

全球での建築物ストックデータの整備に向けた  
リモートセンシングデータと建築物データの関連分析  
村上大輔・山形与志樹

**Relation Analysis of Remotely Sensed Imagery and Building Statistics Toward  
Developing Global Building Stock Data Set  
Daisuke Murakami and Yoshiki YAMAGATA**

**Abstract:** Building stock data (e.g., building density, building height,...) are required to estimate CO<sub>2</sub> emissions, energy demands, and other variables characterizing sustainability. The objective of this study is examining how accurately building stock information can be estimated using global open data, including remote sensing data and OpenStreetMap data. The estimation is performed using the standard linear regression and the gradient boosting, which is an ensemble learning algorithm. The analysis result suggests that the gradient boosting accurately estimates building densities and building heights that are aggregated into 500 m (or coarser) grids.

**Keywords:** リモートセンシング (remote sensing)、PALSAR-2、建築物ストック (building stock)、WUDAPT

## 1. はじめに

CO<sub>2</sub> 排出量やエネルギー需要などを推定するための基礎データとして建物ストックデータは重要である。幸い、わが国では、例えば東京大学空間情報科学研究センターや株式会社ゼンリン (<http://www.zenrin.co.jp/>) などが個別建物毎のデータを推計・公開している。しかしながら、建物データは、必ずしも海外においても十分に整備されてきてはいない。

グローバルな炭素管理の第一歩として、国内だけでなく国外の都市についても建物データを整備する必要がある。そのような背景の下、全球を

対象とした都市データセットの整備が World Urban Database and Access Portal Tools (WUDAPT : <http://www.wudapt.org/>) プロジェクトの一環で近年進められている。

全球の建物データセットを構築するためには、できるかぎりコストを抑えながら精度の良い推計を行う必要がある。幸い近年では、PALSAR-2 や LiDAR のような建物の構造推定に役立つリモートセンシングデータや、道路網や土地被覆といった関連データが、複数のポータルサイトで一般公開されている。具体的には OpenStreetMap (OSM : <https://www.openstreetmap.org/>) や SEDAC (<http://sedac.ciesin.columbia.edu/>)、Natural Earth (<http://www.naturalearthdata.com/>) などがある。

本研究ではグローバルな建築物ストックデータ整備の第一歩として、全球で入手可能なデー

タを用いることで、建築データがどの程度の精度で推計できるかを検証する。

以降の章立ては次のとおりである。第2章では我々が分析に用いるデータと推計手法を説明する。第3章では建物データの推定結果について紹介し、第4章では今後の方針を示す。

## 2. 分析の概要

本章では、建物データを2.1節で、建物データを推計するための説明変数を2.2節で、それぞれ説明する。次に説明変数から建物高さと建物密度を推計する手法を2.3節で導入する。

### 2.1 建物データ

LiDAR データと建物ポリゴンデータを重ね合わせることで、各建物の地上高と面積を評価した。次に、それらを集計することで、100m メッシュ毎の建物高さ（中央値）データと建物密度データを作成した（図1左）。また500m メッシュ毎の両データについても同様に作成した。

本研究ではメッシュ別の建物データを推計するためのモデルを構築する。もし同モデルの精度が良好であれば、LiDAR（処理コストが大きい）を用いたモデルの推定は一部の都市・地域のみで行い、推定されたモデルを広域に適用することで効率良く建物データを推計する、ということが可能となる。

### 2.2 説明変数データ

都市内部の建築物ストックを説明するだけの空間解像度を有する PALSAR2 と OSM の両オープンデータを用いる。

PALSAR2 とは人工衛星 ALOS-2 に搭載された合成開口レーダ（地表面から反射される電波を受信することで情報を得るレーダ）であり、H を水平偏波、V を垂直偏波とすると、次の4種類の偏波データが観測可能である：HH；HV；VH；VV。東京駅周辺において2015年9月に観測した各偏

表1：説明変数

データ	変数	記述
Palsar-2	HH	HH 偏波
	HV	HV 偏波
	VH	VH 偏波
	VV	VV 偏波
OSM	Road	道路密度
	PRoad	主要道路密度

波の強度を図1に示す。この図からもわかるように、各偏波は地物に応じてそれぞれ異なる分布傾向を示す。それらを利用することで建物データが精度良く推計できる可能性がある。PALSAR2 の空間解像度は最高で3mである。Palsar-2 について詳しくは ALOS-2 の HP ([http://www.eorc.jaxa.jp/ALOS/index\\_j.htm](http://www.eorc.jaxa.jp/ALOS/index_j.htm)) を参照されたい。

一方で OSM は自由に利用・編集ができる世界地図データベースであり、図1右に示したように東京都心においても詳細な道路情報が道路区分等の属性とともに整備されている。

### 2.3 手法

本研究では線形回帰モデル(linear regression: LM; 1 式)を推計に用いる。

$$\log(y_g) = \sum_k x_{g,k} \beta_k + \varepsilon_g \quad (1)$$

$y_g$  はグリッド  $g$  における建築物データの値、 $x_{g,k}$  は同グリッドにおける  $k$  番目の説明変数の値、 $\beta_k$  は回帰係数、 $\varepsilon_g$  は独立同分布に従う攪乱である。本分析に用いる説明変数を表1に示す通りである。

LM にはいくつかの短所がある。まず各説明変数間からの影響は必ずしも線形とは限らない。また各説明変数は類似した分布傾向を持つため、多重共線性にも対処する必要がある。そこで、本研究では、多重共線性に対処した上で、以下の非線

形性モデルも推定することとする。

$$\log(y_g) = \sum_m w_m f_m(x_{g,k}) + \varepsilon_g \quad (2)$$

$f_m(x_{g,k})$ はパラメータを無作為に与えた  $m$  番目の  $x_{g,k}$  のスプライン関数を表し、 $w_m$  は  $f_m(x_{g,k})$  に対する重みである。(2)式の推定には勾配ブースティング (gradient boosting: GB ; Freidman, 2001) を用いる。GB とは  $f_m(x_{g,k})$  を多数 (本研究では 200) 発生させて、各々の当てはまりの良さに応じてそれらを重みづけた上で (重み:  $w_m$ ) 加重平均をとることで、モデルを推定するというアンサンブル学習手法の一つである。勾配ブースティングについて、詳しくは Hastie and Tibashirani (2014)を参照されたい。

### 3. 分析結果

建物高さと建物密度の両方に対する LM と GB の推計精度を 10 fold cross-validation (10 fold CV) を用いて比較する。ここで 10-fold CV とは標本を無作為に 10 分割して、そのうちの 9 つでモデルを推定して、残りの 1 つで推計 (予測) 精度を検証するという、アプローチである。今回は、誤差率と相関係数を用いて推計精度を検証した。

建物高さに対する CV の結果を表 2 に、建物密度に対する結果を表 3 に、それぞれ整理した。100m グリッド別データに対しては LM も GB も精度が低い一方で、500m グリッド別に集計した場合では両モデルの精度は比較的良好であることが確認できる。また全 4 ケースで GB の精度が LM を上回った。

GB による建物高さと建物密度と両推計結果 (CV での out-of-sample prediction で得たもの) を図 2~5 にプロットした。500m グリッド毎の推計値は非常に実測値に近い。さらに 100m グリッド毎の結果についても大まかな分布傾向は実測値に非常に類似していることが確認できる。

最後に GB で推定された建物高さに対する各説明変数の寄与率を表 4 に整理した。この表より、

表 2 : Cross-validation の結果 (建物高さ)

集計単位	統計量	LM	GB
100m グリッド	誤差率	2.987	2.725
	相関係数	0.321	0.343
500m グリッド	誤差率	0.449	0.336
	相関係数	0.754	0.790

表 3 : Cross-validation の結果 (建物密度)

集計単位	統計量	LM	GB
100m グリッド	誤差率	3.697	2.669
	相関係数	0.383	0.540
500m グリッド	誤差率	0.753	0.688
	相関係数	0.824	0.883

表 4 : 建物高さに対する各説明変数の寄与率(GB)

単位	HH	HV	VH	VV	Road	PRoad
100m	0.25	0.00	0.17	0.07	0.27	0.24
500m	0.08	0.00	0.23	0.30	0.21	0.18

多くの説明変数が建物高さの推計精度向上に寄与していることを確認した。より多くの説明変数からの効果を安定的に推定することが可能な GB によるアンサンブル学習は、建物データセットを効率よく構築していくうえで有用といえよう。

### 4. まとめ

本研究では東京都心部を対象に、オープンデータを活用した建物データの推計精度の検証を行うことで GB の有用性を確認した。今後は東京以外の都市でも推計精度を検証した後に、より多様なオープンデータを活用しながら、全球の都市を対象とした建物データセットの構築に取り組みたい。

### 参考文献

- Friedman, J. H. (2001) Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Hastie, T. and Tibshirani, R. (2014) 「統計的学習の基礎: データマイニング・推論・予測」, 共立出版..

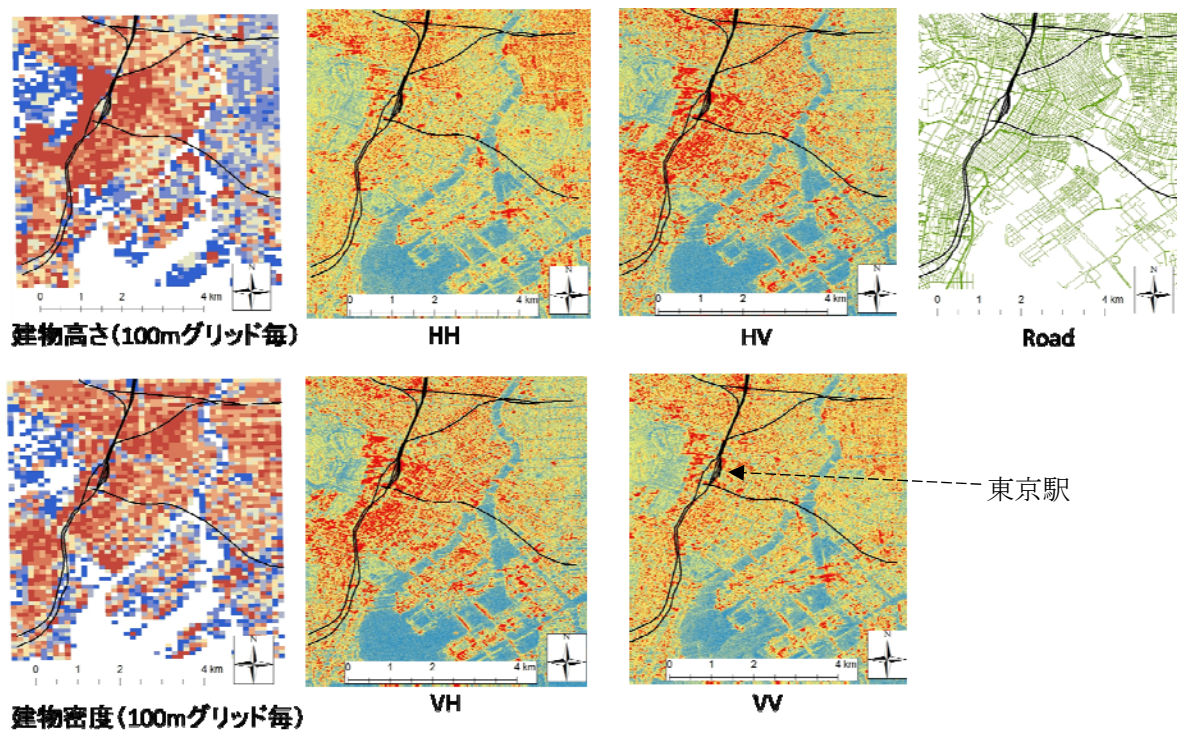


図1：本研究で使用する変数の空間分布

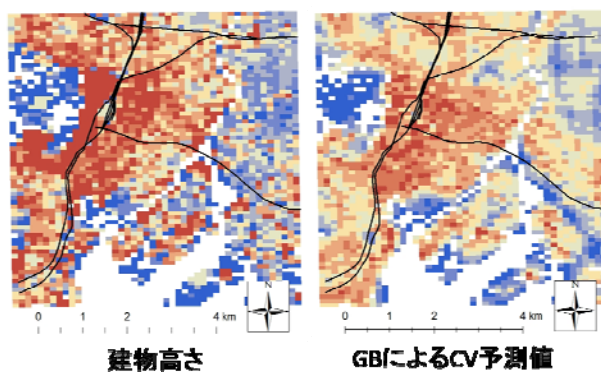


図2：GB による建物高さの CV 予測値  
(100m グリッド)

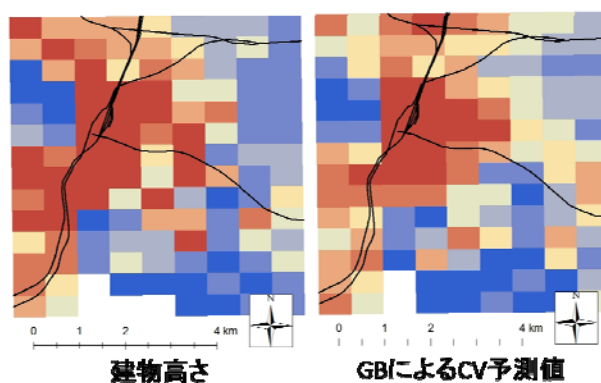


図4：GB による建物高さの CV 予測値  
(500m グリッド)

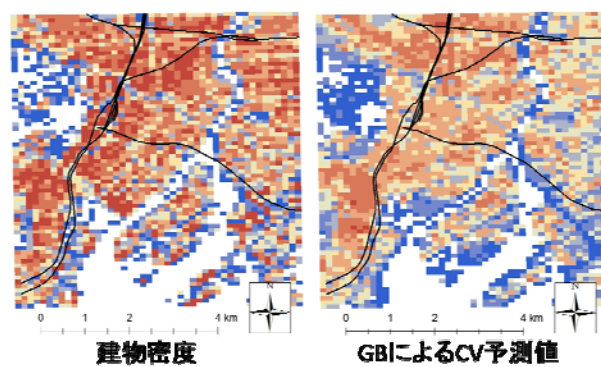


図3：GB による建物密度の CV 予測値  
(100m グリッド)

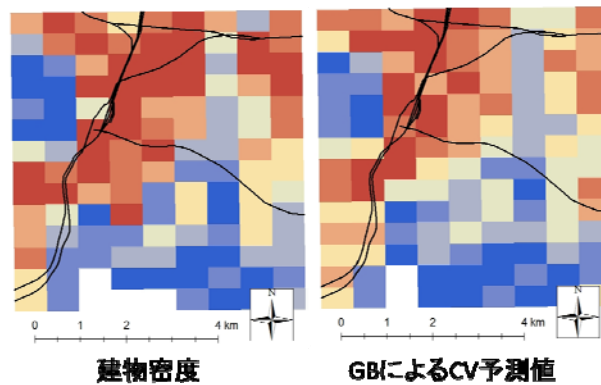


図5：GB による建物密度の CV 予測値  
(500m グリッド)