Estimating the Building Height from Sentinel Data at 10 m Resolution

Xin Yan

LA 558 Final Project

Contents

01. Background

02. Data and Method

03. Results

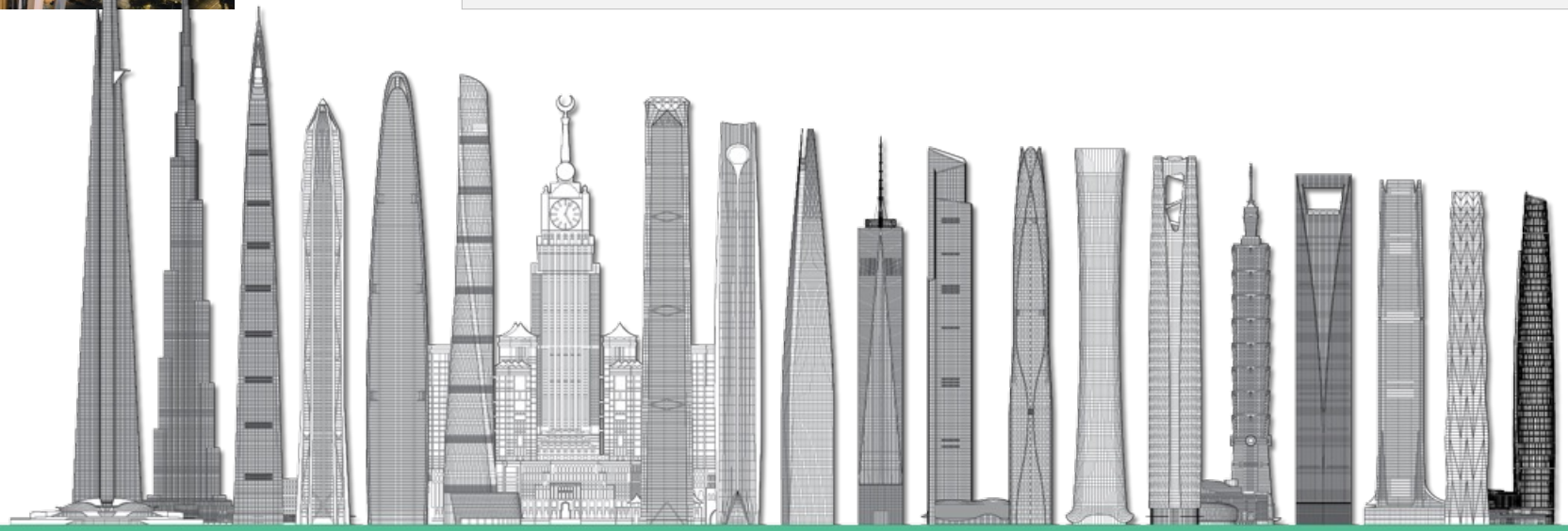
04. Discussion and Summary

Background

Why we need to know the building height?

Building height --- an important indicator for

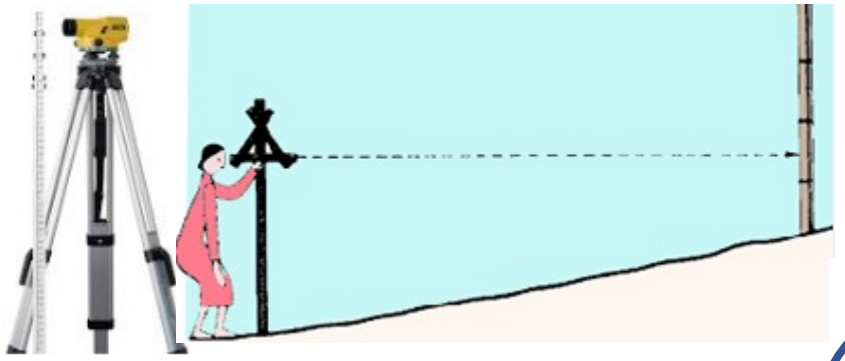
- Estimating energy consumption
- Material stock allocation
- Greenhouse gas emissions
- Urban heat island effects
- Distribution of population
- Urban development plan



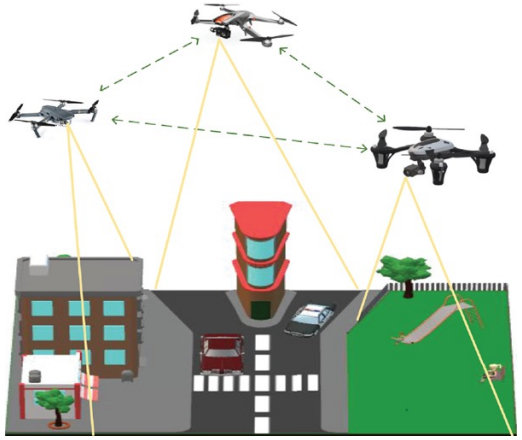
Background

How to measure the height of building? **New technology!**

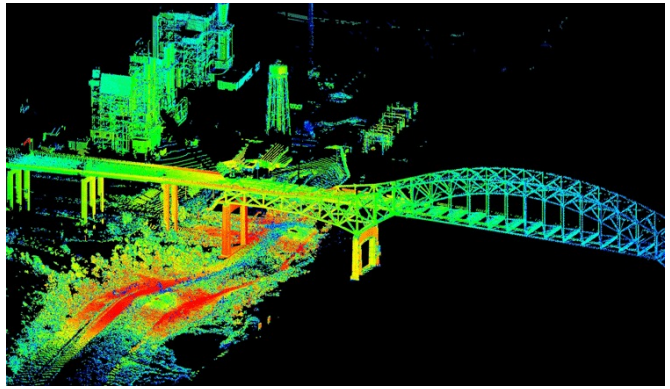
Handy



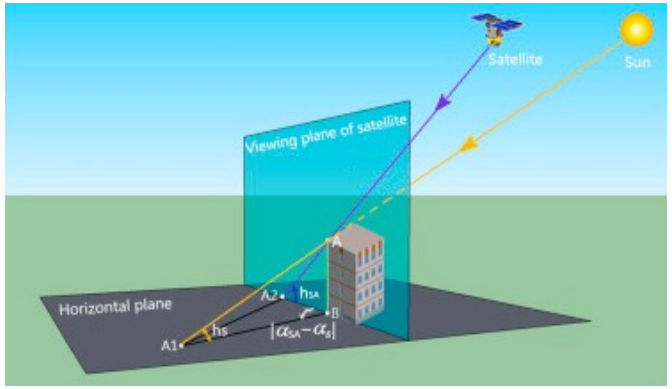
Photogrammetry



Lidar



Satellite



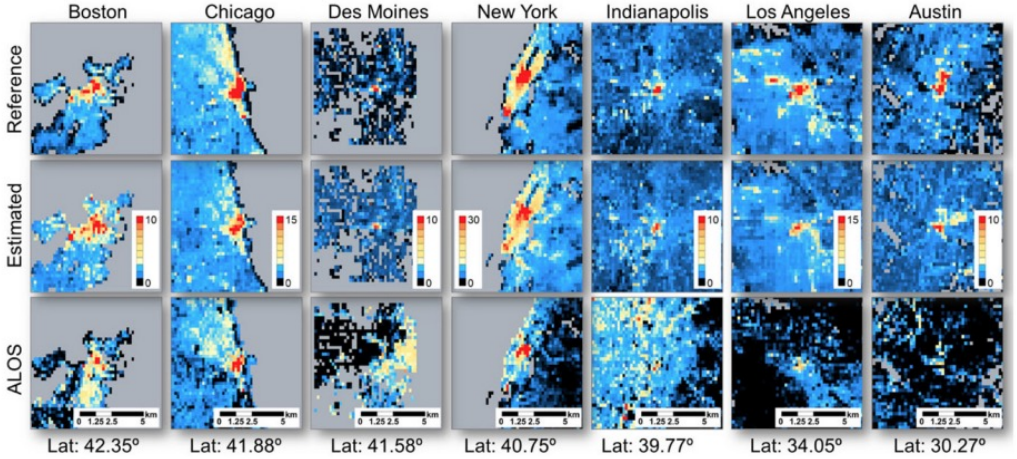
Research Gap

What's the research frontier?

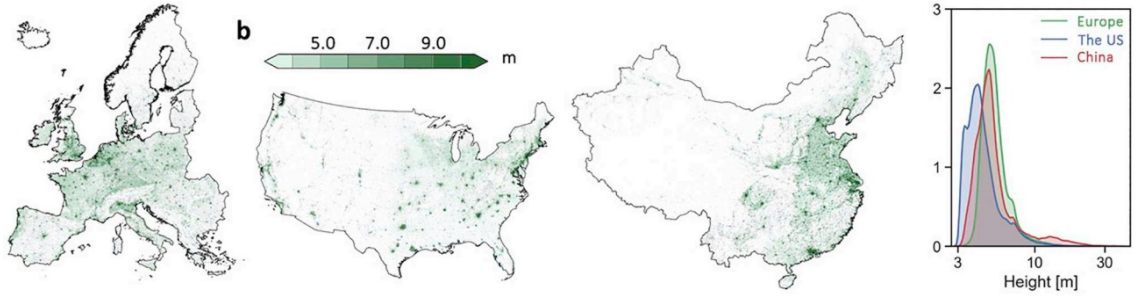
1. Li et al., 2020

$$VVH = VV * \gamma^{VH}$$

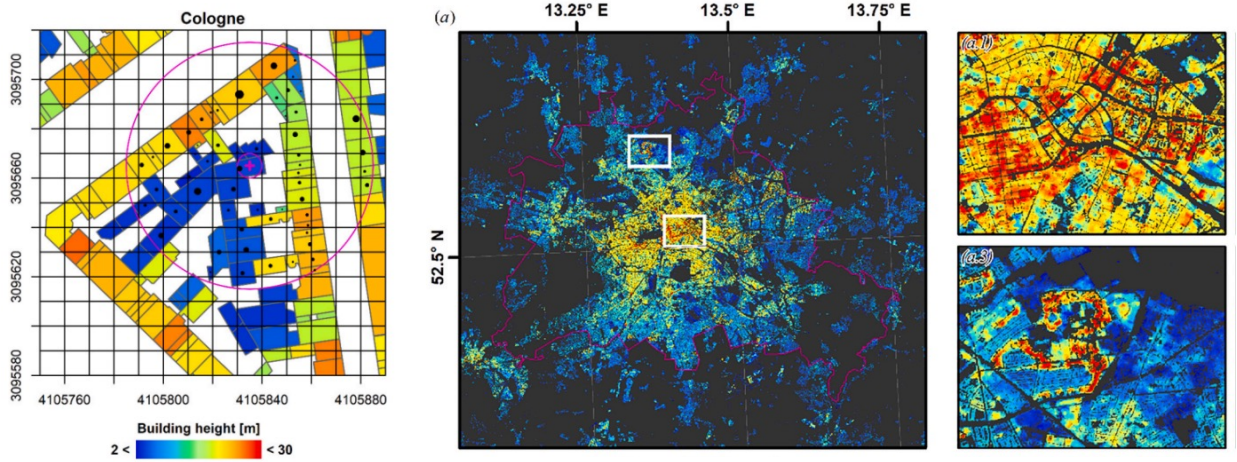
$$\ln H = a * VVH^{b} + c$$



2. Li et al., 2021



3. Frantz et al., 2021



Research goal:

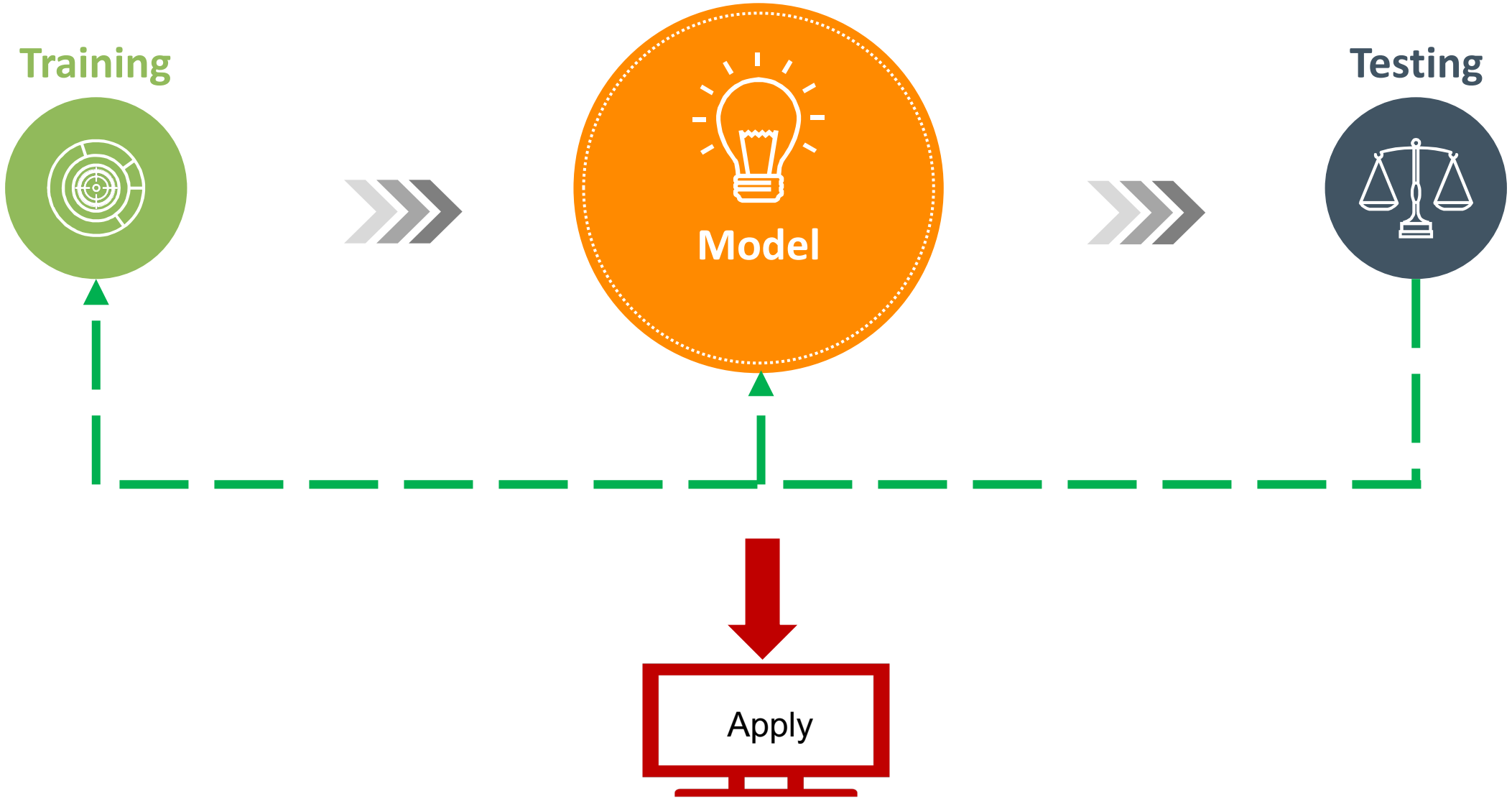
- Develop an effective and accurate model to estimate building height at 10 m resolution.
- Obtain national or global building height data.

Data and Method

Independent variables
Dependent variable
Feature selection
Model ensemble
Model evaluation

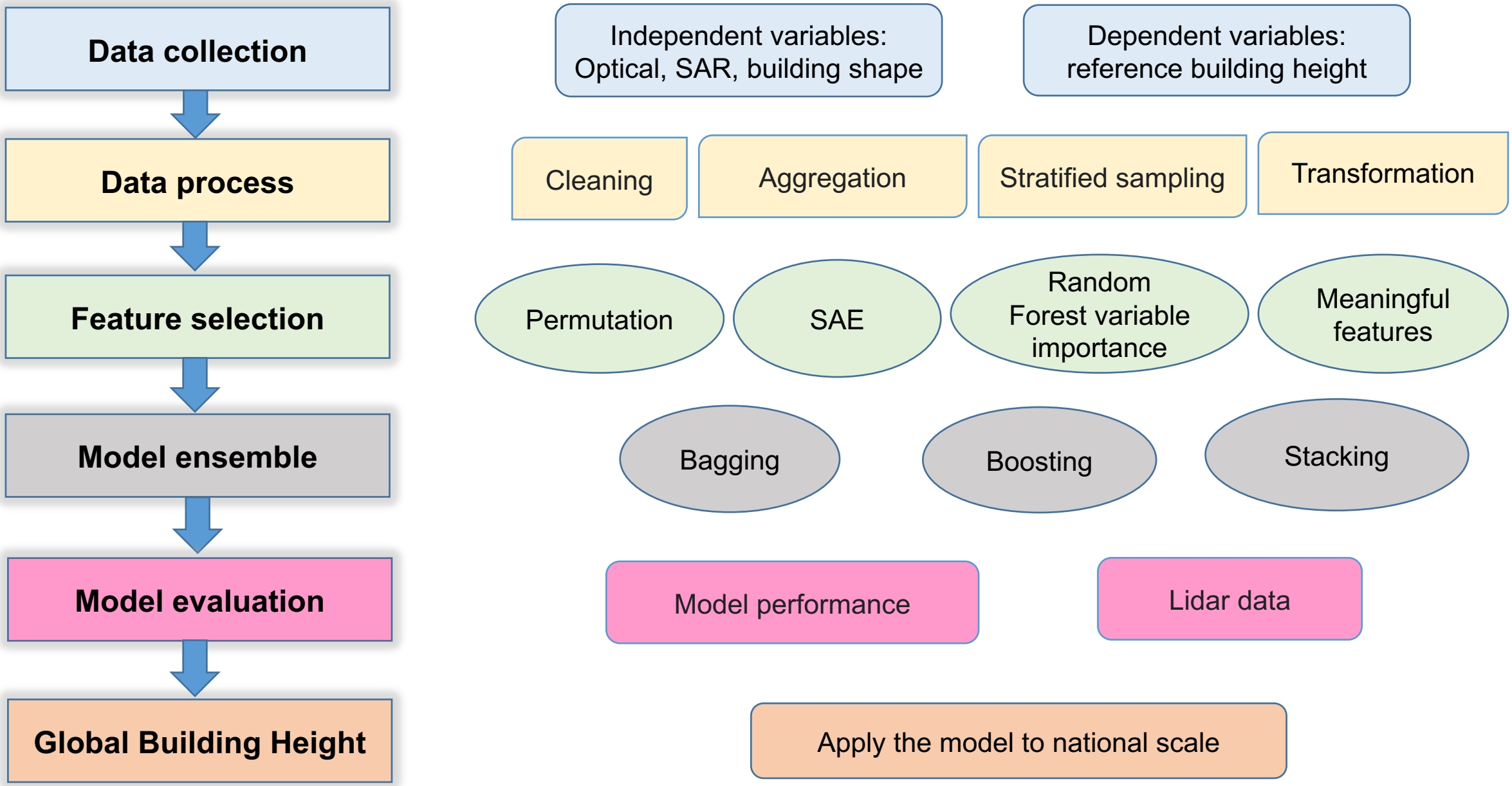
How to build a machine learning model

Basic logic



Flow chart

More detailed flow



Independent variables

Spatial – Spectral - Temporal feature database

Spectral Indexes

Sentinel-2

Band 2,3,4,8 (10m)

Sentinel-1

VV, VH, VVH

Indexes

NDWI, NDVI, LSWI, NDBI, mNDWI

B2	496.6nm (A) / 492.1nm (B)	Blue
B3	560nm (A) / 559nm (B)	Green
B4	664.5nm (A) / 665nm (B)	Red
B8	835.1nm (A) / 833nm (B)	NIR

$$VVH = VV * 5^{VH}$$

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}}$$

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$$

$$LSWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

$$NDBI = \frac{\rho_{SWIR} - \rho_{NIR}}{\rho_{NIR} + \rho_{SWIR}}$$

$$MNDWI = \frac{\rho_{Green} - \rho_{SWIR}}{\rho_{Green} + \rho_{SWIR}}$$

A total of 160 features

Temporal Indexes

Average, standard deviation, 0/10/25/50/75/90/100% quantiles, range, interquartile range, skewness, and kurtosis

156

+

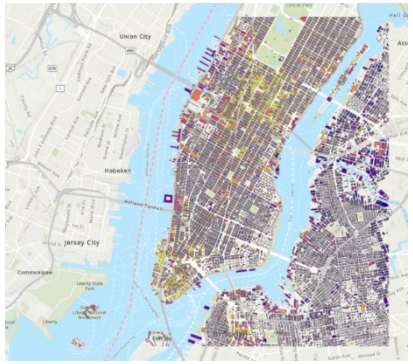
Spatial index

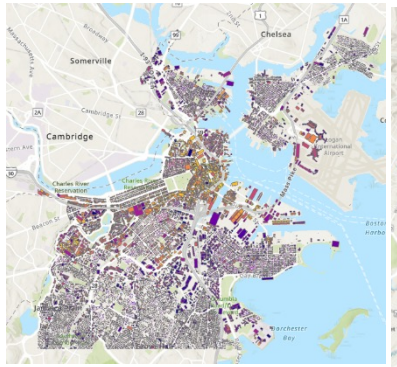
Length
Width
Near distance
Orientation

4

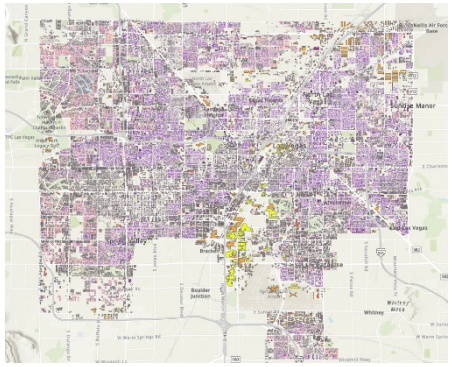
Dependent variable: building height

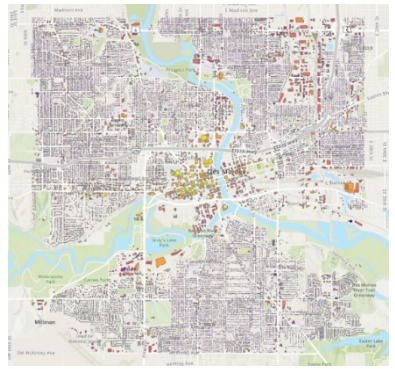
Reference building height data from twelve cities

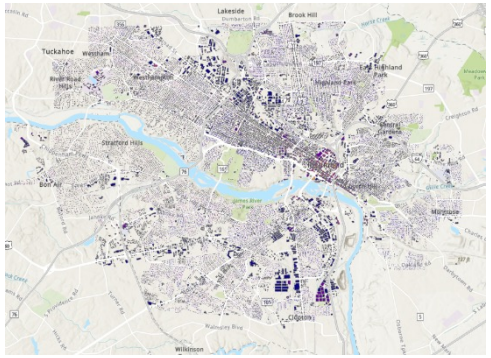

New York

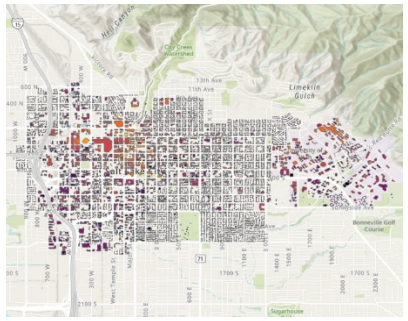

Boston

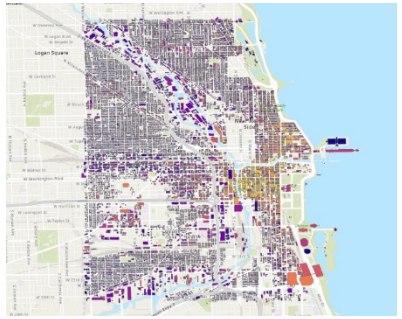

Austin

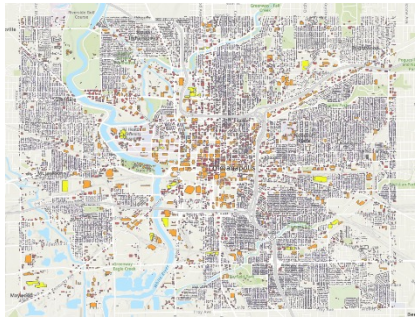

Las Vegas

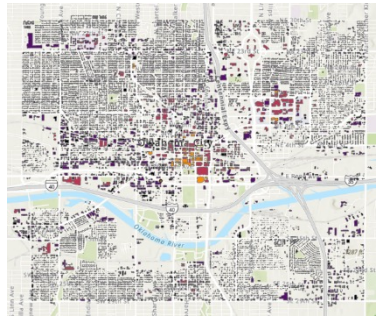

Des Moines

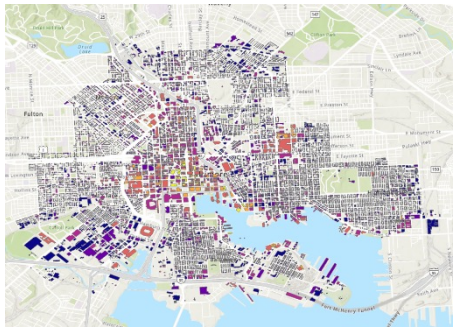

Richmond

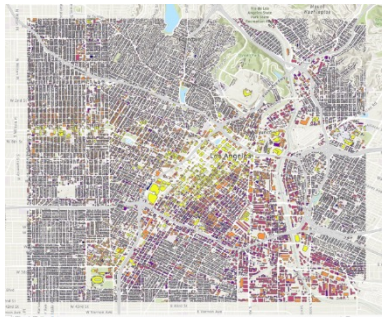

Salt Lake City


Chicago


Indianapolis


Oklahoma


Baltimore


Los Angeles

Building footprint source: https://github.com/microsoft/USBuildingFootprints

Feature selection

Get the stable features

❖ Shapley Additive Explanations

This value is the average marginal contribution of a feature value across all the possible combinations of features.

❖ Permutation feature importance

This approach directly measures feature importance by observing how random re-shuffling (thus preserving the distribution of the variable) of each predictor influences model performance.

❖ Random Forest variable importance

The default method to compute variable importance is the mean decrease in impurity mechanism: At each split in each tree, the improvement in the split-criterion is the importance measure attributed to the splitting variable, and is accumulated over all the trees in the forest separately for each variable.

❖ Meaningful features

Besides above statistic indicators, we can also select the meaningful features manually.

Independent variables: MBG_Length, MBG_Width, LSWI_kurtosis, LSWI_skew, NDVI_skew, NDWI_skew, B3_p10, B4_p10, NDVI_interquatile_range, VV_mean, VH_mean, VVH_mean, VVH_stdDev

Model ensemble

RF VS. SVM VS. Glmnet VS. gamboost

- **Random Forest regression**
- **Support Vector regression**
- **Glmnet**
- **Generalized Additive Models**

Preliminary accuracy comparison of different models

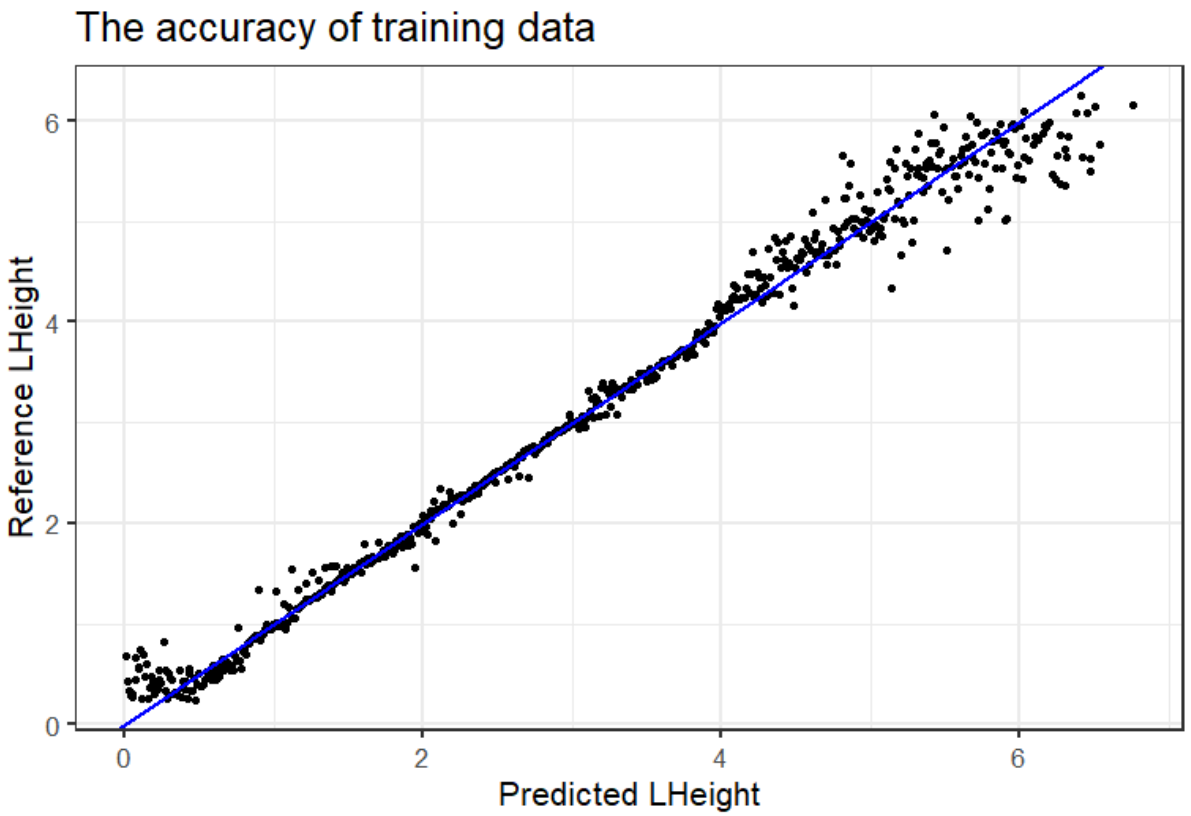
Rsquared

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
rpart	0.4178080306	0.4478954	0.5066016	0.4968194	0.5297325	0.5954782	0
glm	0.0001710916	0.6088913	0.6225619	0.5265694	0.6360930	0.6831387	0
glmnet	0.5973529469	0.6186869	0.6352957	0.6351207	0.6477667	0.6902985	0
lm	0.0001710916	0.6088913	0.6225619	0.5265694	0.6360930	0.6831387	0
rf	0.6804925350	0.7203345	0.7385717	0.7352511	0.7560331	0.7973643	0
knn	0.4489230663	0.4971347	0.5325895	0.5298389	0.5486733	0.6448886	0
svmRadial	0.6760886457	0.7073117	0.7270151	0.7264737	0.7484454	0.7805409	0
gausssprLinear	0.5980602391	0.6171075	0.6351901	0.6348789	0.6474015	0.6919441	0
gamboost	0.6492994480	0.6632784	0.6823294	0.6807147	0.6943563	0.7222680	0

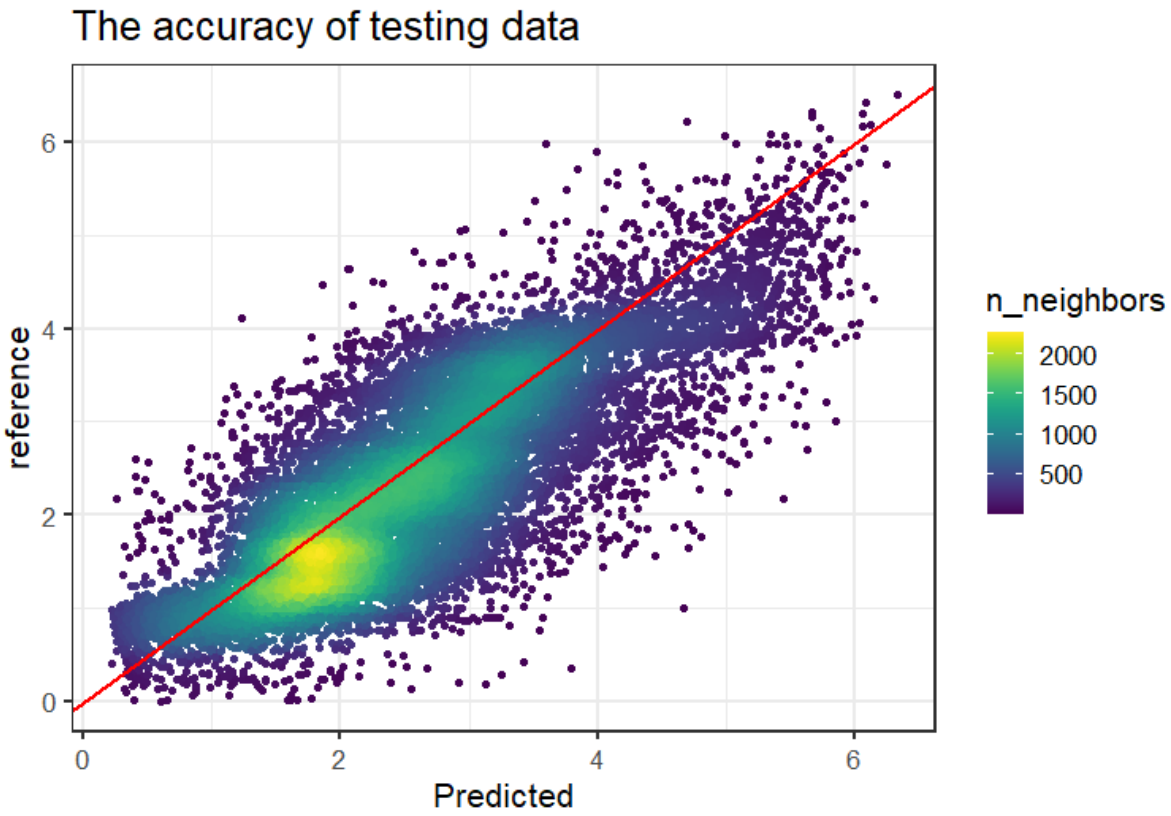
Results

Accuracy
Building height mapping

Test the accuracy



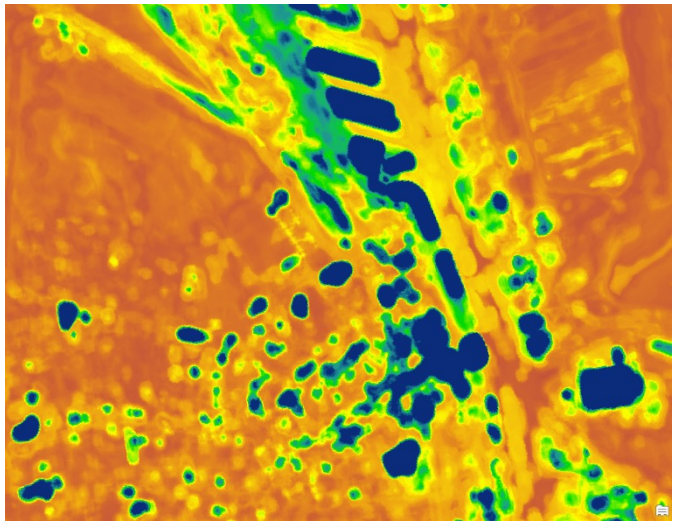
Mean of squared residuals: 0.04895864
% Var explained: 98.72



$R^2 = 0.782$

Building height mapping

Predicted building height in DC, Washington

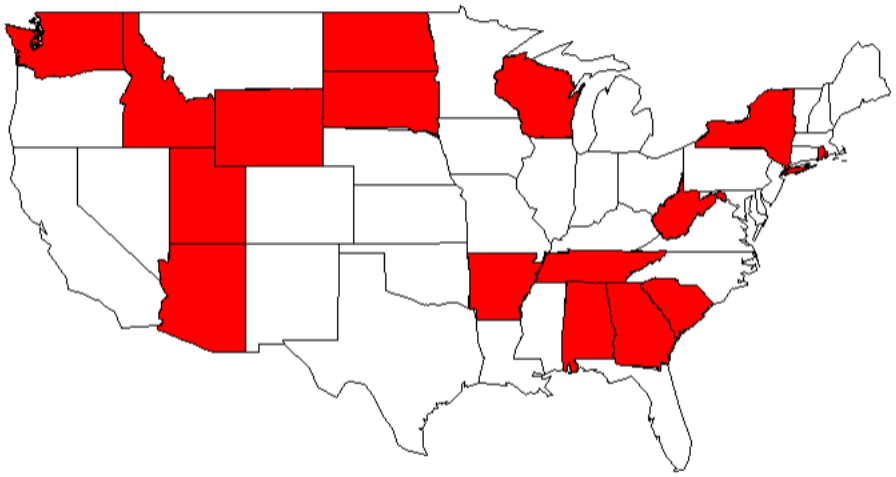


Predicted height in Alabama

Satellite image in Alabama

Building height mapping

Regions where data has been produced



Time consuming……

- Use more and better computer;
- Use server

Discussion and summary

Future works
Summary

Discussion

Deficiency and solutions

- ❑ More systematic accuracy analysis
- ❑ Building height distribution analysis

Summary

TAKEAWAY

Process:

The data are processed in R mostly and displayed in Tableau.

Contributions:

- ❑ Combined optical & SAR imagery, and building shape metrics to propose a set of Spatial – Spectral - Temporal feature database.
- ❑ Build a building height estimation model at 10m resolution.

Applications:

- ➢ Easily estimate and update large-scale (even global) building height.


Thank you
xinyan@iastate.edu