Height Inference for all USA Building Footprints in the Absence of Height Data

P5 Presentation

Imke Lánský

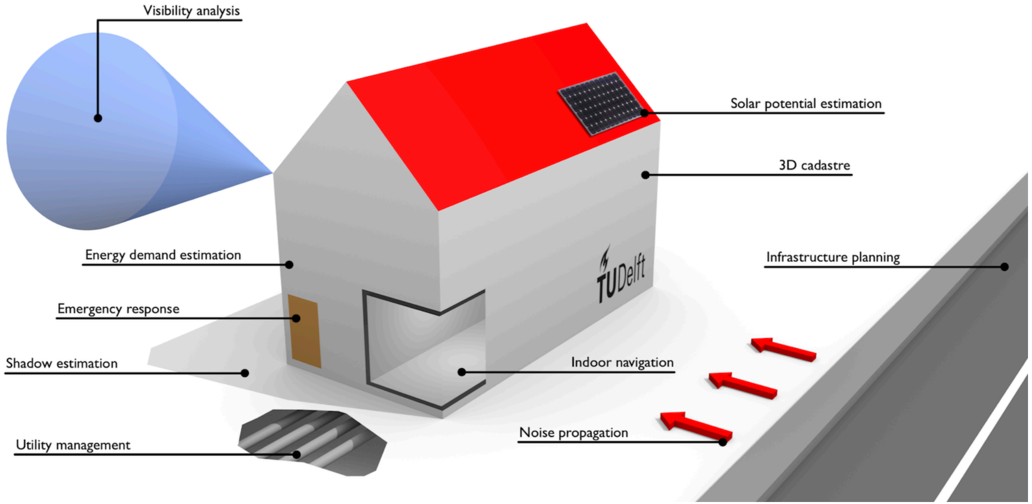
Supervisors: Dr. H. Ledoux & B. Dukai & Dr. F. Biljecki

Co-reader: Prof.dr. J.E. Stoter

Delegate: Prof.dr. W.K. Korthals Altes

Introduction

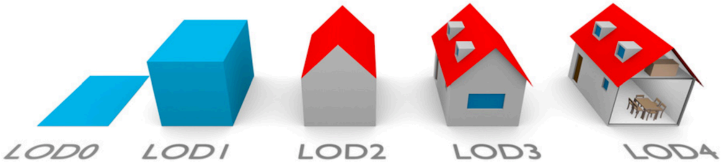
3D City Models



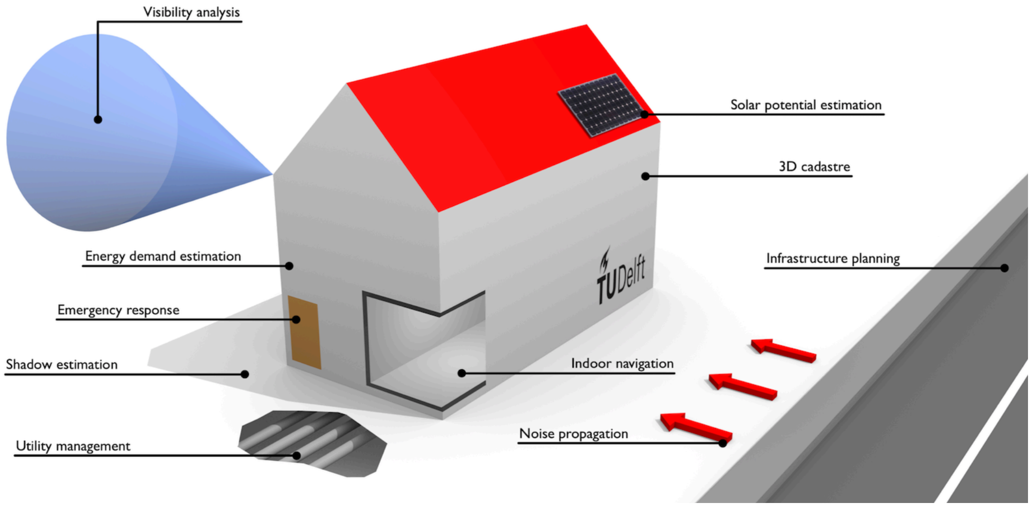
Applications 3D city models

Image: [Biljecki et al., 2015]

3D City Models

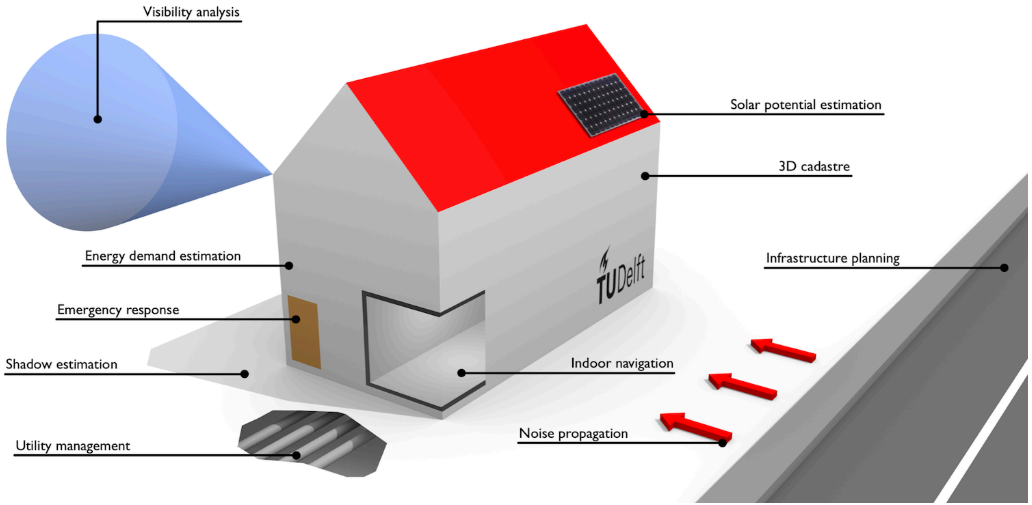


Levels of Detail (LODs)
Image: [Biljecki et al., 2016]



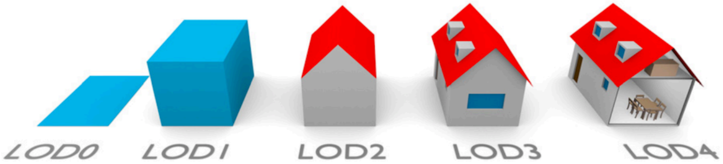
Applications 3D city models
Image: [Biljecki et al., 2015]

3D City Models



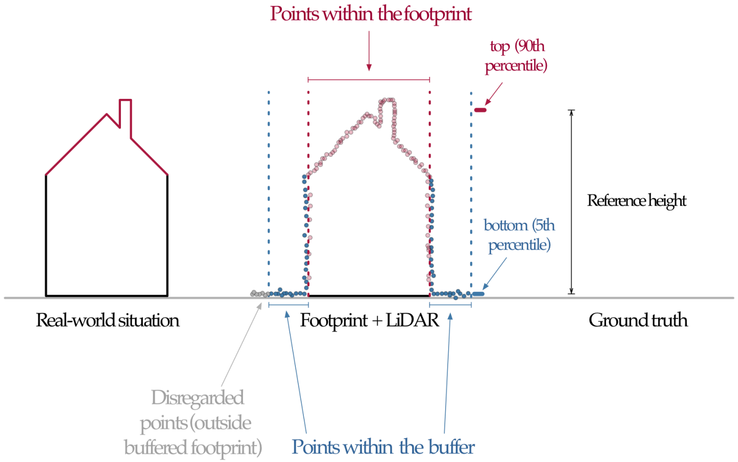
Applications 3D city models

Image: [Biljecki et al., 2015]



Levels of Detail (LODs)

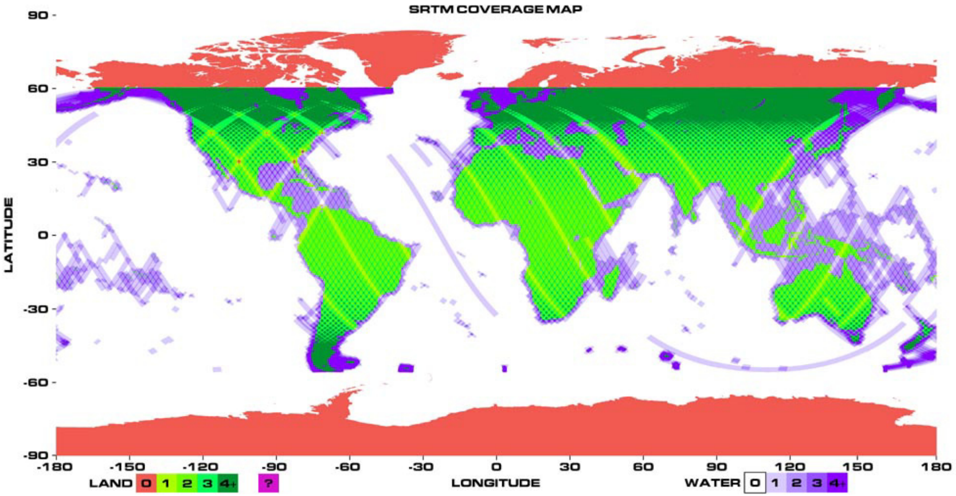
Image: [Biljecki et al., 2016]



Building heights from LiDAR

Image: [Biljecki et al., 2017]

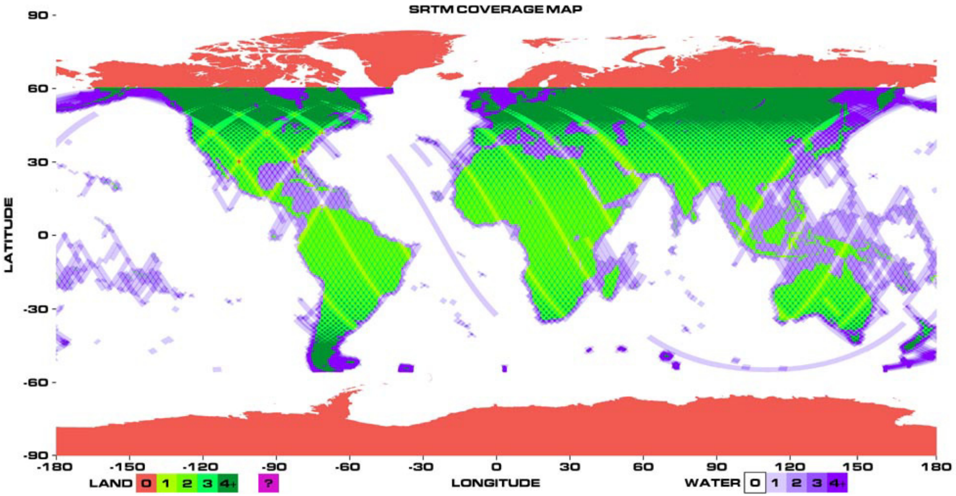
Alternative Options



Shuttle Radar Topography Mission (SRTM) coverage

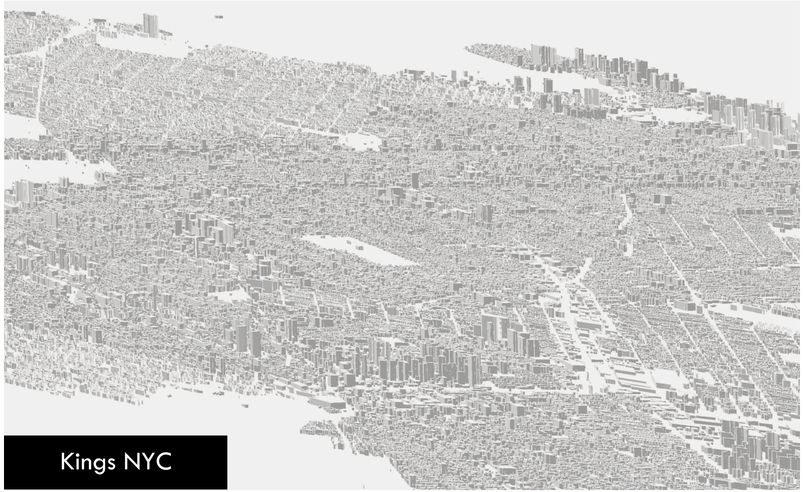
Image: [JPL NASA]

Alternative Options



Shuttle Radar Topography Mission (SRTM) coverage

Image: [JPL NASA]



Open City Model (OCM)

Research Questions

"Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?"

Research Questions

"Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?"

1. What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?

Research Questions

“Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?”

1. What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?
2. What **relations** are present between the different **geometric properties** of the building footprints and the building height? And which **subset** is deemed ‘**optimal**’ for predicting building heights?

Research Questions

"Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?"

1. What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?
2. What **relations** are present between the different **geometric properties** of the building footprints and the building height? And which **subset** is deemed '**optimal**' for predicting building heights?
3. Are the **geometric properties** of the building footprints as training features **sufficient** for meeting the accuracy requirements?

Research Questions

“Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?”

1. What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?
2. What **relations** are present between the different **geometric properties** of the building footprints and the building height? And which **subset** is deemed ‘**optimal**’ for predicting building heights?
3. Are the **geometric properties** of the building footprints as training features **sufficient** for meeting the accuracy requirements?
4. What **other features,** besides the geometric properties of the building footprints, can be used in the machine learning algorithms to estimate building heights? And does including these features, even if they are incomplete, **improve the accuracy** of the estimations?

Research Questions

"Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?"

1. What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?
2. What **relations** are present between the different **geometric properties** of the building footprints and the building height? And which **subset** is deemed '**optimal**' for predicting building heights?
3. Are the **geometric properties** of the building footprints as training features **sufficient** for meeting the accuracy requirements?
4. What **other features**, besides the geometric properties of the building footprints, can be used in the machine learning algorithms to estimate building heights? And does including these features, even if they are incomplete, **improve the accuracy** of the estimations?
5. What **methods** can be used for **scaling** the machine learning techniques to the whole of the USA?

State-of-the-Art

CityGML CityJSON

3D city model encodings

[Gröger et al., 2012; Ledoux et al., 2019]

CityGML

CityJSON

3D city model encodings

[Gröger et al., 2012; Ledoux et al., 2019]

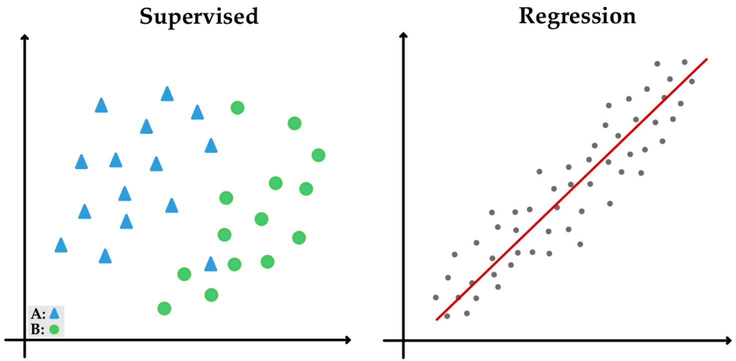
LOD1 roof reference points

[Biljecki et al., 2014]

CityGML {} CityJSON

3D city model encodings

[Gröger et al., 2012; Ledoux et al., 2019]

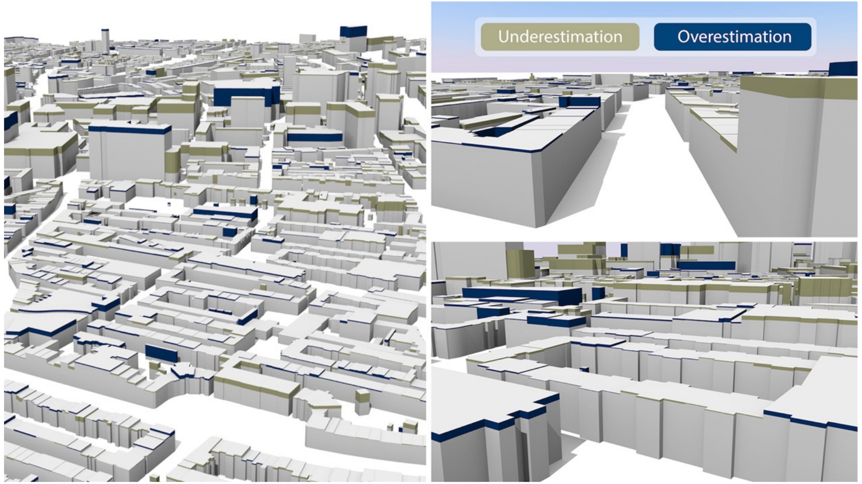


Machine learning

LOD1 roof reference points

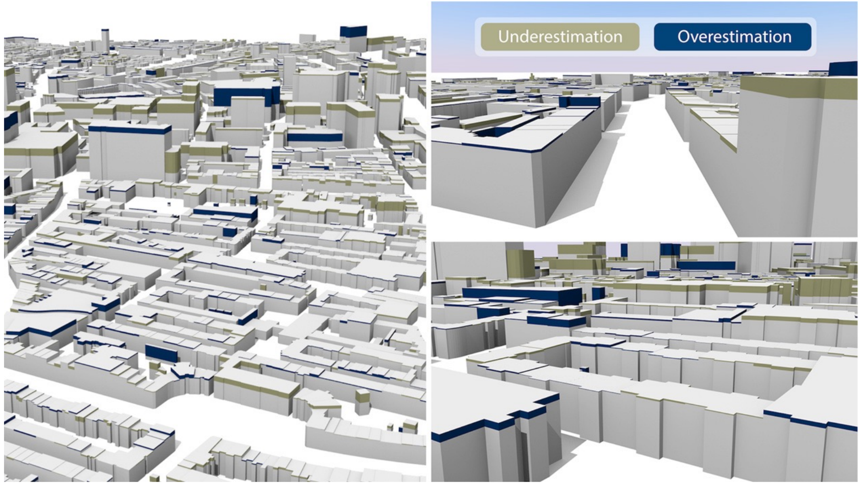
[Biljecki et al., 2014]

Machine Learning

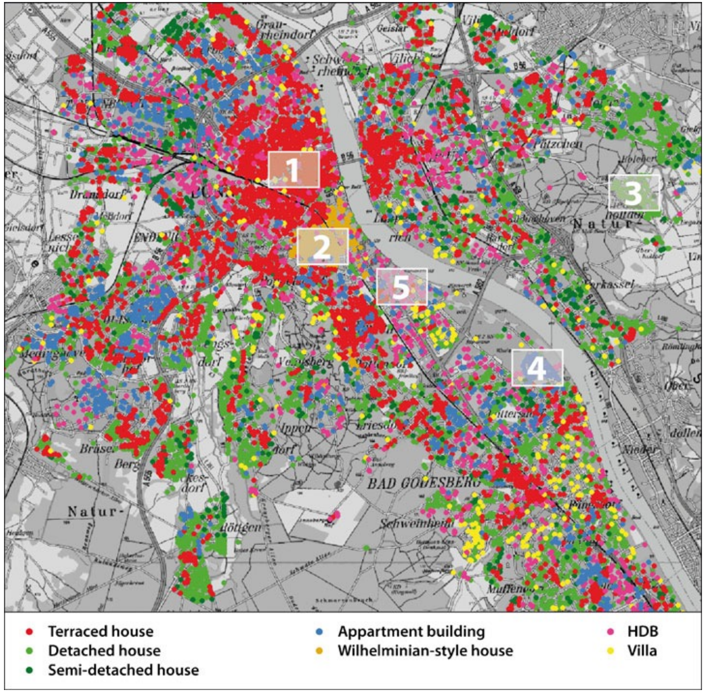


Random Forests – Rotterdam, The Netherlands
[Biljecki et al., 2017]

Machine Learning



Random Forests – Rotterdam, The Netherlands
[Biljecki et al., 2017]



Support Vector Machines – Building Classification
[Henn et al., 2012]

Contributions

1. Scale to a much larger extent than previously done, and deal with the diversity in built-up areas

Contributions

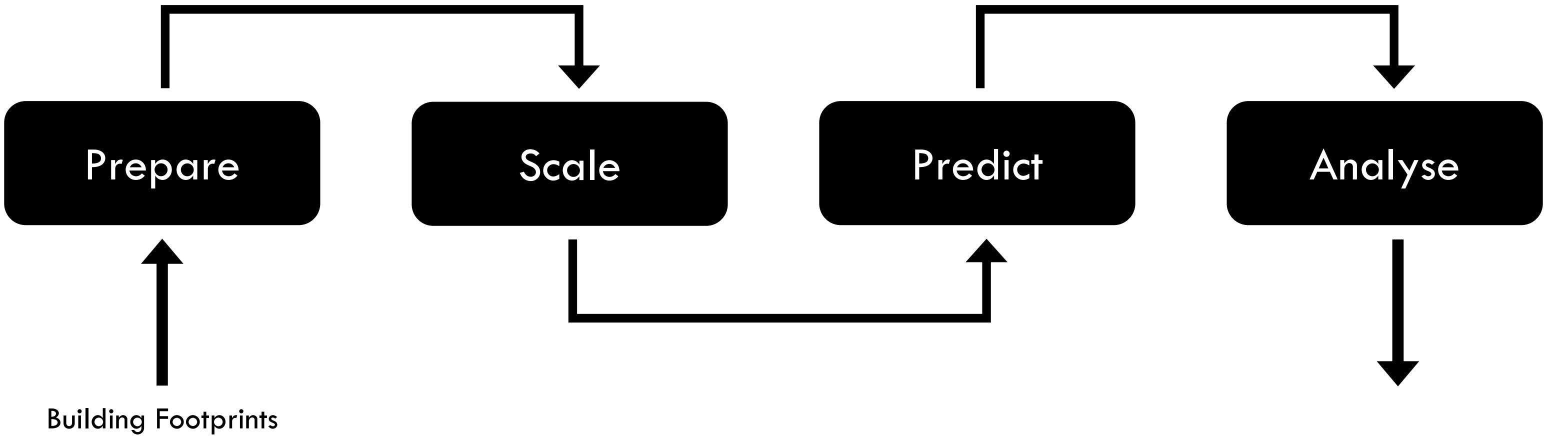
1. Scale to a much larger extent than previously done, and deal with the diversity in built-up areas
2. Investigate the possibility of only using geometric features for inferring the building heights, and try to determine an optimal subset for this purpose

Contributions

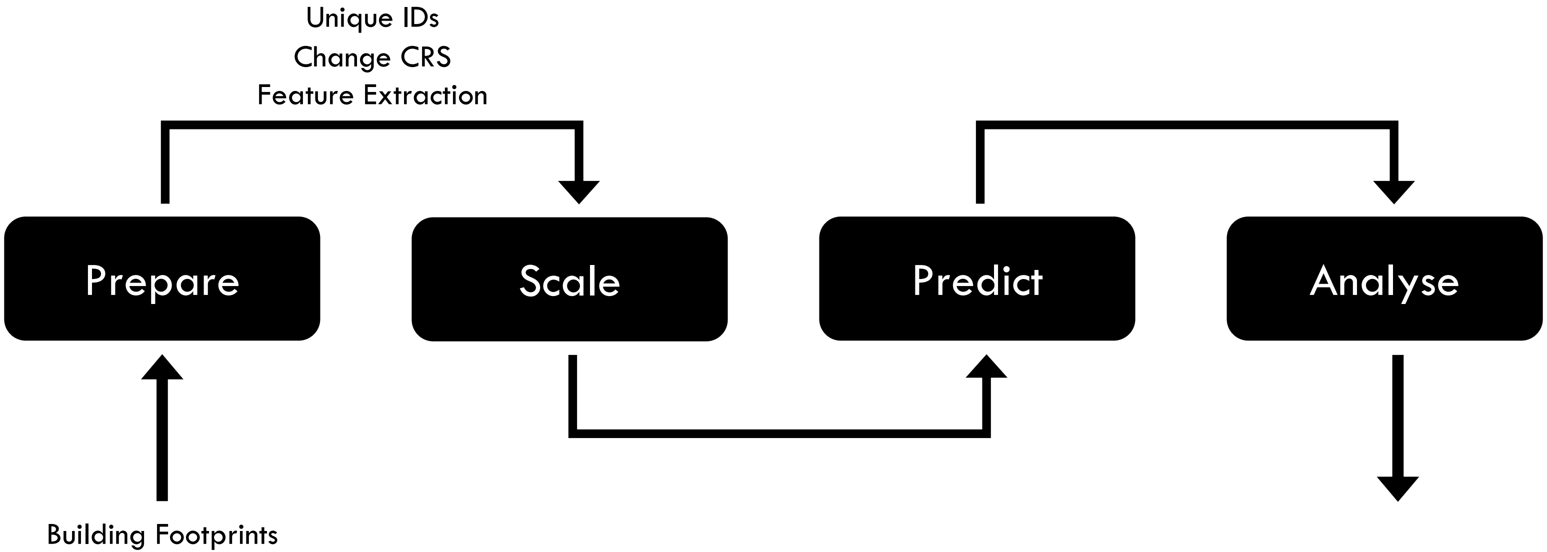
1. Scale to a much larger extent than previously done, and deal with the diversity in built-up areas
2. Investigate the possibility of only using geometric features for inferring the building heights, and try to determine an optimal subset for this purpose
3. Consider the different roof reference points and their influence on the height prediction results

Methodology

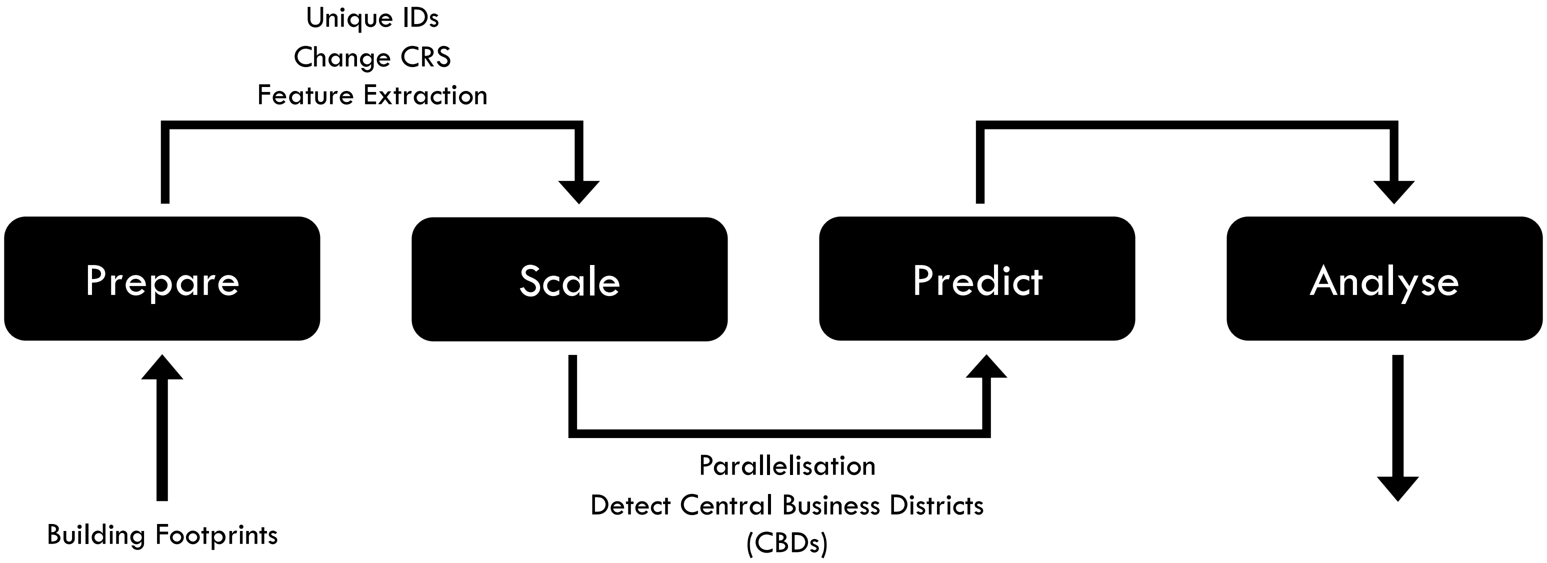
Overview



Overview


Unique IDs
Change CRS
Feature Extraction
Prepare
Scale
Predict
Analyse
Building Footprints

Overview


Unique IDs
Change CRS
Feature Extraction
Prepare
Scale
Predict
Analyse
Building Footprints
Parallelisation
Detect Central Business Districts
(CBDs)

Overview

Building Footprints

Prepare

Unique IDs
Change CRS
Feature Extraction

Scale

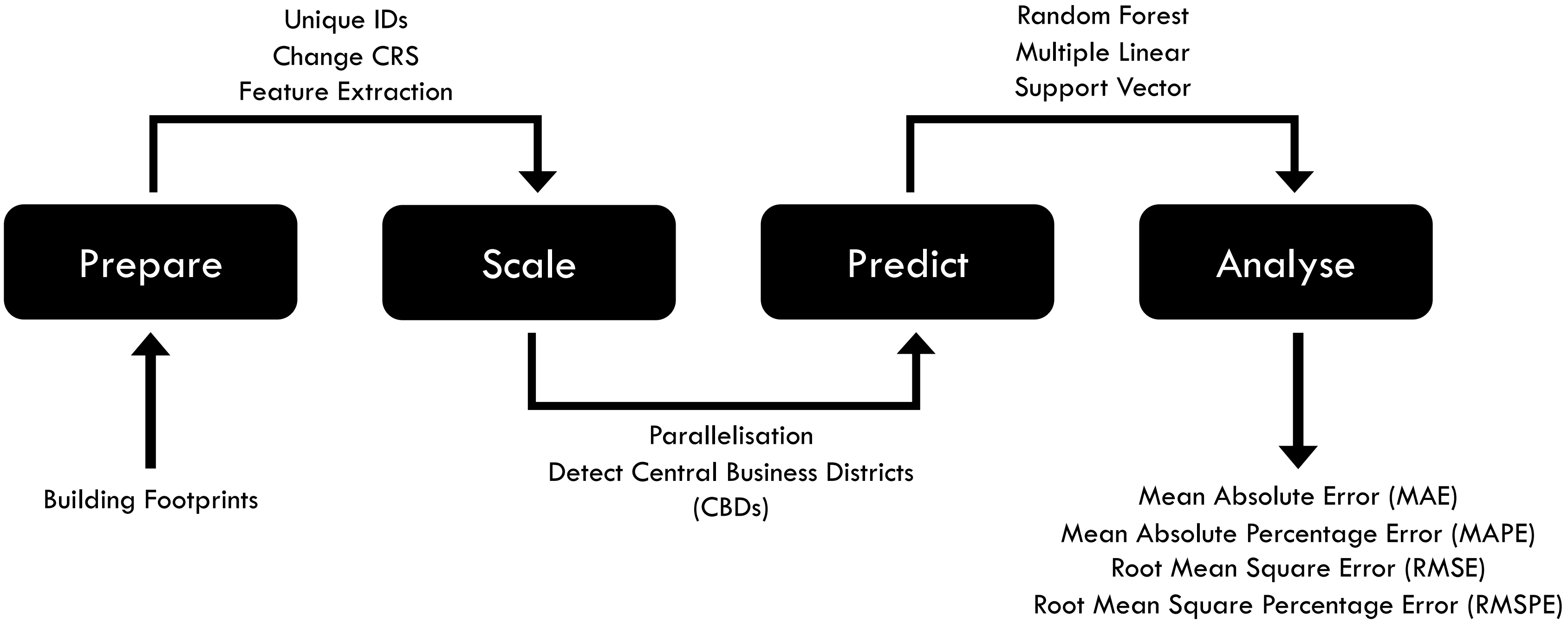
Parallelisation
Detect Central Business Districts (CBDs)

Predict

Random Forest
Multiple Linear
Support Vector

Analyse

Overview


Unique IDs
Change CRS
Feature Extraction
Random Forest
Multiple Linear
Support Vector
Prepare
Scale
Predict
Analyse
Building Footprints
Parallelisation
Detect Central Business Districts (CBDs)
Mean Absolute Error (MAE)
Mean Absolute Percentage Error (MAPE)
Root Mean Square Error (RMSE)
Root Mean Square Percentage Error (RMSPE)

Geometric Features

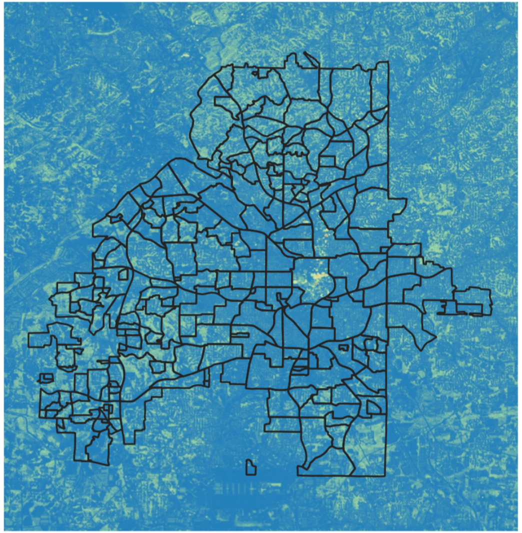
	Feature	Description	Computation
1.	**Area**	Footprint area	-
2.	**Compactness**	Normalised Perimeter Index	$\frac{2\sqrt{\pi A}}{P}$
3.	**Number of neighbours**	Buildings within 100m	Centroids
4.	Complexity	The irregularity of the footprint	$\frac{P}{\sqrt[4]{A}}$
5.	Number of adjacent buildings	Buildings within 1m	Buffers
6.	Length	Longest edge oriented MBR*	-
7.	Width	Shortest edge oriented MBR*	-
8.	Slimness	Side ratio	$\frac{Length}{Width}$
9.	Number of vertices	Number of vertices in footprint	-

* Minimum Bounding Rectangle

Detect CBDs

Atlanta, Georgia

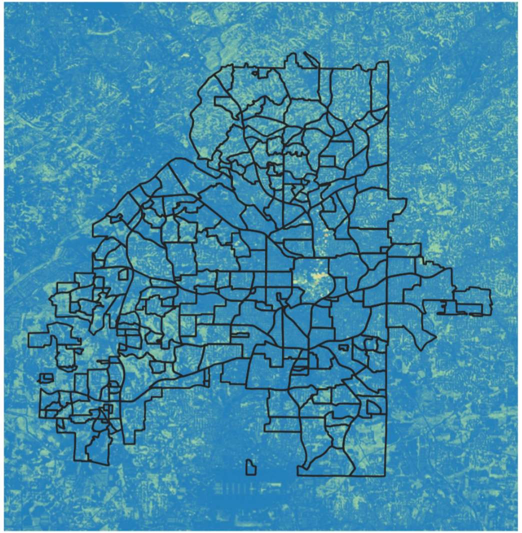
No filter



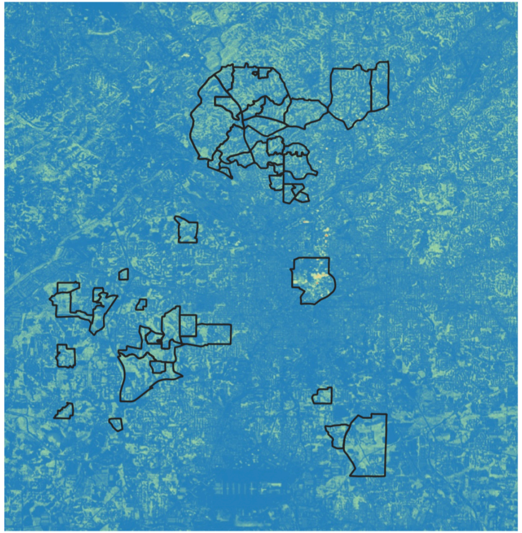
Detect CBDs

Atlanta, Georgia

No filter



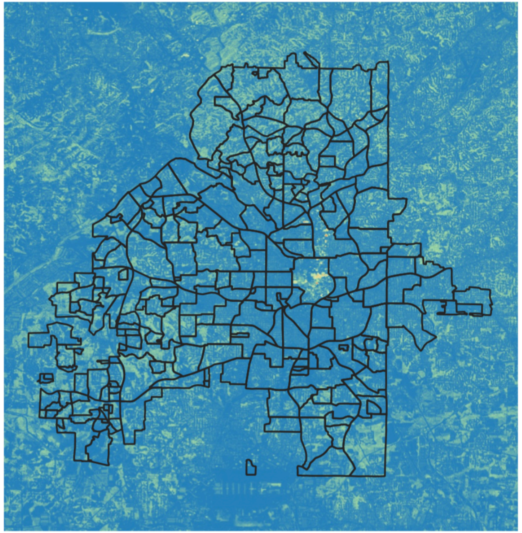
mean >= 11m



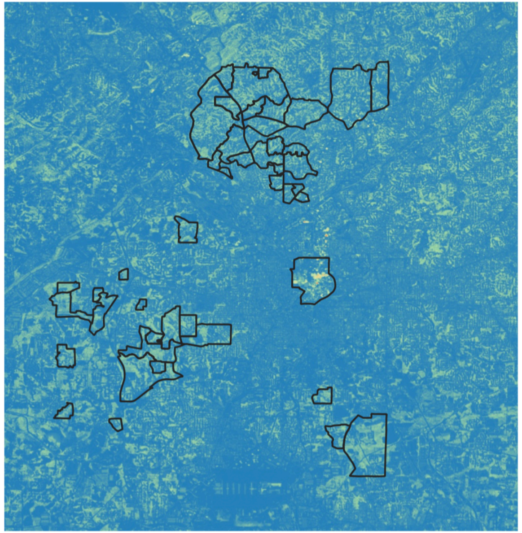
Detect CBDs

Atlanta, Georgia

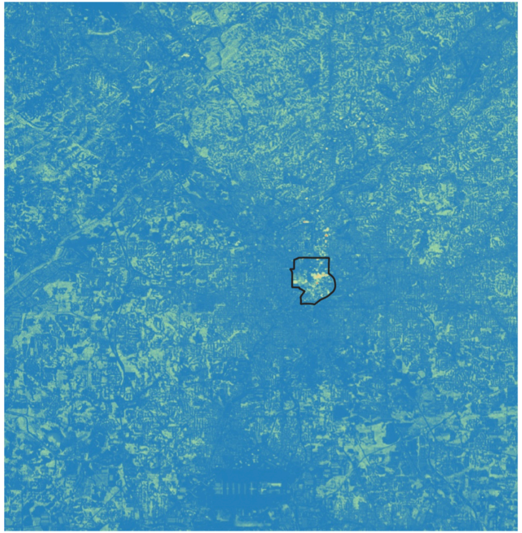
No filter



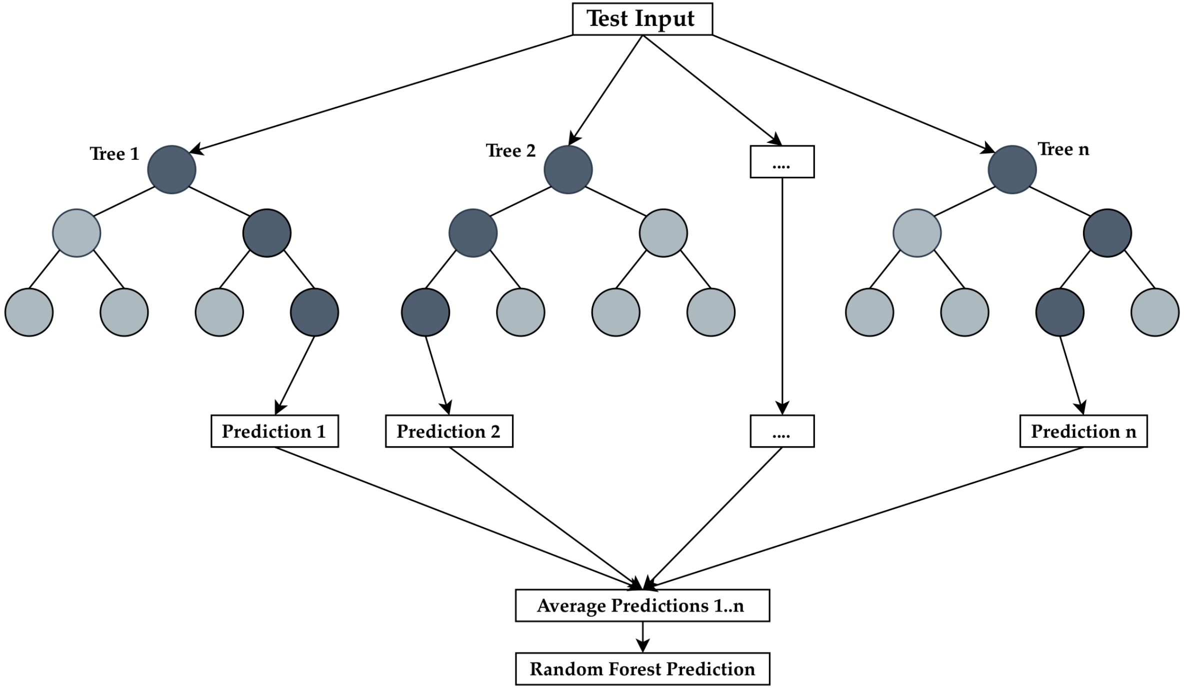
mean >= 11m



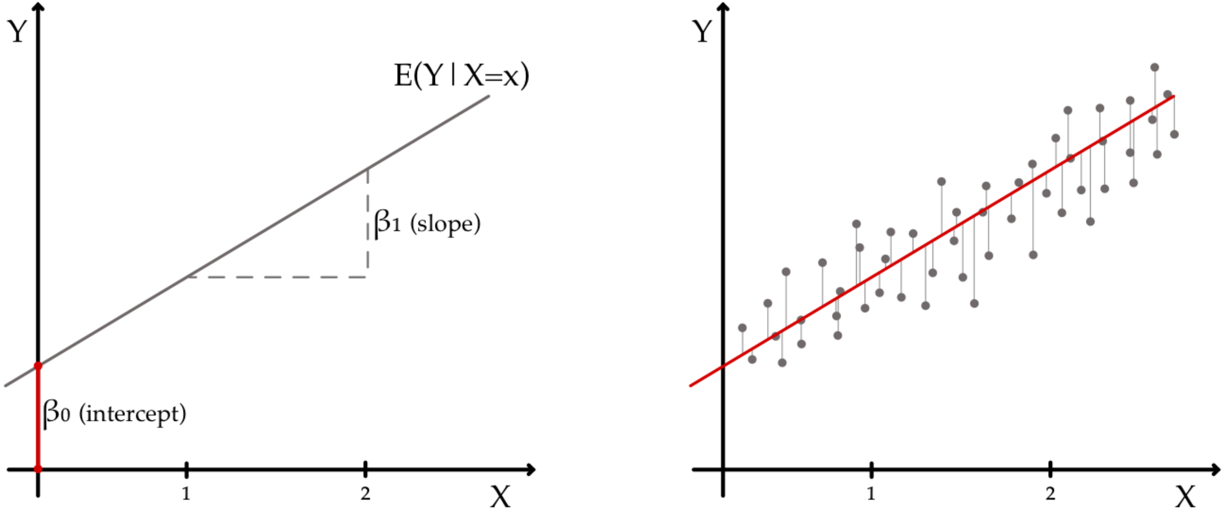
mean >= 11m AND max >= 100m



Random Forest Regression



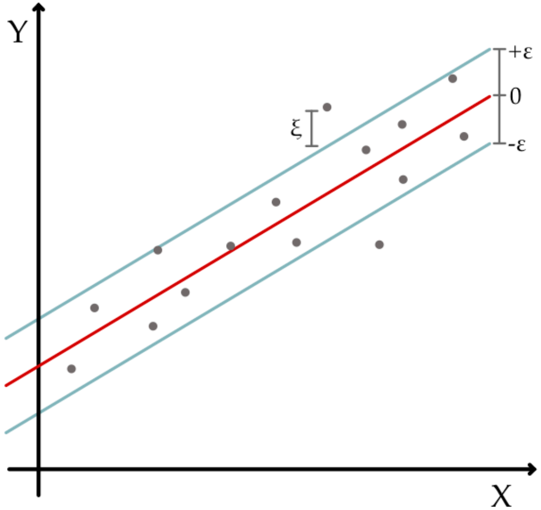
Multiple Linear Regression



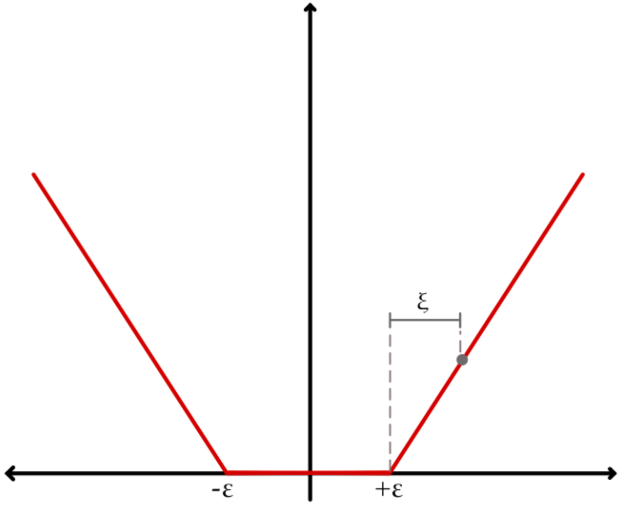
Simple linear regression – 1D example

Support Vector Regression

ε-boundary and slack variables



ε-insensitive loss function



Implementation

Data & Software

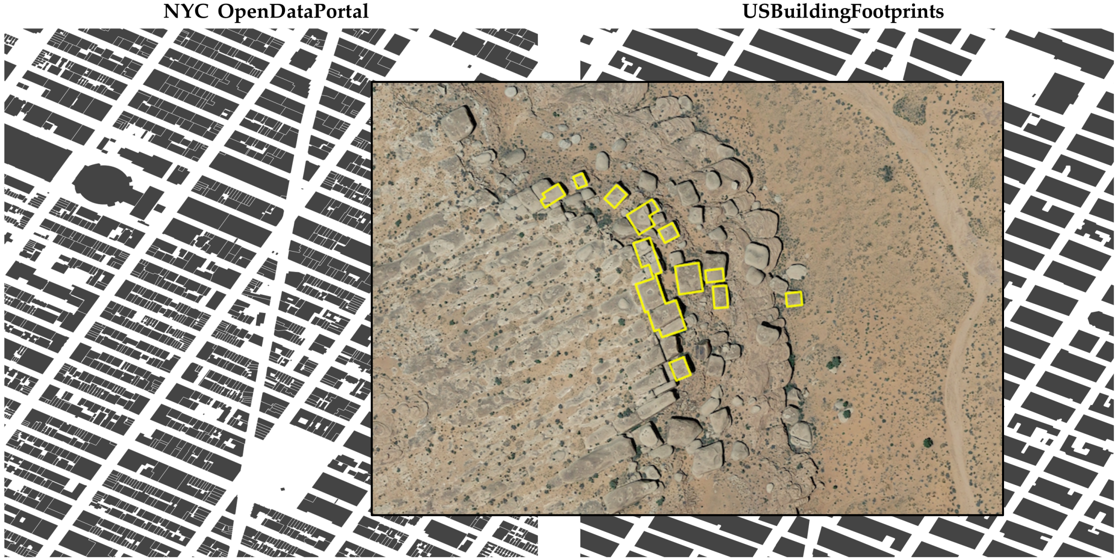
USBuildingFootprints dataset

NYC OpenDataPortal

USBuildingFootprints

Data & Software

USBuildingFootprints dataset

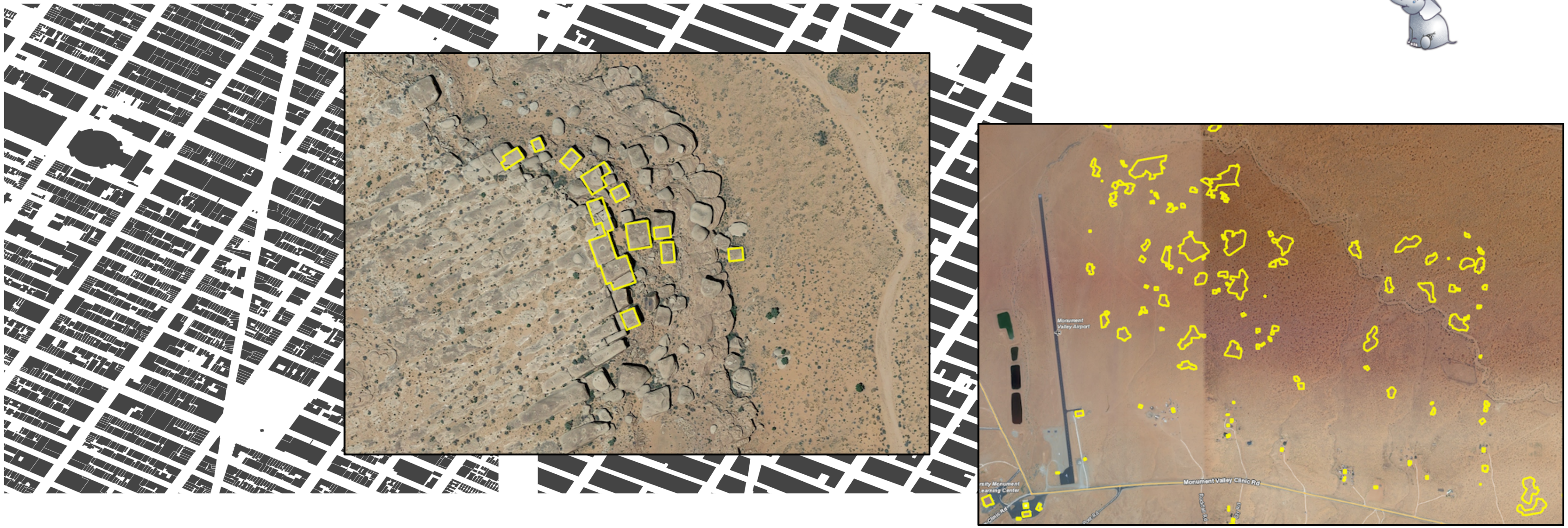


Data & Software

USBuildingFootprints dataset

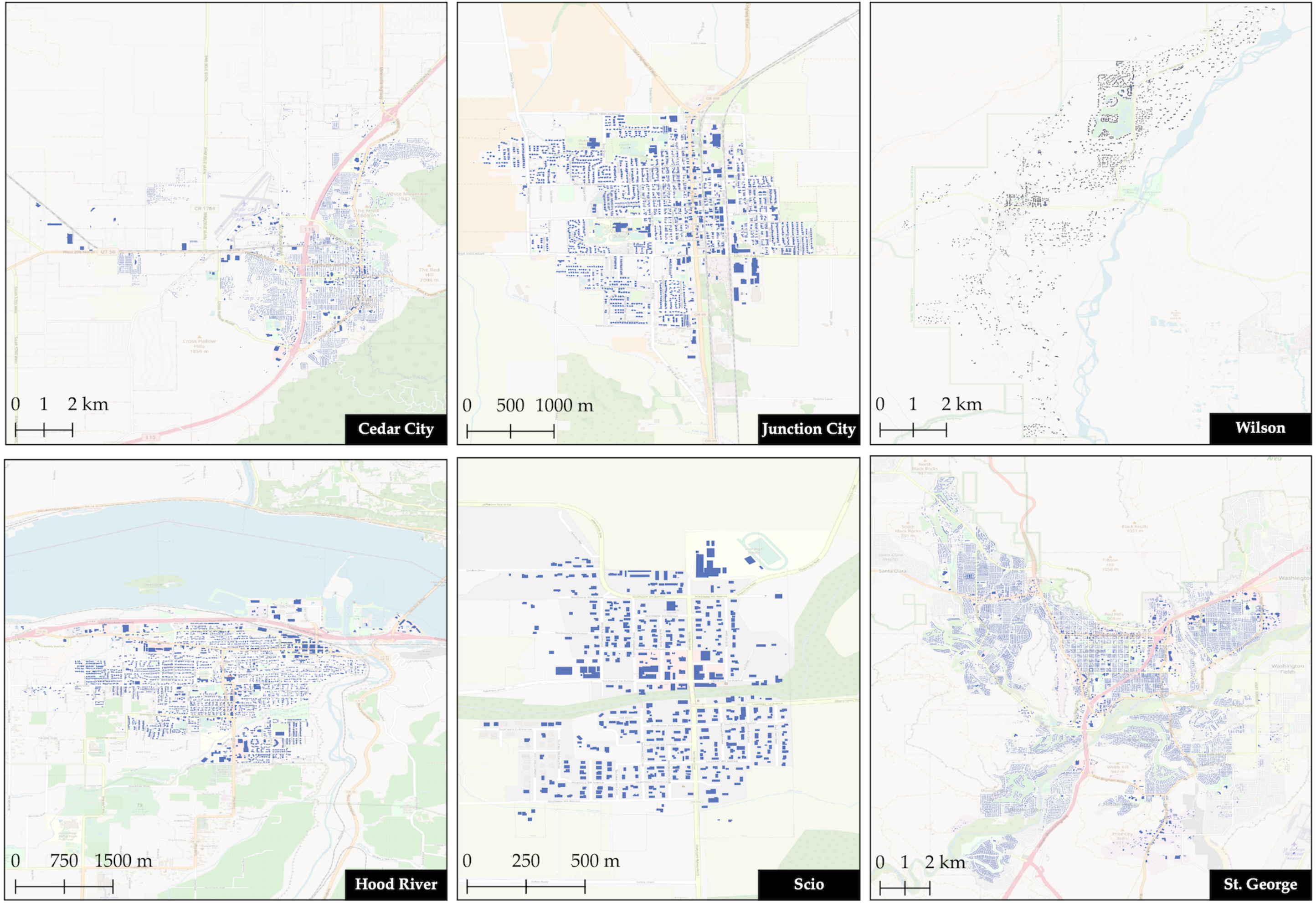
NYC OpenDataPortal

USBuildingFootprints

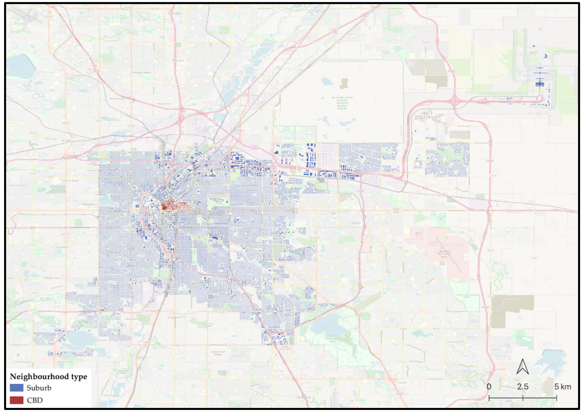


Suburban & Rural Data

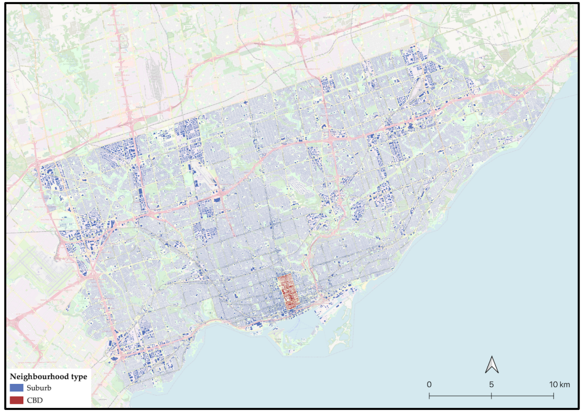
Own models: *USBuildingFootprints* and LiDAR data



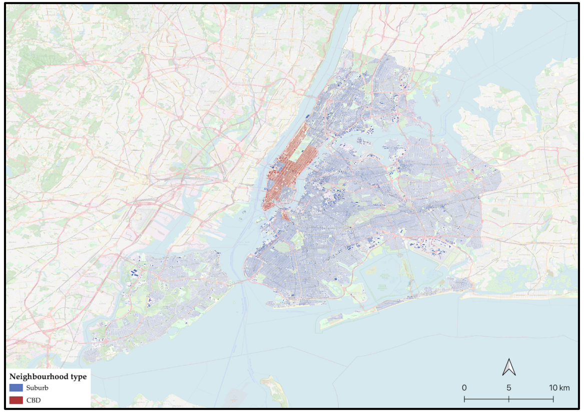
CBD & Suburban Data



Denver



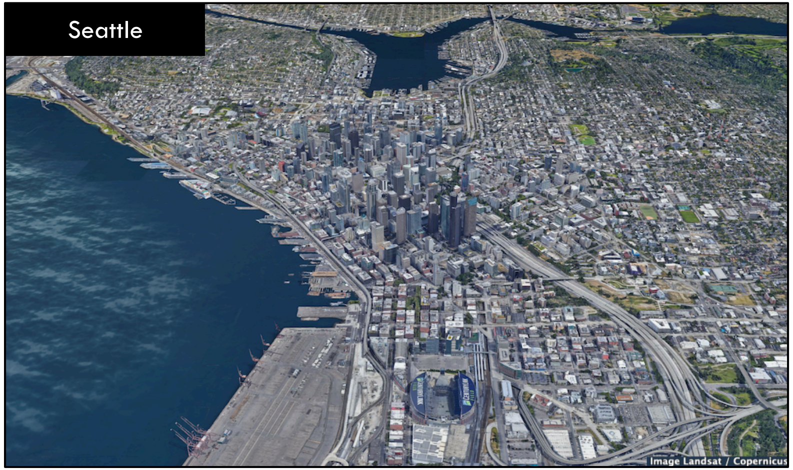
Toronto



New York City

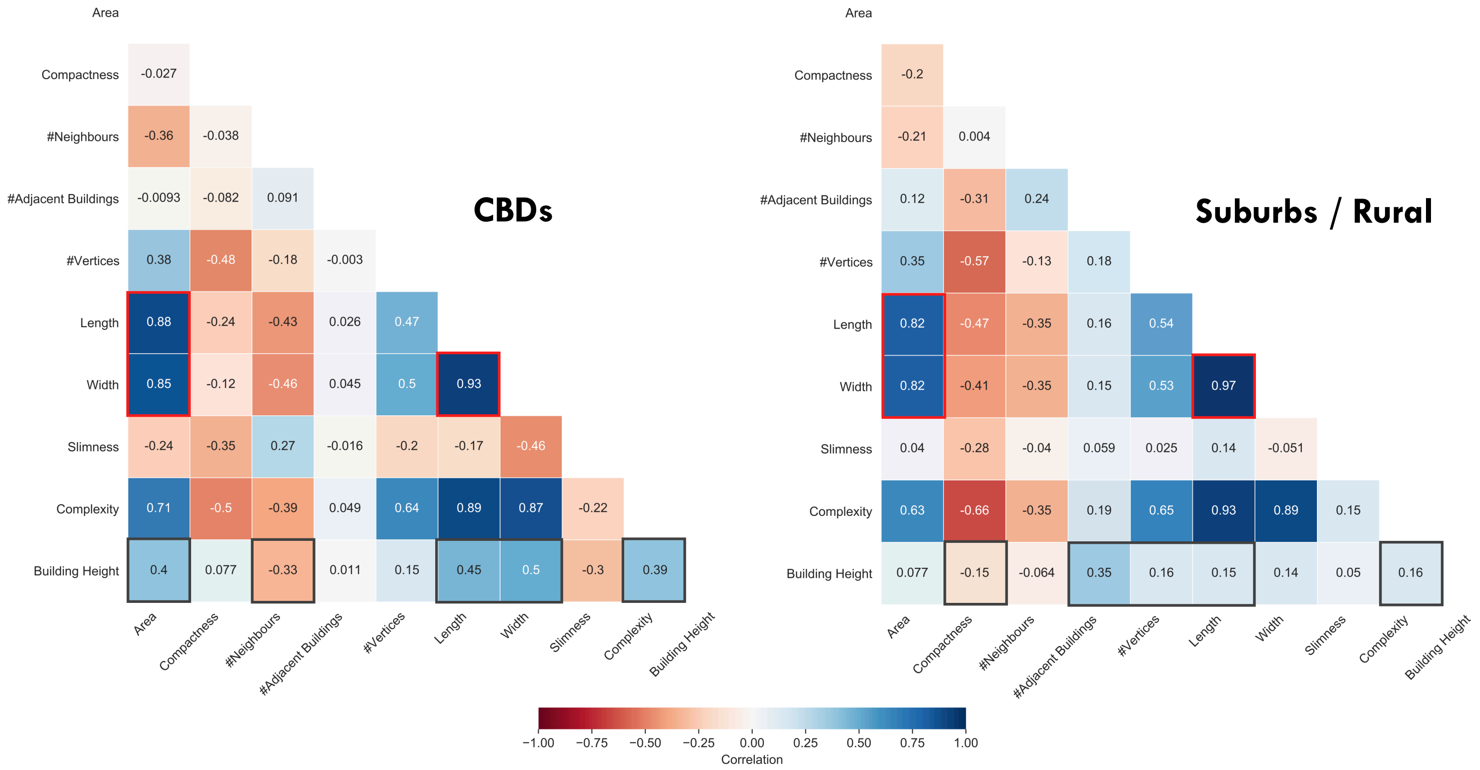
Test Areas


Astoria
Google Earth

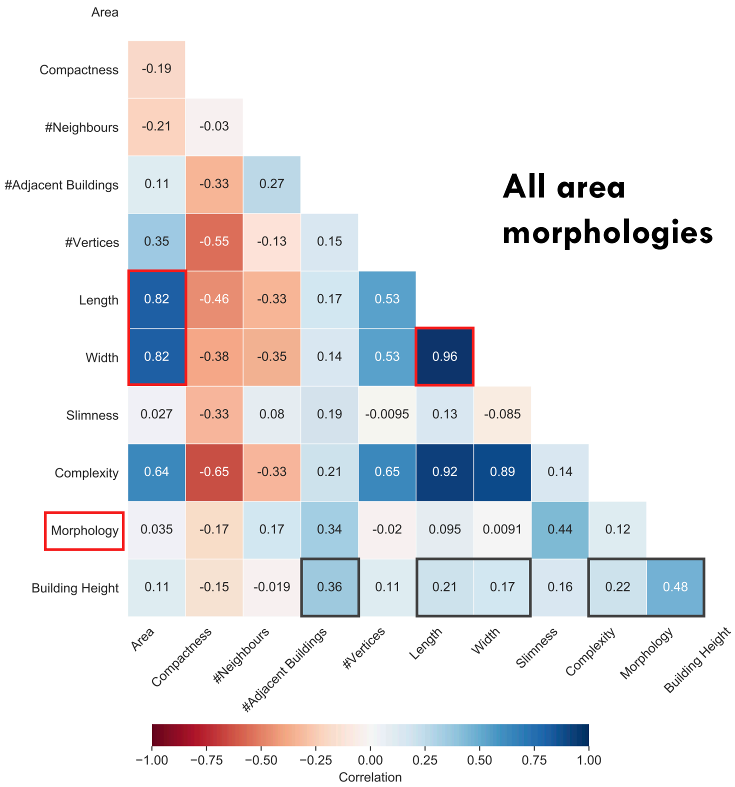

Seattle
Image Landsat / Copernicus


Portland
Image Landsat / Copernicus

Contribution Geometric Features



Contribution Geometric Features



All area morphologies

Results

Methodology Runtime

Regressor	Training time* [s]		
	Suburban / Rural model	*CBD model*	*Combined model*
RFR	13.96	2.62	17.03
MLR	0.12	0.01	0.50
SVR	0.70	0.04	2.23

* Average of 10 runs

Methodology Runtime

Regressor	Training time* [s]		
	Suburban / Rural model	*CBD model*	*Combined model*
RFR	13.96	2.62	17.03
MLR	0.12	0.01	0.50
SVR	0.70	0.04	2.23

* Average of 10 runs

Regressor	Predicting time [mm:ss]	
	Split model	*Combined model*
RFR	03:38.03	05:56.08
MLR	00:09.73	00:23.13
SVR	00:10.55	00:26.91

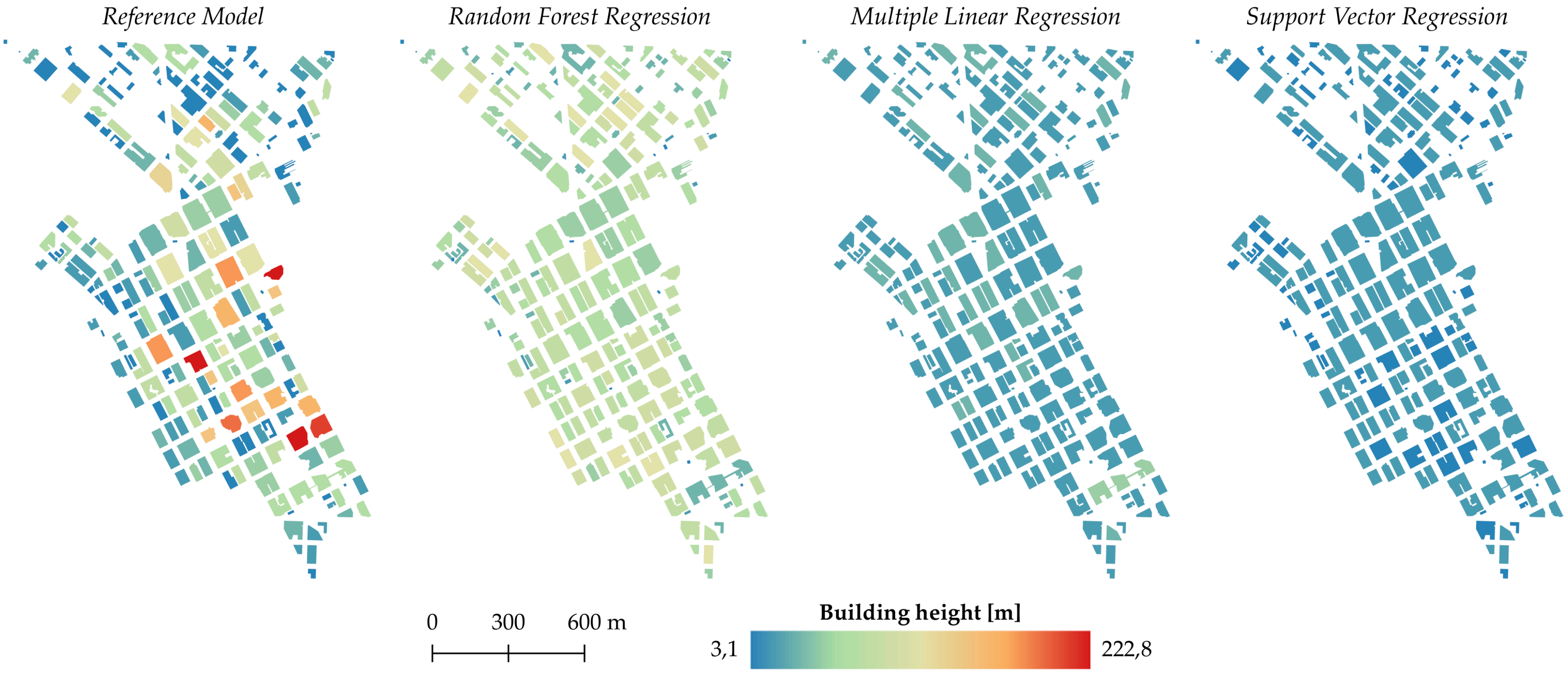
~125 million building footprints

Model Accuracy

Regressor		Seattle		Portland		Astoria	
		CBD	*Combined*	*Suburban / Rural*	*Combined*	*Suburban / Rural*	*Combined*
RFR	*MAE [m]*	40.54	39.74	1.42	1.42	2.29	2.29
	MAPE [%]	224.93	216.87	24.77	24.92	28.90	28.91
	RMSE [m]	48.58	47.91	2.36	2.36	2.99	2.98
	RMSPE [%]	361.21	351.15	32.64	32.76	36.00	35.95
MLR	*MAE [m]*	37.09	32.84	1.67	1.77	2.28	2.30
	MAPE [%]	218.27	117.57	27.27	29.30	29.30	29.91
	RMSE [m]	44.73	49.66	2.61	2.68	2.93	2.94
	RMSPE [%]	341.09	186.50	32.59	36.72	35.04	36.32
SVR	*MAE [m]*	36.83	34.88	1.65	1.41	2.27	2.51
	MAPE [%]	216.12	78.68	26.79	22.64	29.08	30.13
	RMSE [m]	44.44	55.25	2.58	2.39	2.92	3.21
	RMSPE [%]	337.03	107.51	31.91	26.64	34.58	34.21

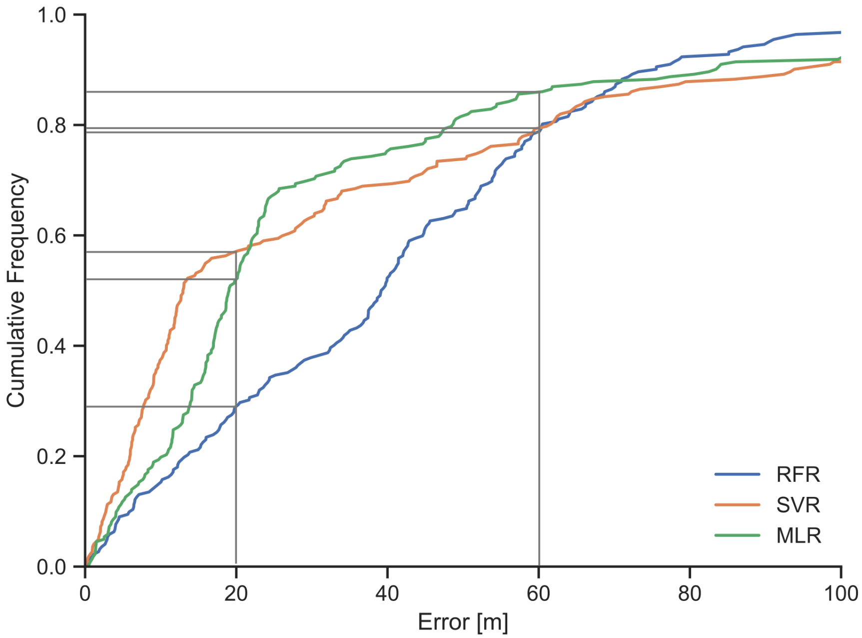
Prediction model with area morphology as additional feature

Model Accuracy – Seattle CBD

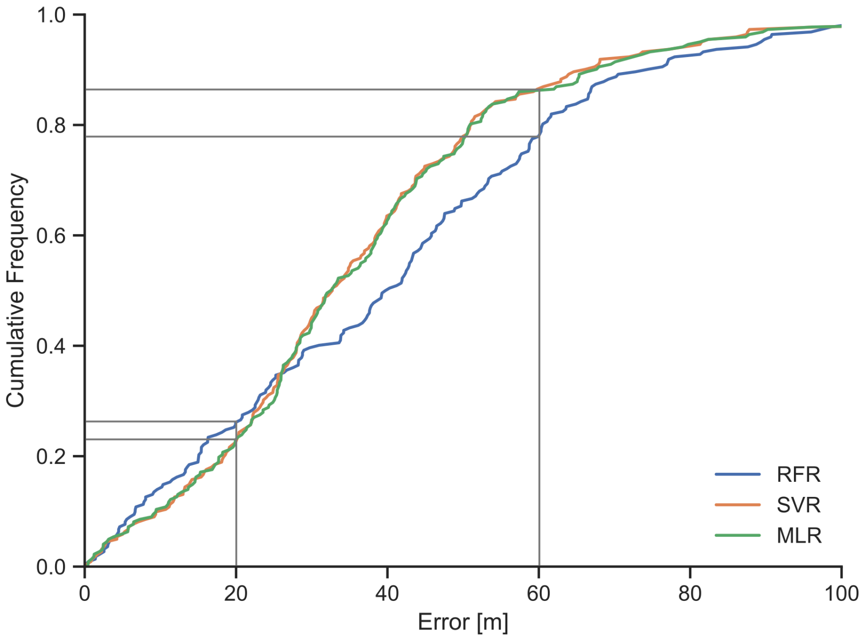


Model Accuracy – Seattle CBD

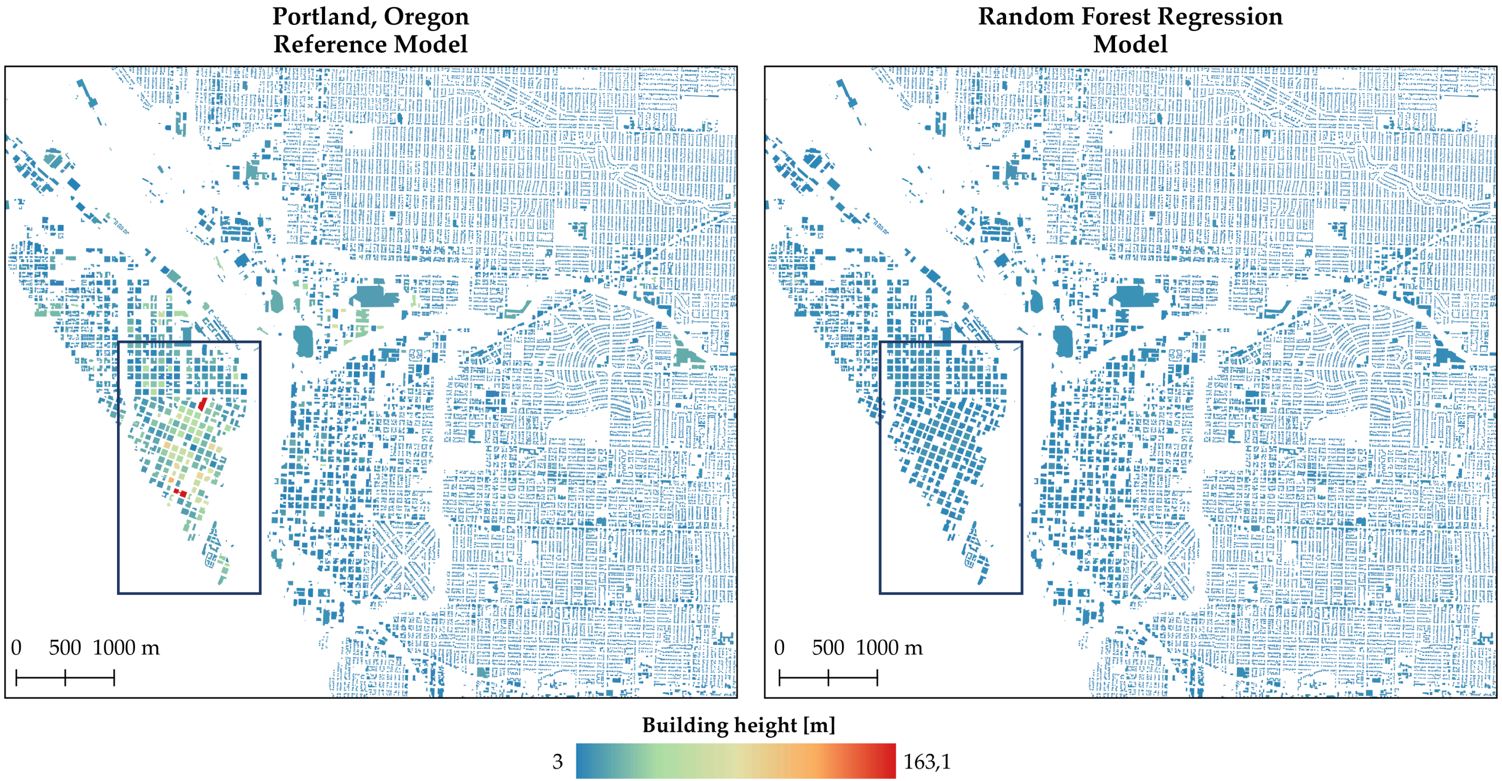
Combined prediction model



CBD prediction model

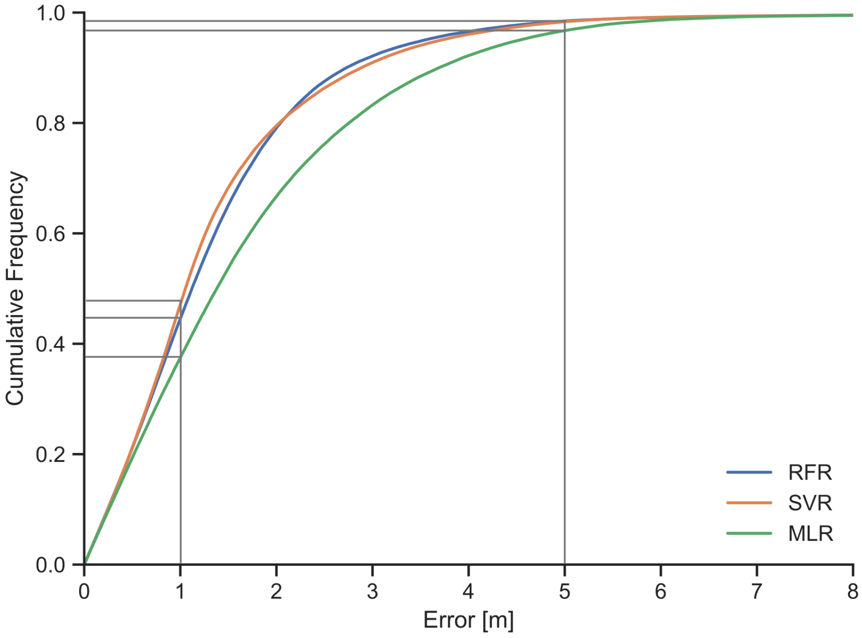


Model Accuracy – Portland

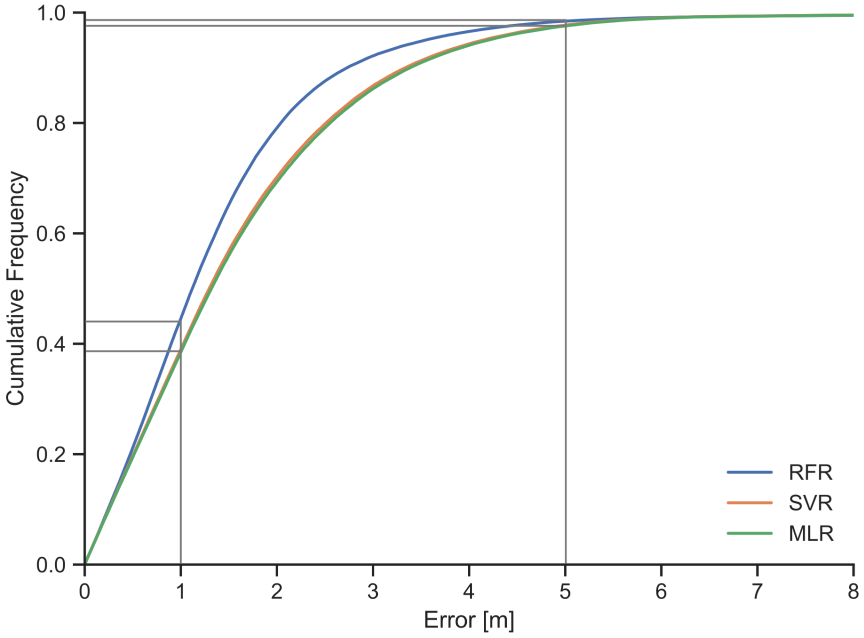


Model Accuracy – Portland

Combined prediction model

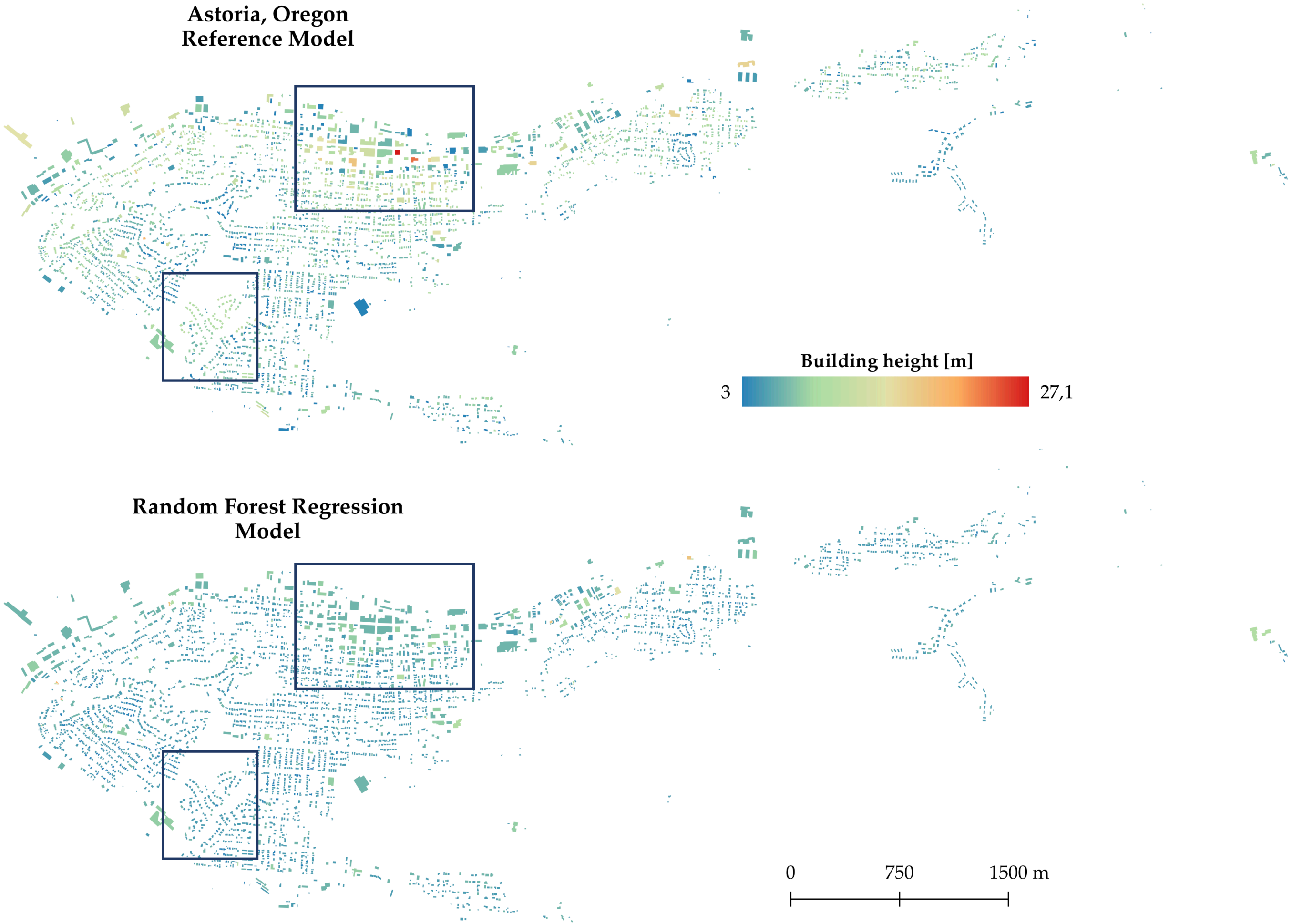


Suburban / Rural prediction model



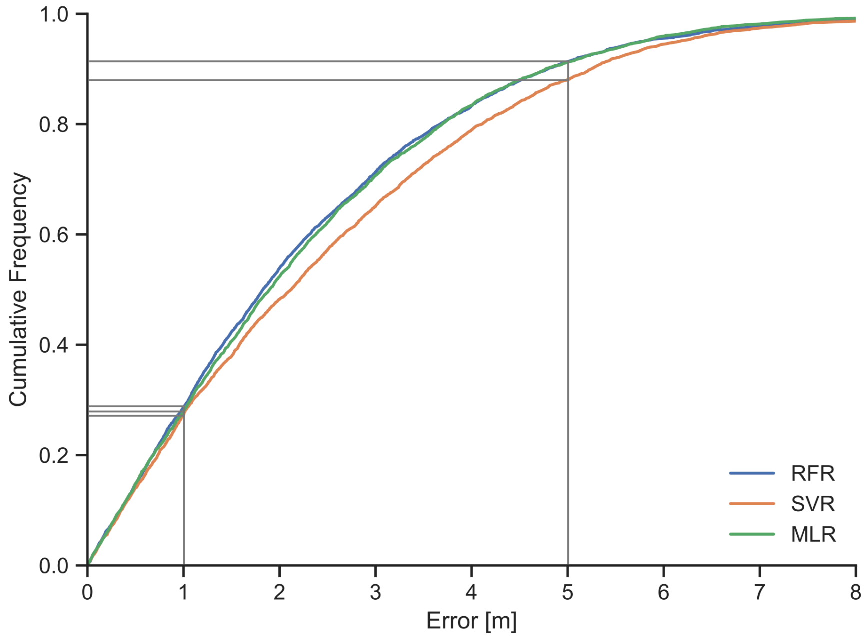
Model Accuracy – Astoria

Prediction model trained on rural and suburban data

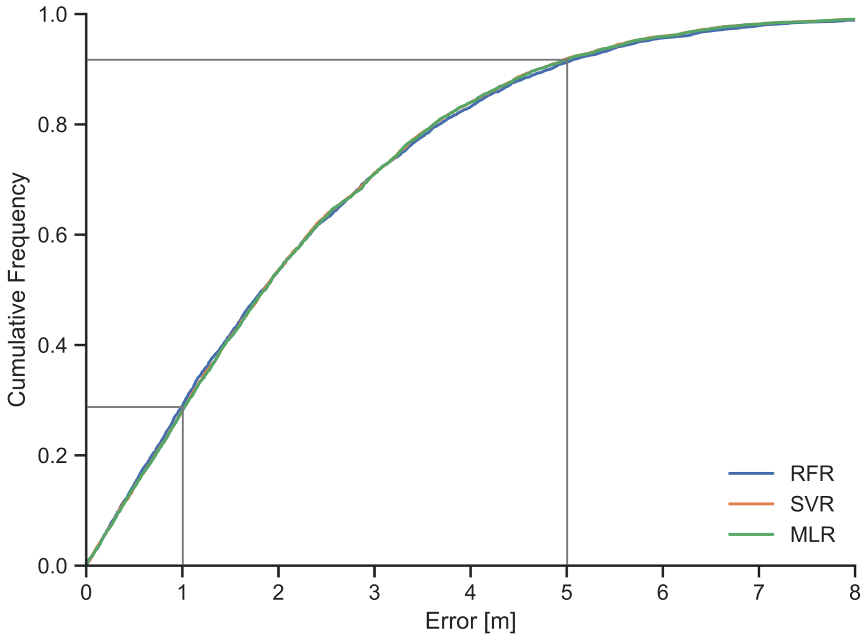


Model Accuracy – Portland

Combined prediction model



Suburban / Rural prediction model



Comparison OCM

Prediction model trained on rural and suburban data

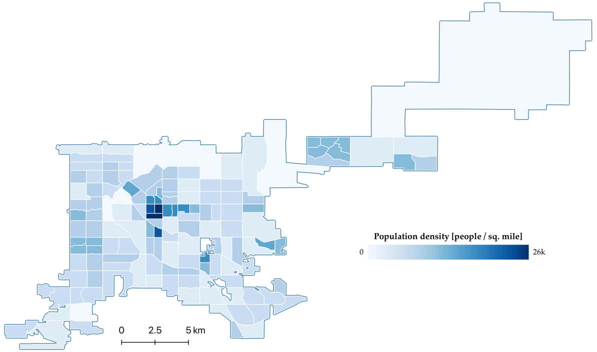
Error Measure	OCM	SVR
MAE [m]	2.16	2.09
MAPE [%]	27.29	27.89
RMSE [m]	2.76	2.64
RMSPE [%]	31.56	33.34

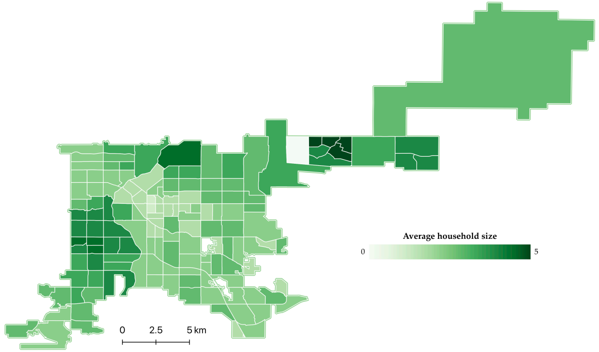
Support Vector Regression Model

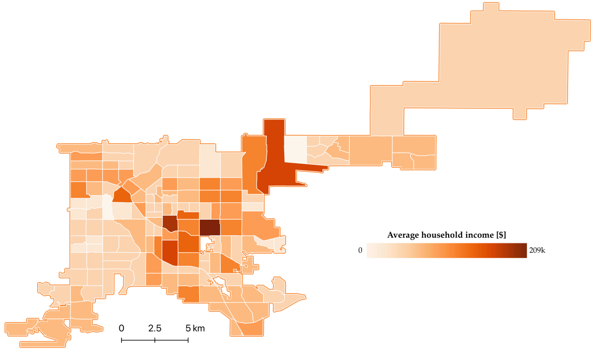
Open City Model

Non-Geometric Features Denver

- *Census*:
  - Population density
  - Avg. household income
  - Avg. household size
- *Cadastral*:
  - Building type
- *Miscellaneous:*
  - #Amenities
  - Raster building height





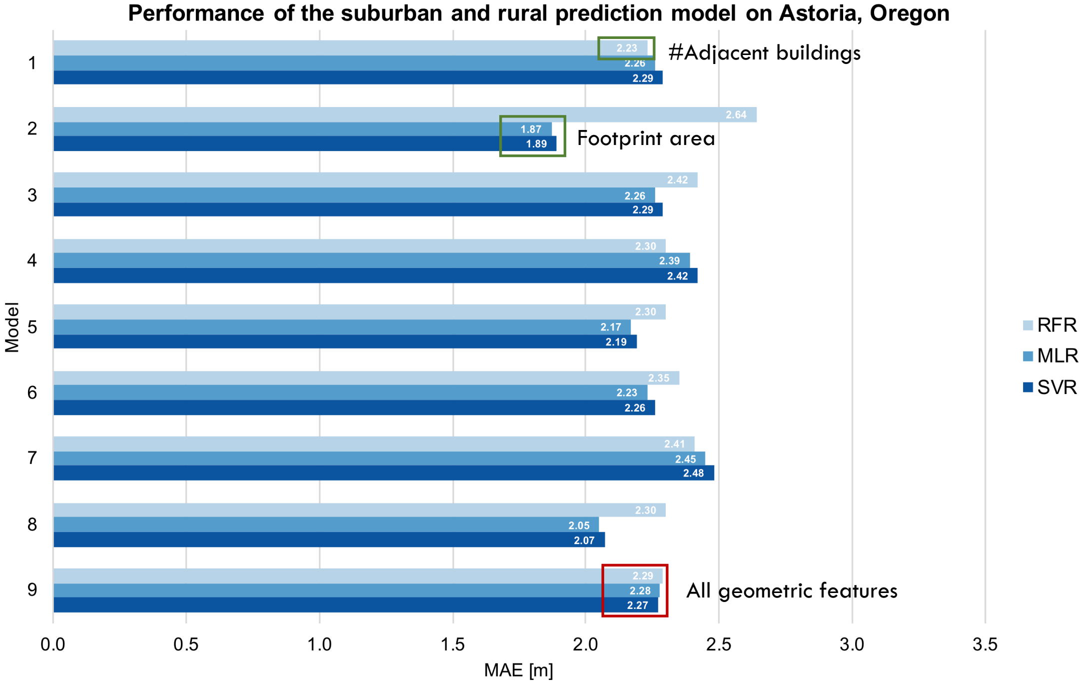


Non-Geometric Features Denver

Regressor		Suburbs		CBDs		Combined	
		Base	*Enriched*	*Base*	*Enriched*	*Base*	*Enriched*
RFR	*MAE [m]*	1.35	0.96	20.84	17.29	1.44	1.03
	MAPE [%]	22.05	15.68	152.40	114.66	22.67	16.02
	RMSE [m]	2.71	2.11	30.68	27.12	3.55	2.93
	RMSPE [%]	33.30	25.08	267.17	208.19	36.97	28.52
MLR	*MAE [m]*	1.59	1.47	21.33	16.87	1.72	1.60
	MAPE [%]	26.81	25.16	158.71	109.09	27.99	26.66
	RMSE [m]	2.93	2.58	31.55	28.57	3.80	3.37
	RMSPE [%]	35.05	33.73	246.51	200.89	40.21	37.75
SVR	*MAE [m]*	1.55	1.46	26.10	25.80	1.68	1.59
	MAPE [%]	23.94	23.79	87.49	89.80	25.12	24.93
	RMSE [m]	2.98	2.64	41.45	39.83	3.84	3.41
	RMSPE [%]	31.21	31.07	88.19	95.24	36.17	34.84

Additional Results

- **Feature subsets**
  - Highly dependent on test area and prediction model
- **Height percentiles**
  - CBDs: 90^{th} percentile
  - Other: 50^{th} percentile



Conclusions

1.

*“Can the **125 million** USA building footprints be assigned a height **without** making use of **height data**, and what **accuracy** can be achieved?”*

YES

CBDs problematic (MAE 32.84m)

Suburban / **Rural** areas more promising (MAE 1.41m)

1.1

*“What **methods** can be used to **assess the accuracy** of the building height estimations? And when are the estimations deemed **accurate enough**?”*

MAE, MAPE, RMSE, RMSPE

CityGML specification: 5m suggestion

1.2

*“What **relations** are present between the different **geometric properties** of the building footprints and the building height? And which **subset** is deemed ‘**optimal**’ for predicting building heights?”*

CBDs: clear (linear) relations (length, width)
Suburban/ **Rural**: less clear (#adjacent buildings)
Combined: less clear (area morphology)

Not one subset best for all test areas and prediction models

1.3

*“Are the **geometric properties** of the building footprints as training features **sufficient** for meeting the accuracy requirements?”*

PARTIALLY

CBDs: no
Suburbs / Rural: yes

1.4

*"What **other features**, …, can be used in the machine learning algorithms to estimate building heights? And does including these features, even if they are incomplete, **improve the accuracy** of the estimations?"*

YES

Census & cadastral information
#Amenities & raster heights

1.5

*“What **methods** can be used for **scaling** the machine learning techniques to the whole of the USA?”*

Parallelisation of processes

Detect area morphologies:
differently trained prediction models

Discussion

1. Methods partially reliable on building height data

Discussion

1. Methods partially reliable on building height data
2. Reference models introduce uncertainty
   i. Both during training and testing phase

Discussion

1. Methods partially reliable on building height data
2. Reference models introduce uncertainty
    i. Both during training and testing phase
3. Only small area for non-geometric features

Discussion

1. Methods partially reliable on building height data
2. Reference models introduce uncertainty
   i. Both during training and testing phase
3. Only small area for non-geometric features
4. Prediction model without area morphology

Discussion

1. Methods partially reliable on building height data
2. Reference models introduce uncertainty
   i. Both during training and testing phase
3. Only small area for non-geometric features
4. Prediction model without area morphology
5. Comparison Open City Model

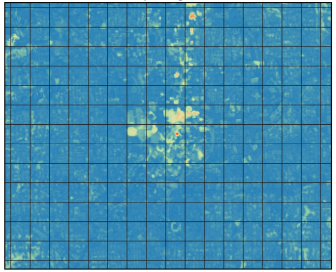
Discussion

1. Methods partially reliable on building height data
2. Reference models introduce uncertainty
   i. Both during training and testing phase
3. Only small area for non-geometric features
4. Prediction model without area morphology
5. Comparison Open City Model
6. Area morphology detection

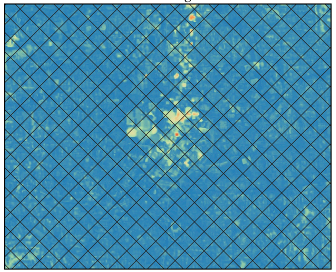
Suggestions for Future Work

1. Improve CBD detection process [suggested at P4]

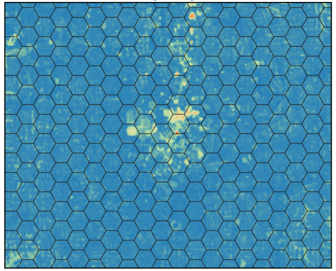
Square grid



Diamond grid

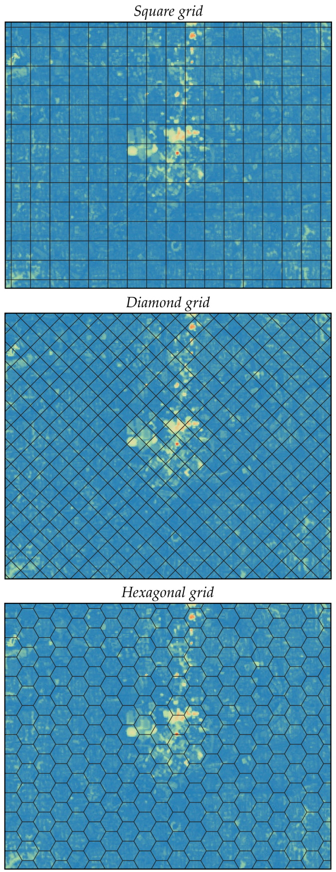


Hexagonal grid



Suggestions for Future Work

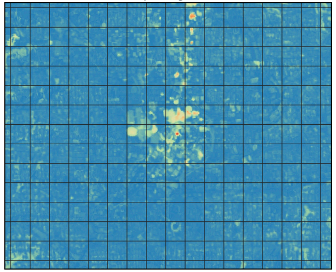
1. Improve CBD detection process [suggested at P4]
2. Additional area morphologies [suggested at P4]



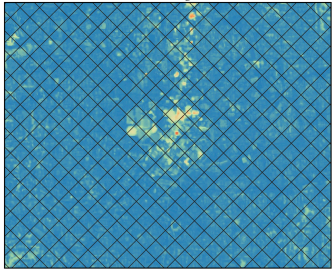
Suggestions for Future Work

1. Improve CBD detection process [suggested at P4]
2. Additional area morphologies [suggested at P4]
3. Extra non-geometric features, e.g.
   i. US-wide census and cadastral data
   ii. Shadows in satellite imagery

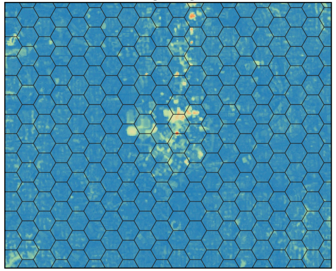
Square grid



Diamond grid



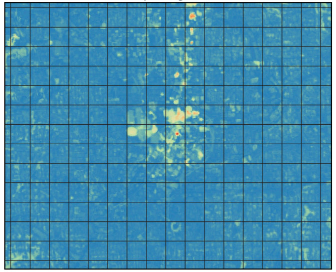
Hexagonal grid



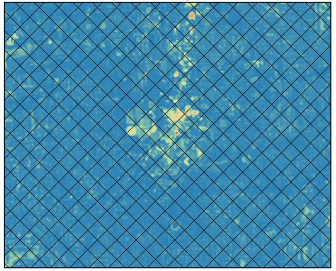
Suggestions for Future Work

1. Improve CBD detection process [suggested at P4]
2. Additional area morphologies [suggested at P4]
3. Extra non-geometric features, e.g.
   i. US-wide census and cadastral data
   ii. Shadows in satellite imagery
4. Test more feature subsets

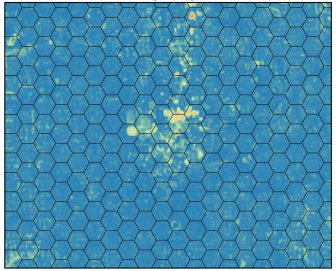
Square grid



Diamond grid

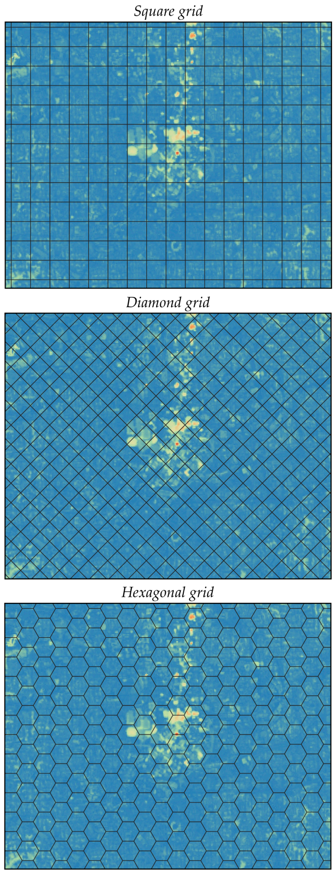


Hexagonal grid



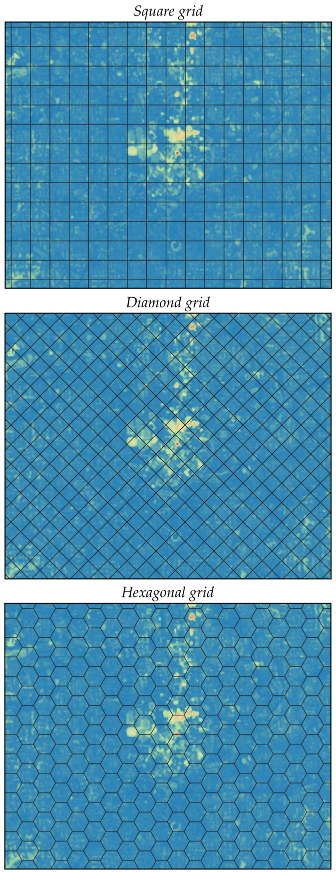
Suggestions for Future Work

1. Improve CBD detection process [suggested at P4]
2. Additional area morphologies [suggested at P4]
3. Extra non-geometric features, e.g.
   i. US-wide census and cadastral data
   ii. Shadows in satellite imagery
4. Test more feature subsets
5. Higher diversity in training data
   i. Mainly for CBDs



Suggestions for Future Work

1. Improve CBD detection process [suggested at P4]
2. Additional area morphologies [suggested at P4]
3. Extra non-geometric features, e.g.
   i. US-wide census and cadastral data
   ii. Shadows in satellite imagery
4. Test more feature subsets
5. Higher diversity in training data
   i. Mainly for CBDs
6. Extra testing areas
   i. Experiment with more granular footprints



Thank you for your attention. Are there any questions?

Icon: [freepick]

References

- Biljecki, F., Ledoux, H., and Stoter, J. (2014). Height references of CityGML LOD1 buildings and their influence on applications. *Proceedings. 9th ISPRS 3DGeoInfo Conference*.
- Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., and Çöltekin, A. (2015). Applications of 3D City Models: State of the Art Review. *ISPRS International Journal of Geo-Information*, 4:2842–2889.
- Biljecki, F., Ledoux, H., and Stoter, J. (2016). An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59:25–37.
- Biljecki, F., Ledoux, H., and Stoter, J. (2017). Generating 3D city models without elevation data. *Comput- ers, Environment and Urban Systems*, 64:1–18.
- Gröger, G., Kolbe, T. H., Nagel, C., and Häfele, K.-H. (2012). *OGC City Geography Markup Language (CityGML) Encoding Standard*. Open Geospatial Consortium.
- Henn, A., Römer, C., Gröger, G., and Plümer, L. (2012). Automatic classification of building types in 3D city models. *GeoInformatica*, 16(2):281–306.
- Ledoux, H., Ohori, K. A., Kumar, K., Dukai, B., Labetski, A., and Vitalis, S. (2019). CityJSON: a compact and easy-to-use encoding of the CityGML data model. *Open Geospatial Data, Software and Standards*, 4(1-12).