Mercator Research Institute on
Global Commons and Climate Change gGmbH

Estimating energy demand of buildings…
by learning their heights

Nikola Milojevic-Dupont, *Peter-Paul Pichler,*
Lynn H. Kaack, Steffen Lohrey, and Felix Creutzig

1. Infrastructure data in OSM are of high relevance for the sustainability science community.

2. Currently, OSM offers limited coverage of the features that most influence a building’s energy use – e.g. the height.

3. Machine learning could predict these features at scale.

Infrastructure frames energy demand in cities

Mobility

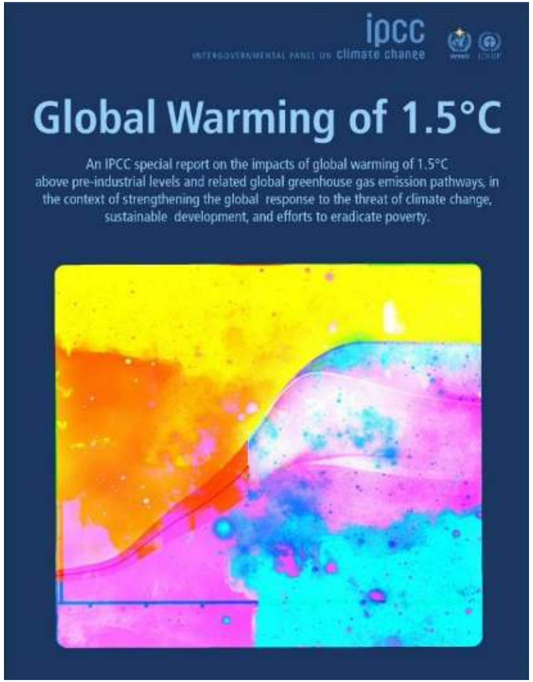
Left: Vikas Taank via Pinterest; Right: http://www.dw.dk/xiamen-bicycle-skyway

Space heating and cooling

Left: Zillow (Southern Living); Right: Elle Interior via Pinterest

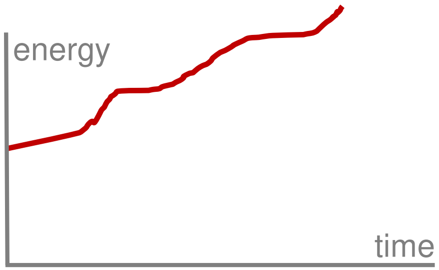
Climate emergency

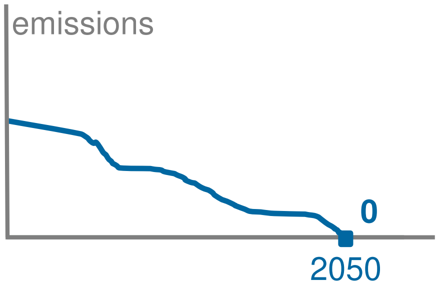
- Buildings and transportation ≈ 60% of our energy demand.
- Without action, energy demand is expected to keep growing.
- We need zero net emissions in 2050 to have good chance to limit global warming to 1.5C.



Climate emergency

- Buildings and transportation ≈ 60% of our energy demand.
- Without action, energy demand is expected to keep growing.
- We need zero net emissions in 2050 to have good chance to limit global warming to 1.5C.





City solutions

- Infrastructure planning can **reframe** energy demand in cities.
- Infrastructure models are the basis for urban planners to design climate mitigation policies.
- Common model framework applicable to different cities would enable comparisons.

Our model

- Gap in literature: building-level energy models at the city scale
- Two main blocks: **3D building model + energy use model**
- Modular framework: simple, increasing complexity as more data gets available

OSM data to develop climate solutions

3D models and height data

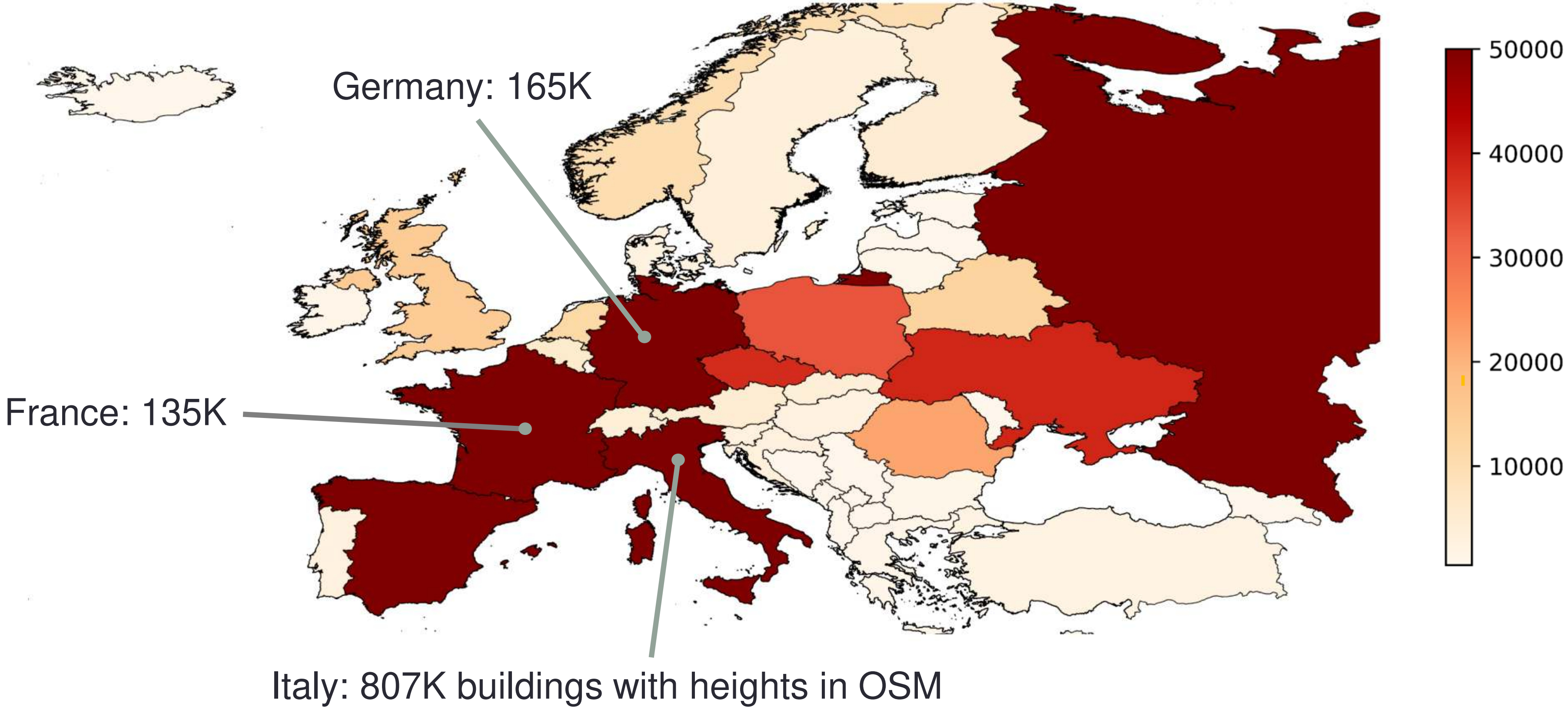
- Level of detail: LOD1 used here

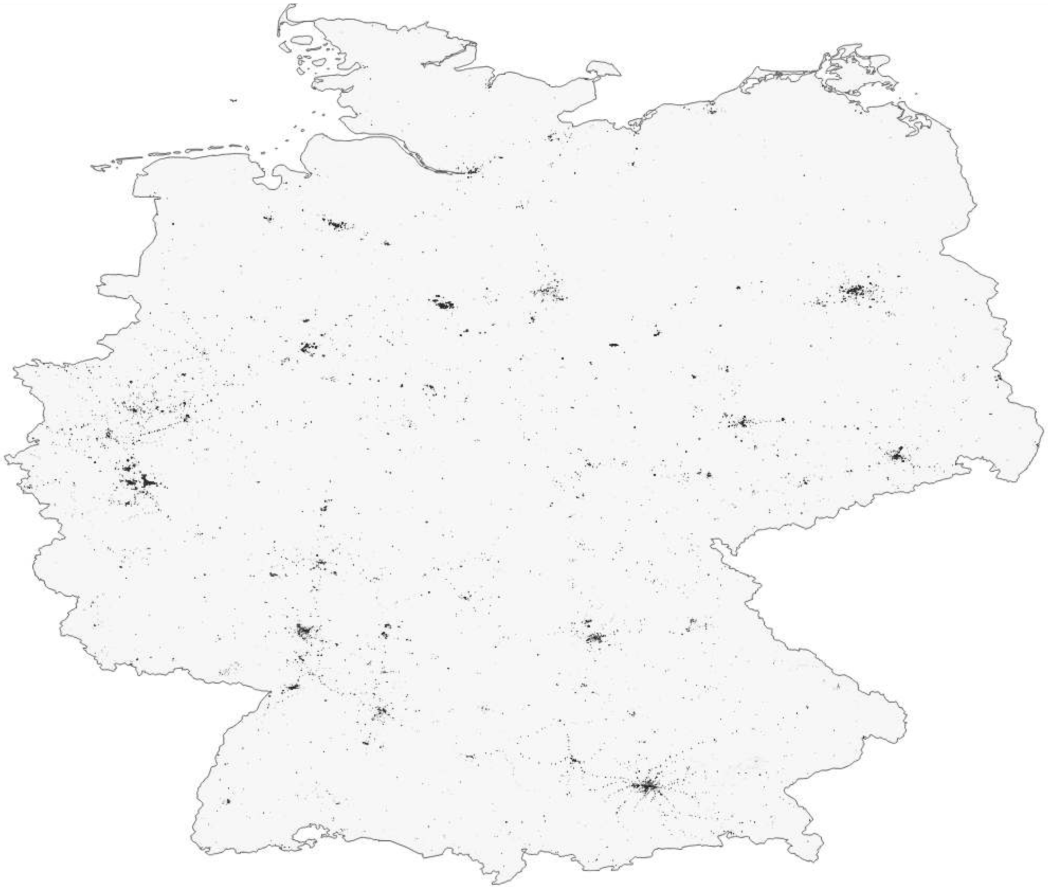
- Sources of height data:
  cadaster, LiDAR, satellite, …

- In OSM, `key:height` ≈ **12M** worldwide ≈ **25** x **Berlin**.

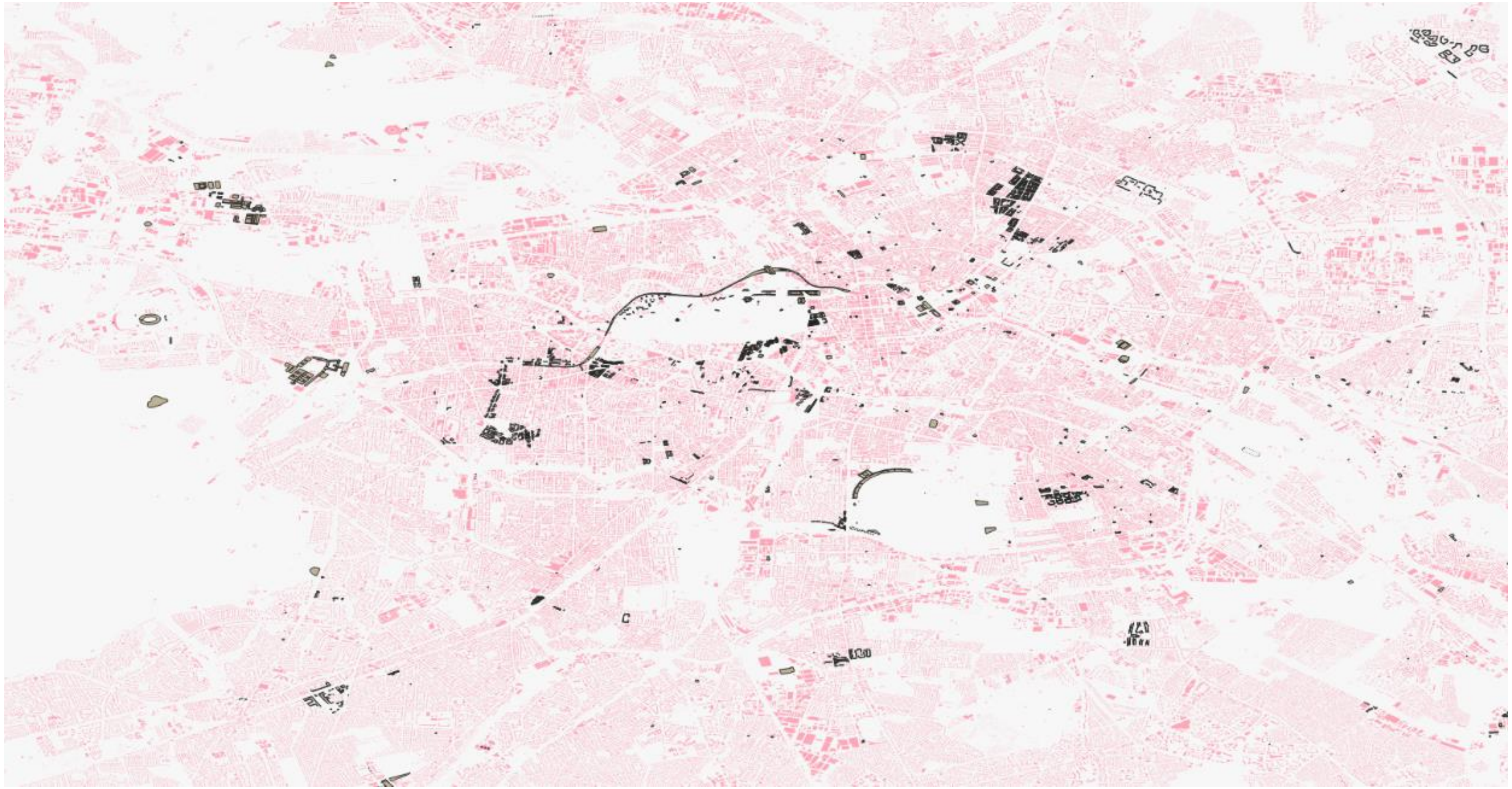
  → Could we predict others?

Europe overall: **1.5M**





Buildings with heights can be fairly well distributed within countries.

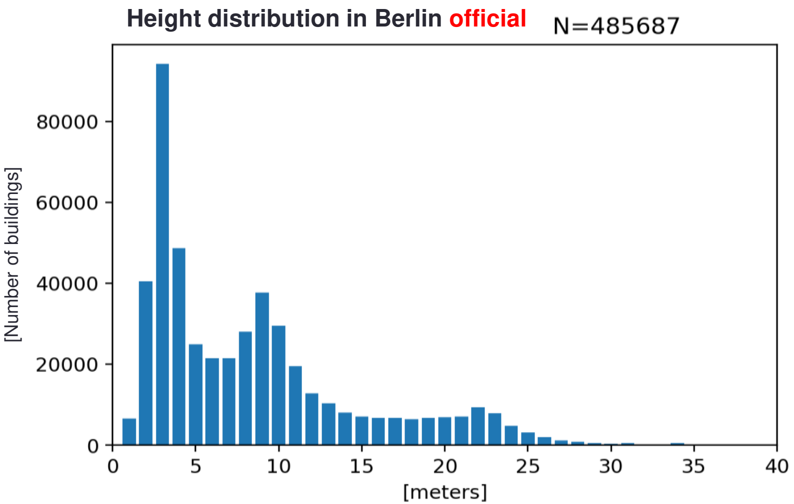


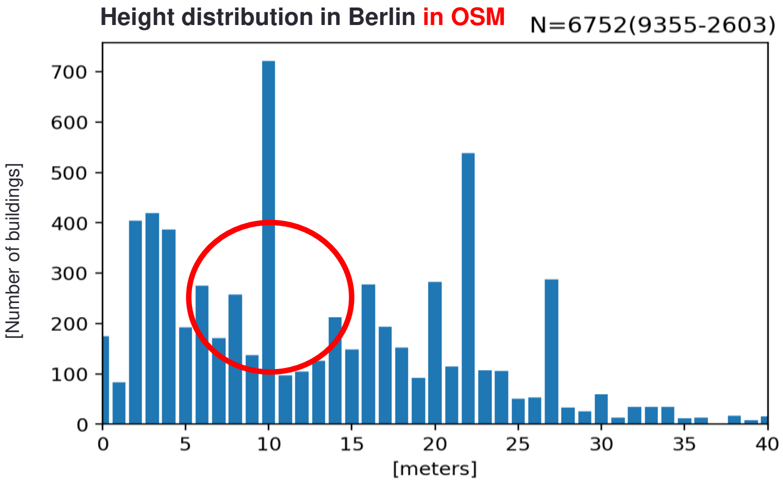
Berlin: ~**1.5**% mapped (6~7K out of ~470K buildings)

Across Europe and
in various contexts,
cities have 100s to 1000s
buildings with heights.

Data management

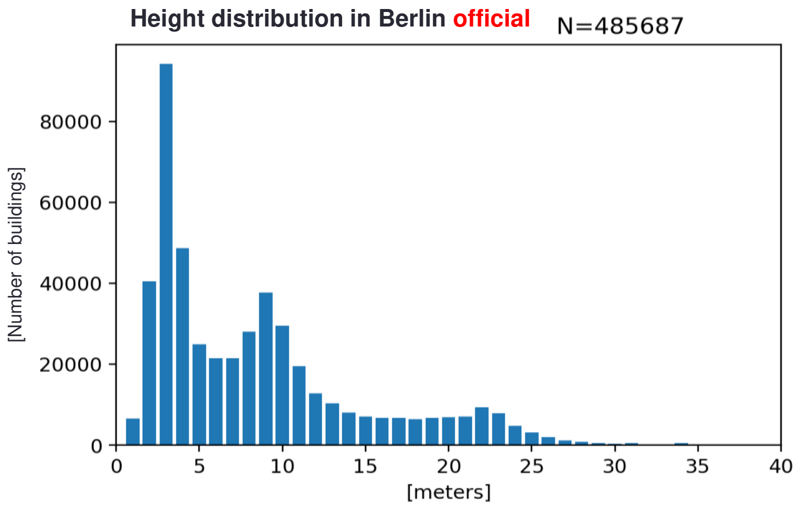
- Dealing with selection bias
- Merging building parts and removing non-building artefacts
- Wrong entries, level of precision e.g. [5m,10m,15m,…] vs continuous

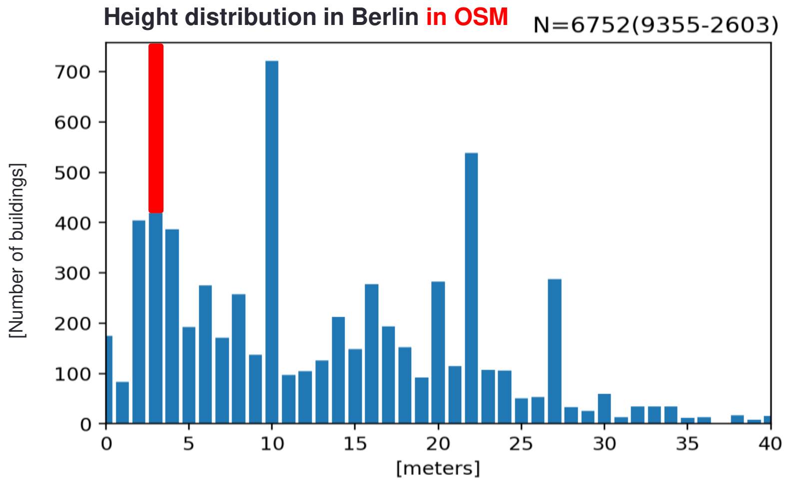




Data management

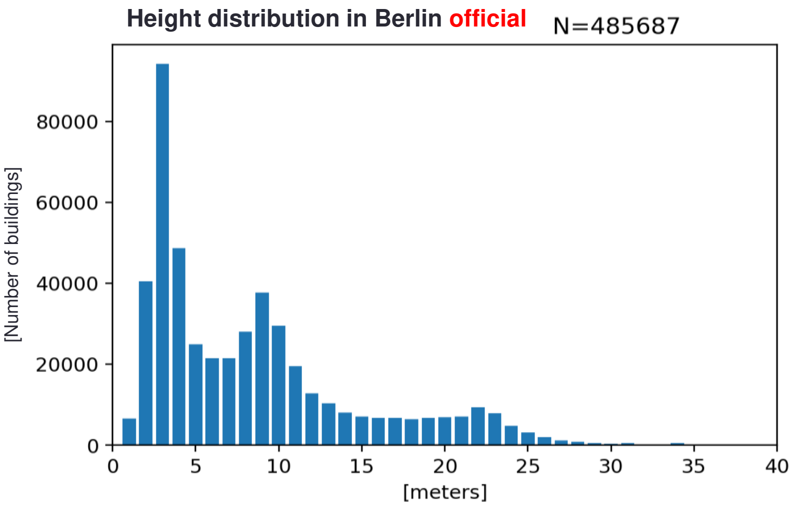
- Dealing with selection bias
- Merging building parts and removing non-building artefacts
- Wrong entries, level of precision e.g. [5m,10m,15m,…] vs continuous

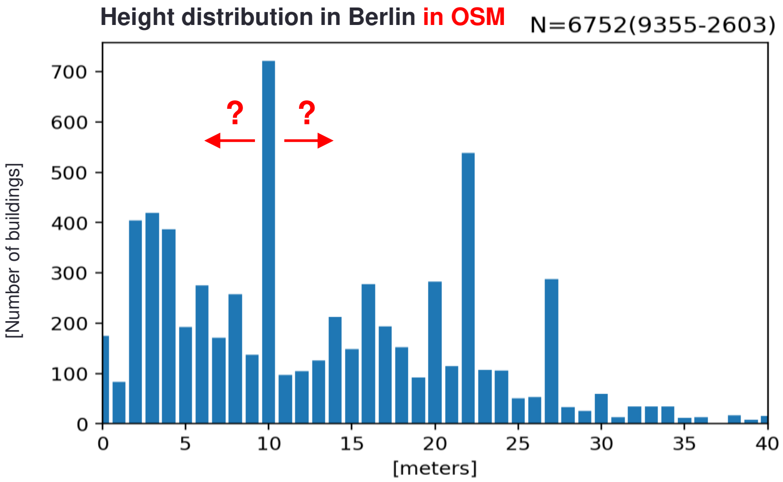




Data management

- Dealing with selection bias
- Merging building parts and removing non-building artefacts
- Wrong entries, level of precision e.g. [5m,10m,15m,…] vs continuous





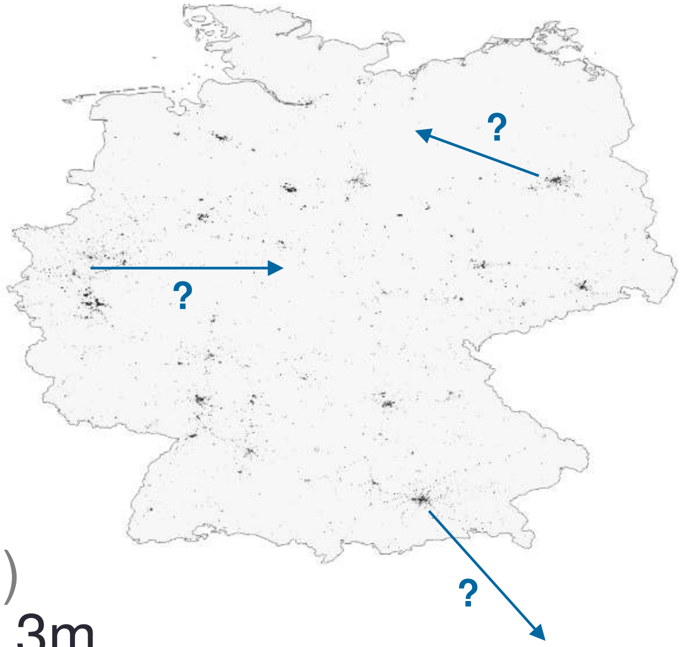
Upscaling building heights in OSM

Height prediction

Goal: Predicting building heights across Europe, using **only** OSM features, and training on **all** available height data

Similar work:

- Whole USA by 'Open City Model' but no accuracy reported
- Two Dutch cities by Biljecki et al. (2017) with MAE = 0.8~3.1m; RMSE = 1.8~4.3m



Features that can be computed for any building in OpenStreetMap

Building

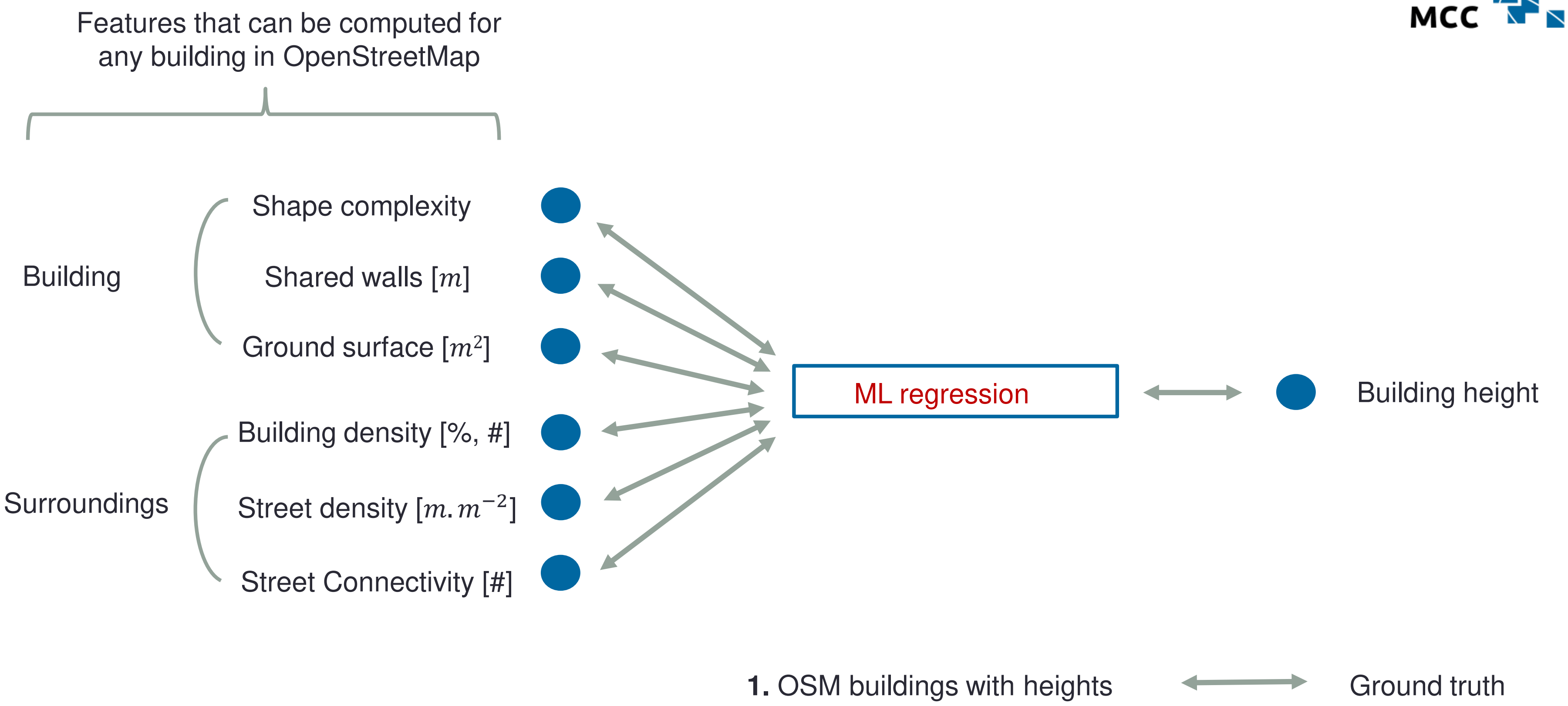
- Shape complexity
- Shared walls $[m]$
- Ground surface $[m^2]$

Surroundings

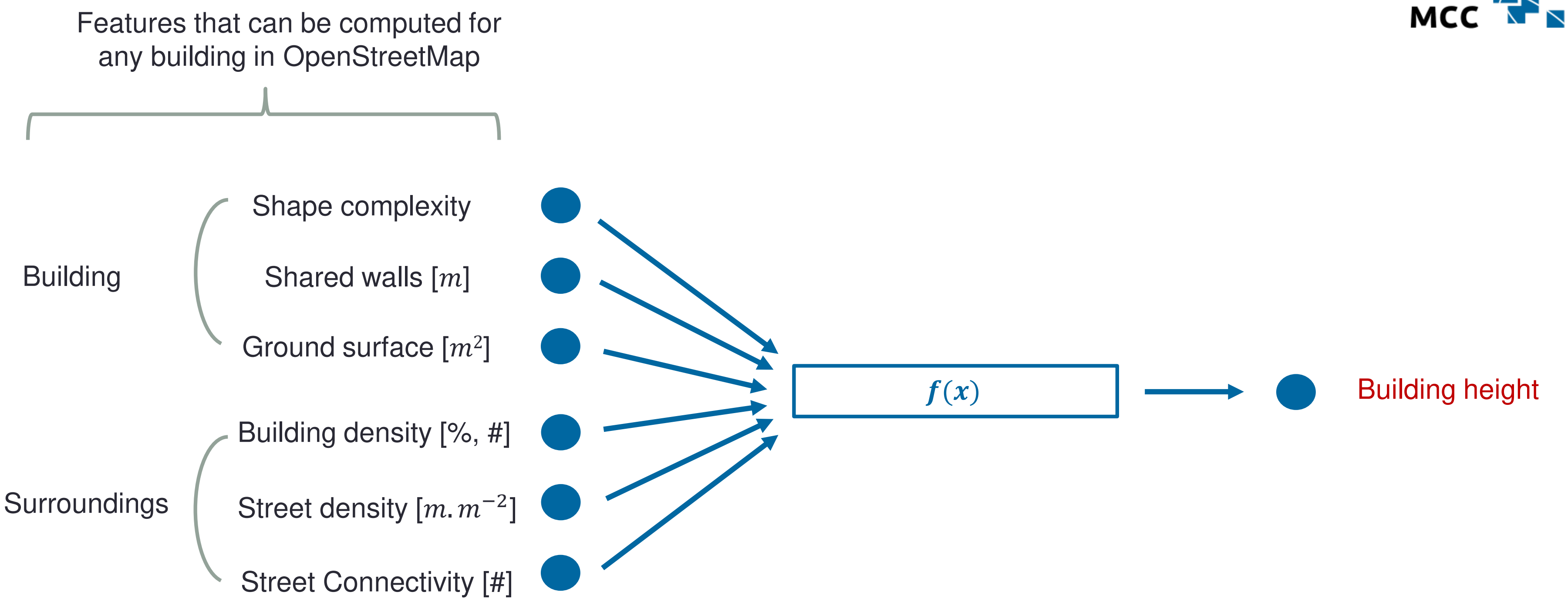
- Building density [%, #]
- Street density $[m.m^{-2}]$
- Street Connectivity [#]

ML regression

Building height



Features that can be computed for any building in OpenStreetMap
Building
Shape complexity
Shared walls $[m]$
Ground surface $[m^2]$
Surroundings
Building density [%, #]
Street density $[m.m^{-2}]$
Street Connectivity [#]
ML regression
Building height
1. OSM buildings with heights
Ground truth



1. OSM buildings with heights $\longleftrightarrow$ Ground truth

2. OSM buildings without heights $\longrightarrow$ Predicted heights

Example

- Random forest,

within and across cities

- Hradec Kralove

$n = 29\text{K}$, $st = 4.6\text{m}$

- Udine

$n = 23\text{K}$, $st = 4.3\text{m}$

Example

- Random forest,

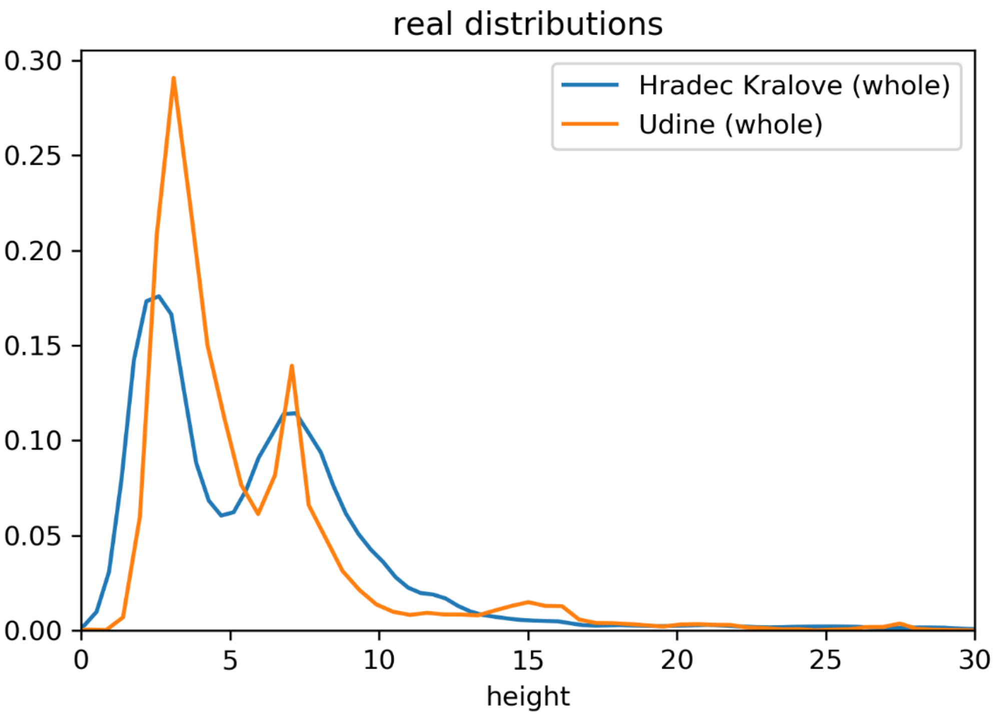
within and across cities

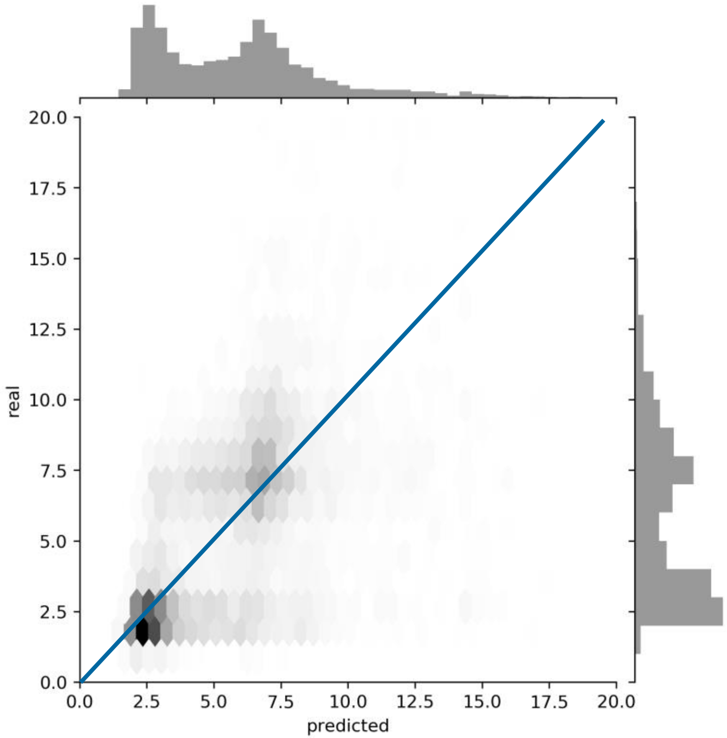
- Hradec Kralove

$n = 29\text{K}$, $st = 4.6\text{m}$

- Udine

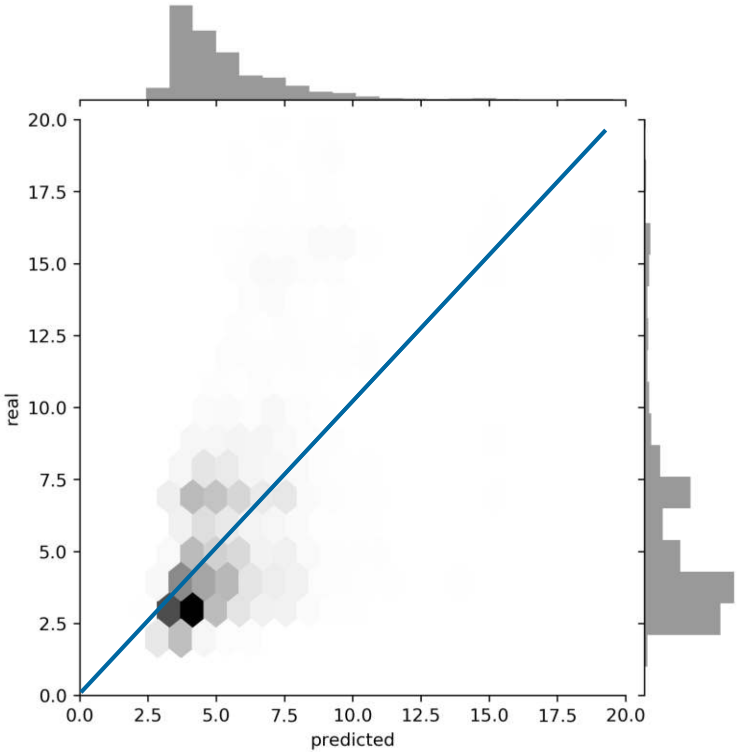
$n = 23\text{K}$, $st = 4.3\text{m}$





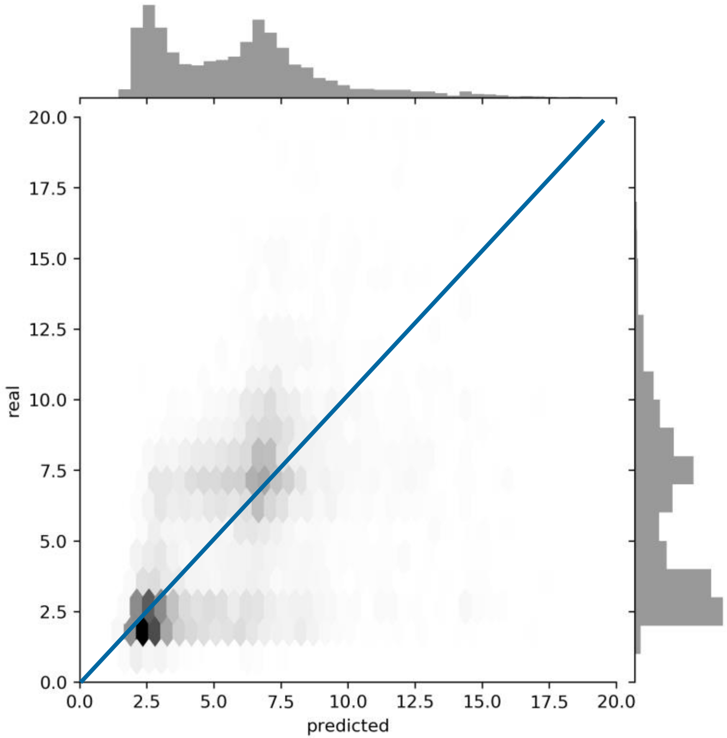
Hradec Kralove

$MAE = 2.46$, $RMSE = 3.79$

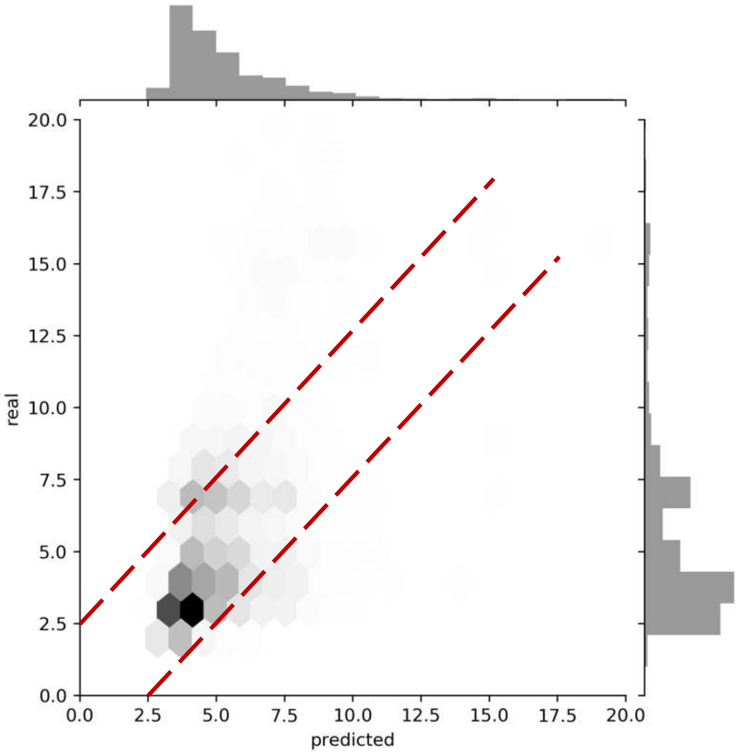


Udine

$MAE = 2.25$, $RMSE = 3.79$



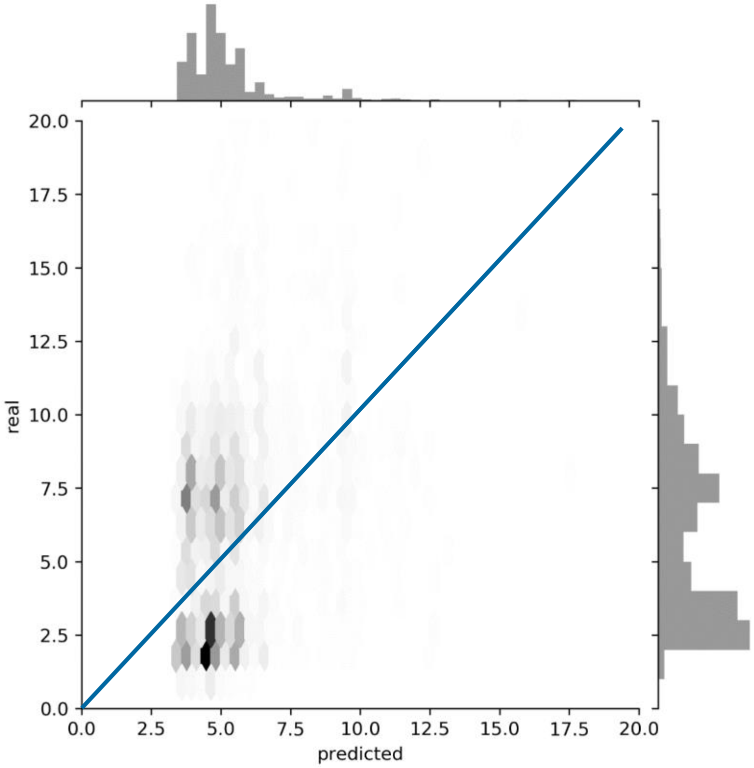
Hradec Kralove



Udine

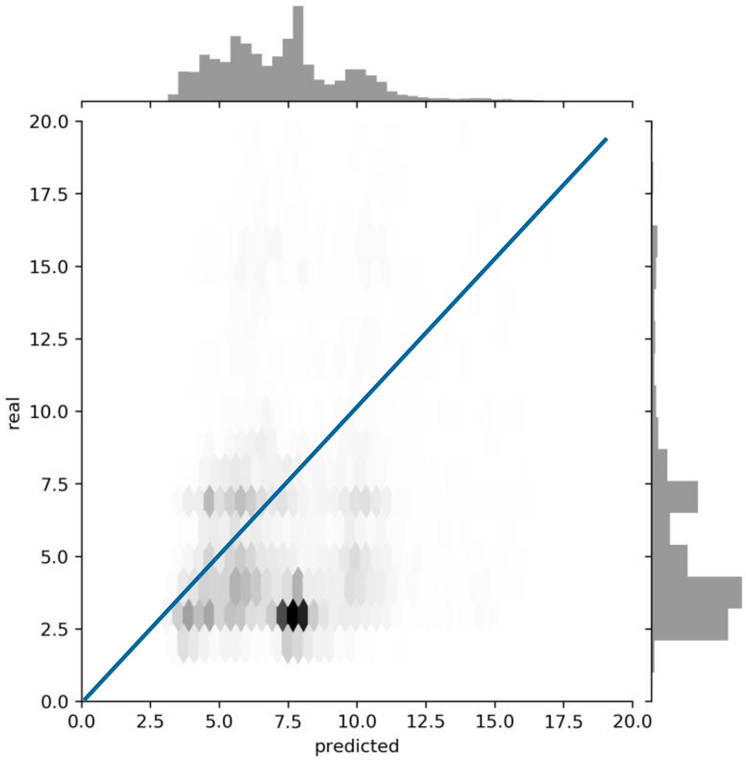
$MAE = 2.46$, $RMSE = 3.79$

$MAE = 2.25$, $RMSE = 3.79$



Hradec Kralove (model Udine)

$MAE = 2.89$, $RMSE = 4.0$



Udine (model Hradec Kralove)

$MAE = 3.64$, $RMSE = 4.99$

What matters for the prediction?

- Generalization in spatial models Meyer et al. (2019)
- Sampling: urban *vs.* rural, region-specific *vs.* whole Europe
- Algorithm, model and feature selection

Implications for the OSM community

- Every building mapped counts.

- Buildings in low-data settings are most important, especially in the Global South.

- ‘Number of floors’ are easier to map and obvious predictors (currently also ~12M worldwide).

1. Infrastructure data in OSM are of high relevance for the sustainability science community.

2. Currently, OSM offers limited coverage of the features that most influence a building's energy use – e.g. the height.

3. Machine learning could predict these features at scale.

Nikola Milojevic-Dupont milojevic@mcc-berlin.net @Nikola_MD