Green)} \\ NIR - Green \\ NIR \\ Red \\ NIR \\ Green \end{array}$$

(3) 機械学習モデル

本研究ではランダムフォレスト(Random Forest)を利用した。コンピュータ言語Pythonでsklearnパッケージを使用し、48種類のデータ(10: Sentinel2-MSI スペクトルバンド, 9: Sentinel-2 MSI 植生指標, 4: PALSAR2 後方散乱係数, 9 PALSAR2 レーダー後方散乱係数指標, 16: PALSAR2 テクスチャ)を入力とし、航空機ライダー地上部バイオマスデータを80%, 10%, 10%に分割し、それぞれ学習、検証、テストサンプリングに使用した。ランダムフォレストは、たくさんの決定木を構築し、変数ごとの重要性を確認しながら、学習データとしてのバイオマスデータを再現する最良のモデルを求める。モデルの評価には、平均平方誤差RMSE、R2、平均絶対誤差MAE、相対RMSEを利用した。学習データの中で、250 Mg/ha以上の非常に大きなバイオマス量を測定エラーとして除去した。

(4) 結果と考察

ランダムフォレストモデルでは、入力変数のカテゴリ数と、説明変数の量に影響を受けて、入力変数の重要度(importance)が変化するが、入力変数の量を増やすと計算時間がかかりすぎてしまうという問題がある。そこで我々はPALSAR-2とSentinel-2の2つのカテゴリごとにGini係数が0.05以上の変数のみを利用して、計算を行うことにした。その結果は、スギのPALSAR-2については、VH mean、VH variance、HV variance、VH correlation、HV correlationが、スギのSentinel-2についてはBand 5、Band 9、Band 8a、Band 11、SR、NDVI、Band 12が、ヒノキのPALSAR-2については、VH mean、VH variance、HV mean、HV varianceが、ヒノキのSentinel-2についてはBand 12、Band 4、Band 9、Band 11、Band 5、Band 8a、Band 6が重要な変数として選択された。

ランダムフォレストアルゴリズムにおいて、ハイパーパラメータの調整は、モデルの性能を決める重要な課題である。繰り返し計算の結果、Maximum depth of the decision tree (MD)、Minimum number of samples (MS)、Minimum number of samples required to be at every leaf node (ML)、Number of trees in the forest (EST)を、それぞれ10, 2, 1, 200に設定した。

構築したモデルは、スギ林に対して作成したモデルは、 $R^2 = 0.31$, root-mean-square error (RMSE) = 54.38 Mg ha⁻¹, mean absolute error (MAE) = 40.98 Mg ha⁻¹, relative RMSE (rRMSE) = 0.35, ヒノキ林に対して作成したモデルは $R^2 = 0.37$, RMSE = 98.63 Mg ha⁻¹, MAE = 76.97 Mg ha⁻¹, rRMSE = 0.33 のある程度高い精度を示した(図4)。それらモデルを最終的に茨城県全体に対して適用した結果、17の県内市町村におけるスギの地上部バイオマス量は5.27 Pgで、単位面積あたりの平均値は146.50 Mg ha⁻¹、標準偏差は44.37 Mg ha⁻¹、ヒノキ林は、3.56 Pg、293.12 Mg ha⁻¹、78.48 Mg ha⁻¹であった(図5)。

我々が計算した地上部バイオマス量を、従来からの方法である森林簿(対象林小班における樹種・林齢を変数とする経験式)から求めたものについて、17の市町村(水戸、日立、石岡、常陸丘、高萩、北茨城、笠間、ひたちなか、常陸宮、那珂、鉾田、小美玉、茨城、大洗、下里、東海、大子)の平均値を求めて比較した(図6)。その結果は、両者の相関係数は0.99と非常に高く、森林簿はスギで25%の過大評価、ヒノキで53%の過小評価をしていることがわかった。

このように、航空機ライダーによる森林バイオマス量を学習データとし、衛星データを入力データとした機械学習モデルは、広域で高い精度で我が国の主要な商業用針葉樹であるスギとヒノキの地上部バイオマス量を再現することができた。また森林簿による従来の推定方法は、正確ではない可能性が高いことも示唆した。

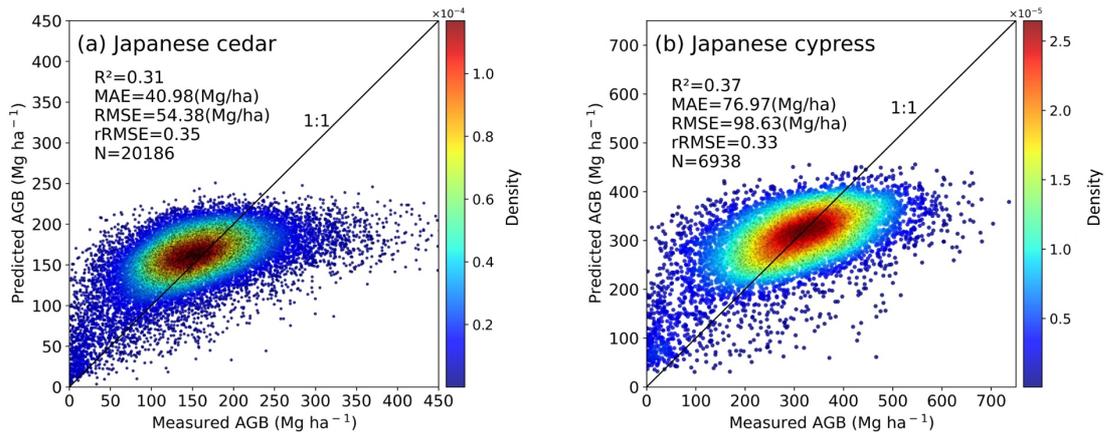


図 4. 茨城県における観測と衛星データによる推定された森林バイオマス ((a)スギ、(b)ヒノキ)

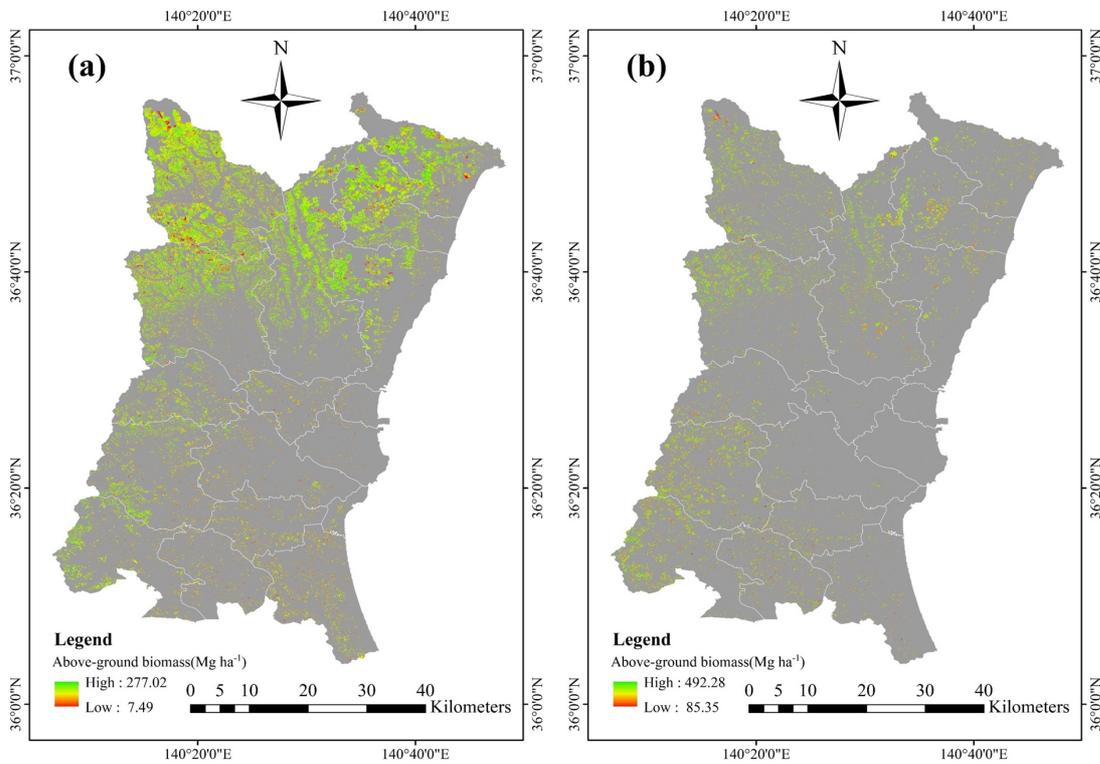


図 5. 茨城県における衛星データから推定された森林バイオマス ((a)スギ、(b)ヒノキ)

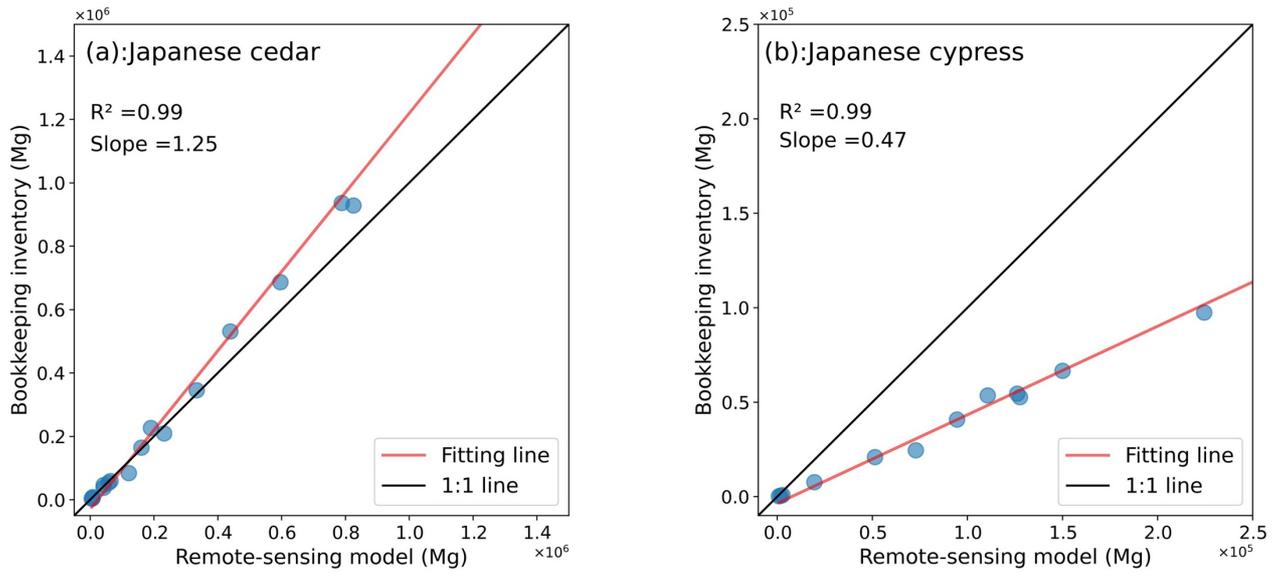


図 6. 茨城県北部 17 市町村における衛星データと森林簿から推定された森林地上部バイオマス ((a) スギ、(b) ヒノキ)

2-2 大分県の落葉広葉樹林における森林炭素蓄積量の推定

(1) はじめに

我が国の森林タイプの代表的なものとして落葉広葉樹林があるが衛星データを利用した森林地上部バイオマス (=炭素蓄積量) の推定はほとんど行われていない。そこで、大分県のクヌギに代表される落葉広葉樹林を対象 (図 7) とし、航空機ライダー観測に基づいた地上部バイオマス量を再現する機械学習モデルの構築を目的とした。

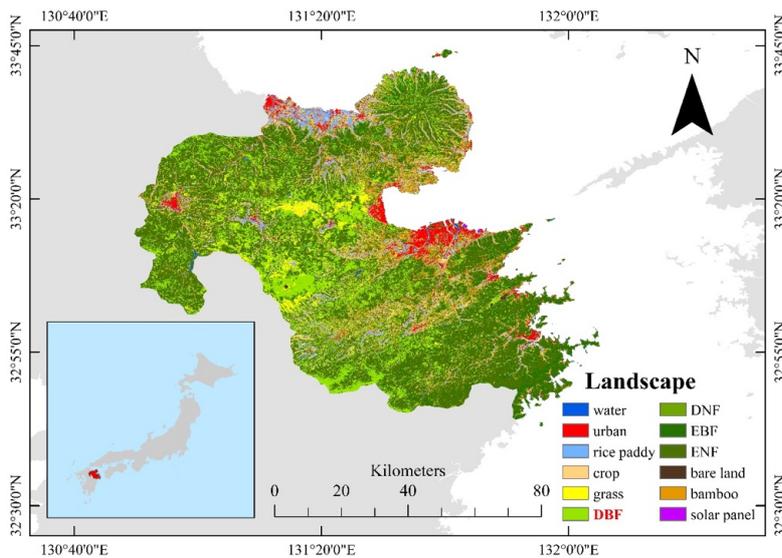


図 7. 大分県における対象地 (落葉広葉樹 DBF: 赤部分)

(2) データの整備

入力に利用する衛星データとして用意した Sentinel-2(ESA)による各種植生指標(NDVI, EVI, DVI, SR)及び PALSAR-2(JAXA) L-band SAR 画像の他に、PALSAR-2-FNF(森林分布図: JAXA; Shimada et al., 2014), SRTM (スペースシャトルによる地形情報, NASA)、Sentinel-1(ESA)なども利用することにしたが、それらのデータ容量が非常に大きいためランダムフォレスト回帰モデルを操作する上での大きな妨げとなる。そのため本研究では主成分分析(PCA)を利用して複数種類のデータをいくつかの軸にまとめる作業を行い、データ処理時間を減少させることにした。3x3 平均フィルターののちに、5x5 の Lee フィルターによってスペckルノイズを低減させた。

航空機ライダー観測に基づいた落葉広葉樹林の森林地上部バイオマスを学習データに利用した。それらはライダーデータより求められた群落高(DCHM)に対して地上 1 1ヶ所で計測された値より構築したアロメトリー式を利用して求めた材積量に、材積密度 0.646 Mg/m³ 及び拡大係数 1.33 をかけたものであり、10mx10m のメッシュで 89,324 ピクセルのデータ容量であった。地上部バイオマス密度は、0.1 から 701.5 Mg/ha の範囲をとり、平均 121.267Mg/ha で標準偏差は 60.47Mg/ha であった。

(3) 機械学習モデル

ここでは入力データの種類の異なる 4 つのランダムフォレスト(RF)回帰モデルを試した。Model 1 - 土地基礎情報 (PALSAR-2-FNF, SRTM, Sentinel-1) と PCA グループ (Sentinel-2 の PCA と植生指標のテクスチャ), Model 2 - 土地基礎情報と VI グループ(Sentinel-2 の植生指標とそのテクスチャ), Model 3 - 土地基礎情報とオリジナルバンドグループ(Sentinel-2 のマルチバンドデータとそのテクスチャ), All data - 上記全て (表 3)。

表 3. ランダムフォレスト回帰モデルの概要

RF-based model	入力データ	入力変数の総数
Model1	土地基礎情報, 3 PCA bands, and 30 relative textures.	70
Model2	土地基礎情報, 10 VIs, and 30 relative textures.	77
Model3	土地基礎情報, 10 original bands, and 30 relative textures.	77
All data	土地基礎情報, 3 PCA bands, 10 VIs, and, 10 original bands, 90 relative textures.	150

(4) 結果と考察

4 つのランダムフォレストモデルを試したところ、Model1 が最も良いパフォーマンスを示した (図 8; $R^2 = 0.69$, root-mean-square error ($RMSE$) = 33.59 Mg/ha, mean absolute error (MAE) = 25.01 Mg/ha, relative RMSE ($rRMSE$) = 0.27)。Recursive Feature Elimination (REF) プロセスによって選択された変数は、34 個のうち 19 個が PCA とそのテクスチャに関するものであった。

Model1 により 34 個に絞り込まれた変数でモデル構築し、計算したところ、落葉広葉樹林の森林地上部バイオマスは 32.69 から 437.72 Mg/ha の範囲をとり、平均 128.84Mg/ha で標準偏差は 32.24Mg/ha であった。大分県での総計は 14,507,596.8 Mg に登った (図 9)。

本研究では、新たに PCA を利用して変数容量を減らすことなどにより効率的なモデルを構築するとともに、これまであまり挑戦されてこなかった、我が国での落葉広葉樹林の森林地上部バイオマス推定について成功した。

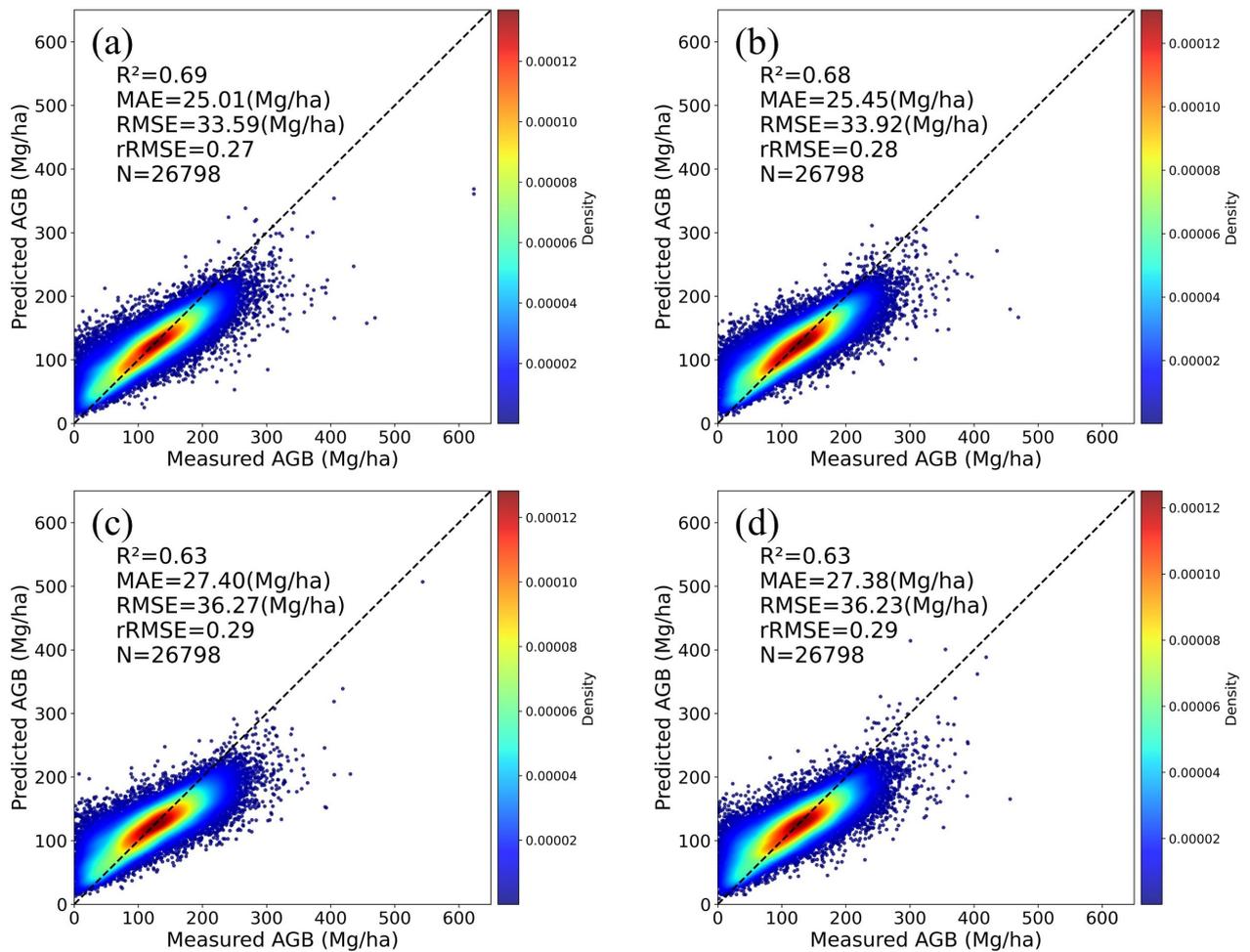


図 8: 大分県の落葉広葉樹林における地上部バイオマスの観測とモデルの比較: (a) model1, (b) model2, (c) model3, (d) All model.

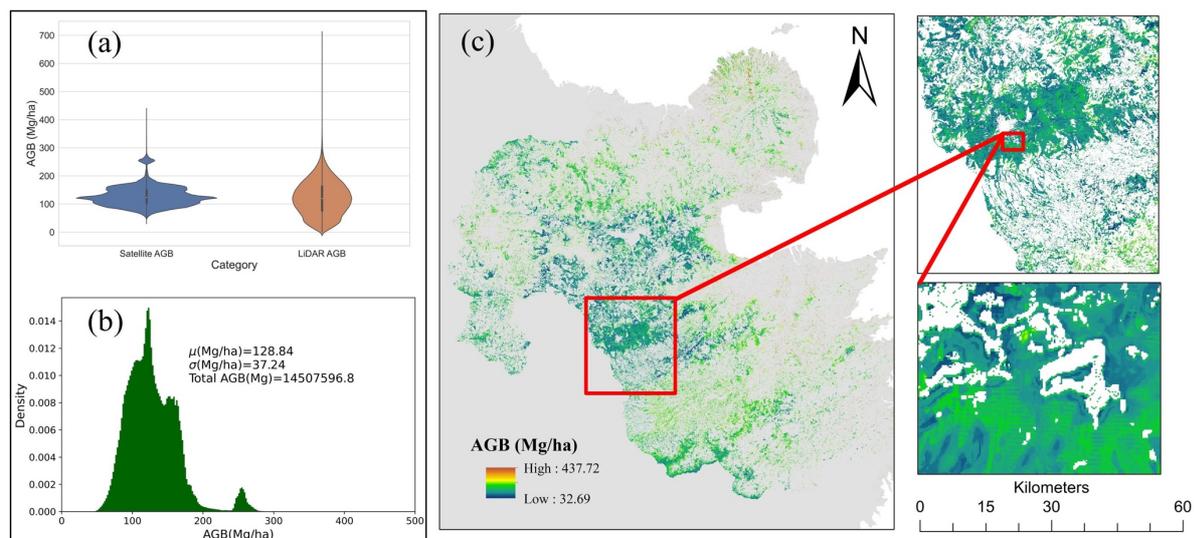


図 9: 大分県の落葉広葉樹林における地上部バイオマス. (a) AGB distributions derived from satellite+RF-RFE (left) and airborne LiDAR (right), (b) histogram of mapping AGB in 10m grid cell, (c) AGB map of DBF in study area with close-up images (d) and (e).

3 おわりに

本研究では、茨城県のスギ・ヒノキ林及び大分県の落葉広葉樹林を対象とし、人工衛星を利用した森林地上部炭素蓄積量（＝森林地上部バイオマス）を推定するための機械学習モデルの構築に成功した。これまで我が国ではあまり行われてこなかった地域スケールでの推定であり、PCA によるデータの圧縮を利用した計算の効率化など手法的な新規性も見出すことができたことは非常に大きな成果であった。今後はより広域なスケールへ拡張していくべくすでに研究を進めているが、一方による広域化で樹木タイプが増えることによる情報要求量の増加や、その大容量データの効率的な処理について、さらなる技術的な開発が必要であることが予想される。我が国は南北に長いための樹種タイプの豊富さや、地形の複雑さを克服して森林地上部炭素蓄積量の推定を目指すことは、森林に国土の7割が覆われる日本だけでなく同様な各国にとって非常に意義深いものであると考えられる。

【参考文献】

- Saatchi S, Buermann W, ter Steege H et al (2008) Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens Environ* 112(5):2000–2017. <https://doi.org/10.1016/j.rse.2008.01.008>
- IPCC, 2013: Climate Change 2013: The Physical Science Basis.
- Fang, J.Y., Kato, T., Guo, Z., Yang, YH., Hu, HF., Shen, HH, Zhao, X., Kishimoto-Mo, AW., Tang, YH. Houghton, RA., 2014: Evidence for environmentally enhanced forest growth. *Proc. Nat. Aca. Sci.*, 111(26), 9527-9532, doi: 10.1073/pnas.1402333111
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraiishi, T., Thapa, R., Lucas, R., 2014. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sensing of Environment* 155, 13-31.

〈発表資料〉

題名	掲載誌・学会名等	発表年月
Using multisource satellite products to estimate forest aboveground biomass in Oita prefecture: An approach with improved accuracy and computational efficiency	Geocarto International	2022年10月
Estimation of Forest Aboveground Biomass of Two Major Conifers in Ibaraki Prefecture, Japan, from PALSAR-2 and Sentinel-2 Data	Remote Sensing	2022年1月