Application of machine learning methods for embodied carbon estimation

Aaron Qiyu Liu | Chalmers University of Technology | 19/09/2023

Background

- Embodied carbon from the construction of buildings and transport infrastructures corresponds to 13% of global CO2 emissions
- Embodied carbon corresponds to lifecycle stages A1-5
- Detailed bottom-up models are needed to support policy making

Source: Skanska USA

How to estimate embodied carbon?

Step 1: Quantify material stock

$$Stock = Inventory \ast Material\ intensity$$

Step 2: Material flow analysis

$$MFA(Stock): Inflows, Outflows$$

Step 3: Estimate embodied carbon

$$EC = Inflows \ast Emission\ factor$$

Inflow

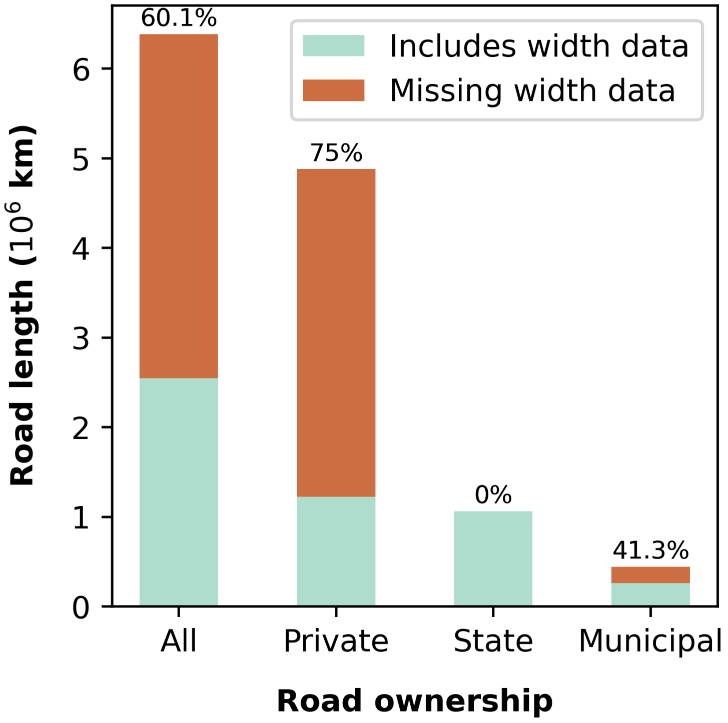
Outflow

What is the challenge?

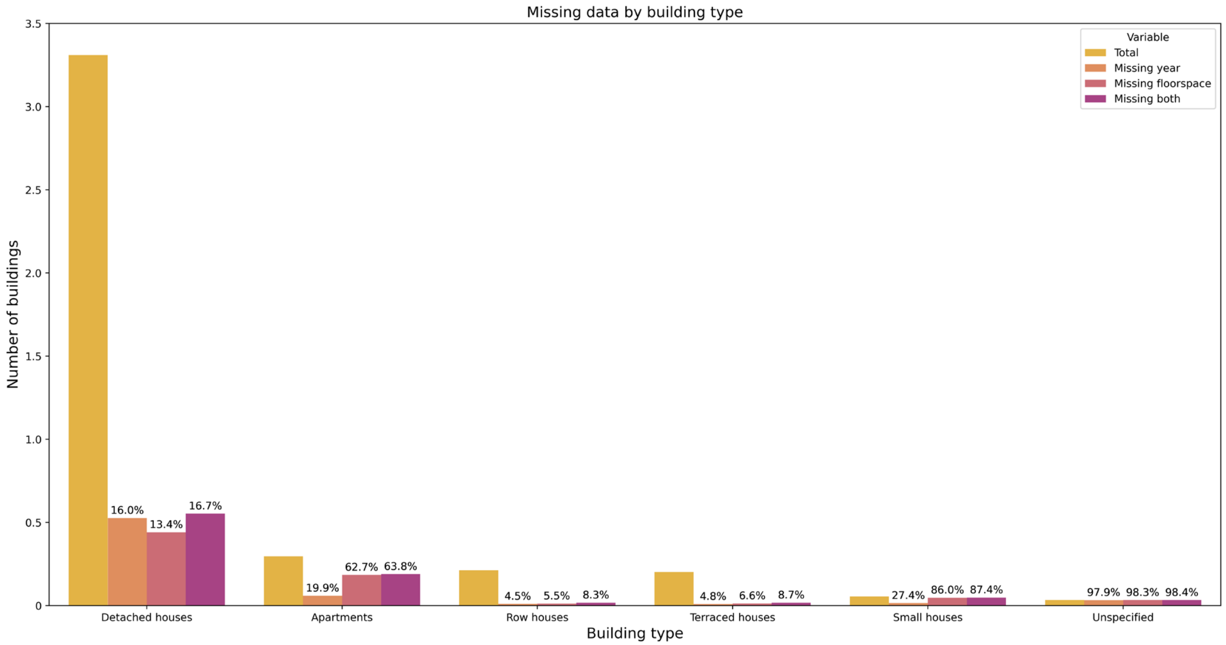
- The quantification of stock requires data on each building/road in the analyzed area
- Statistical data are often incomplete
- To conduct national level analysis, methods to impute or predict missing data is required

Missing data - roads

- Includes all asphalt paved roads and gravel roads, cycleways are excluded from the analysis
- Data from NVDB
- Width data is important for material stock estimation
- Private and municipally owned roads have large proportion of missing data



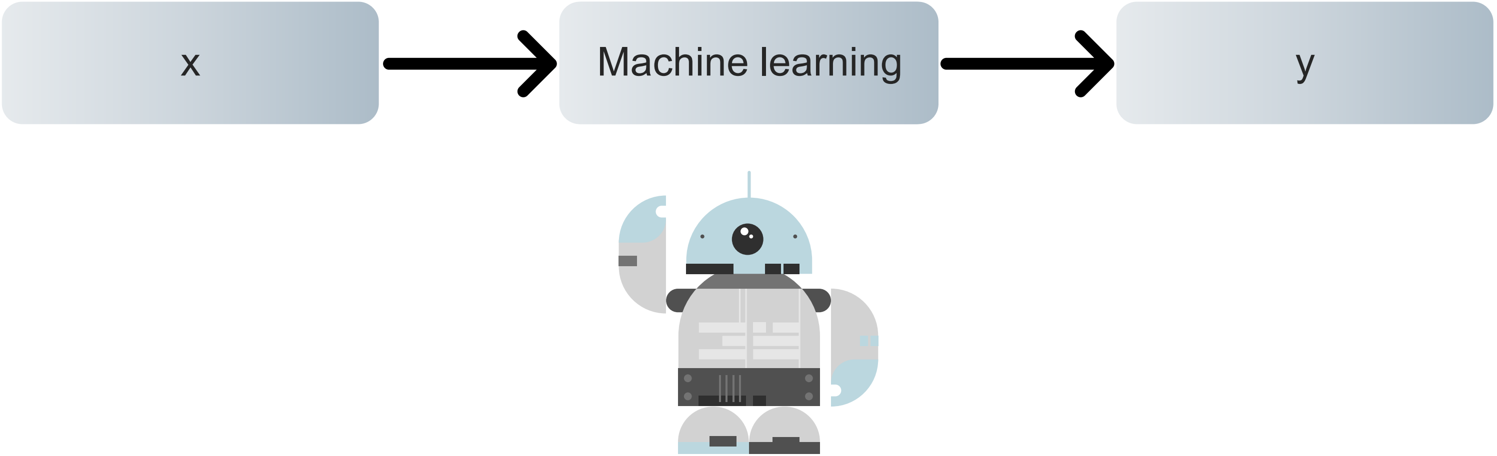
Missing data – residential buildings



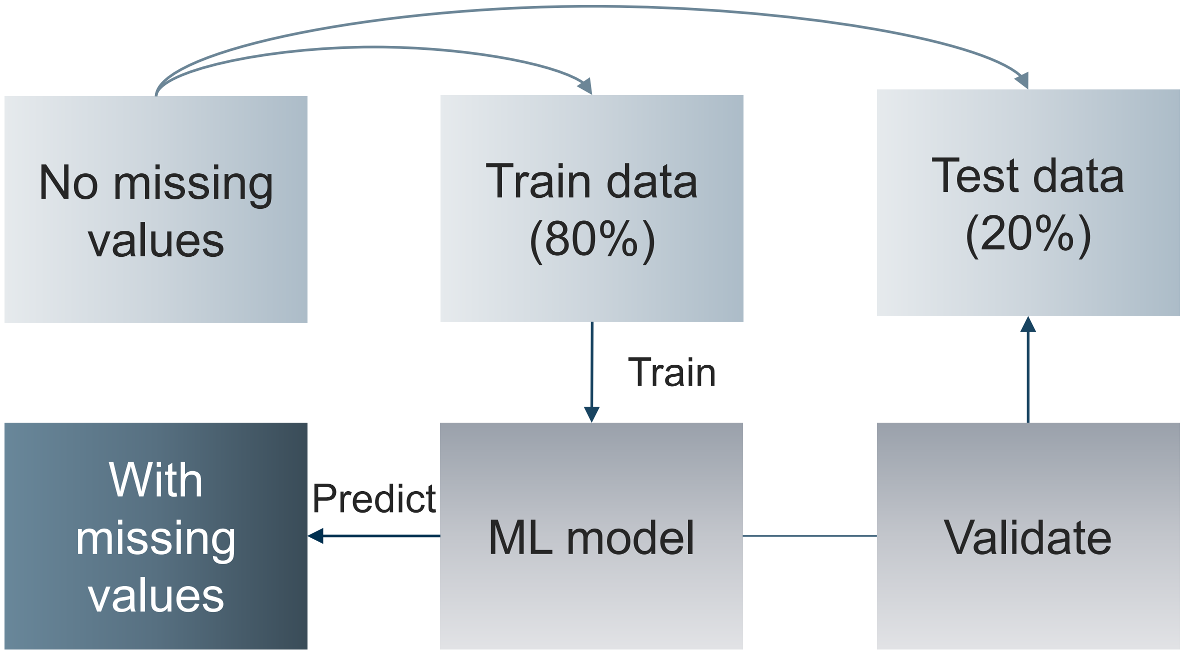
How to deal with missing data?

1. **Drop all missing data**:
   - Makes the analysis incomplete
2. **Imputing with mean, median, or nearest neighbors**:
   - Could be worse than just random guess
   - Data is very heterogenous
3. **Randomly sample from distribution**:
   - Basically, random guesses, can be seen as baseline
4. **Machine learning!**:
   - Good at tasks like this, but it is a ‘black box’

What is machine learning?



How does (supervised) ML work?



How is ML applied?

Roads

- Predicting missing road widths with regression models

Residential buildings

- Predicting missing building age with classification models
- Predicting missing building floor space with regression models and predicted age as a feature

Machine learning results

Road results

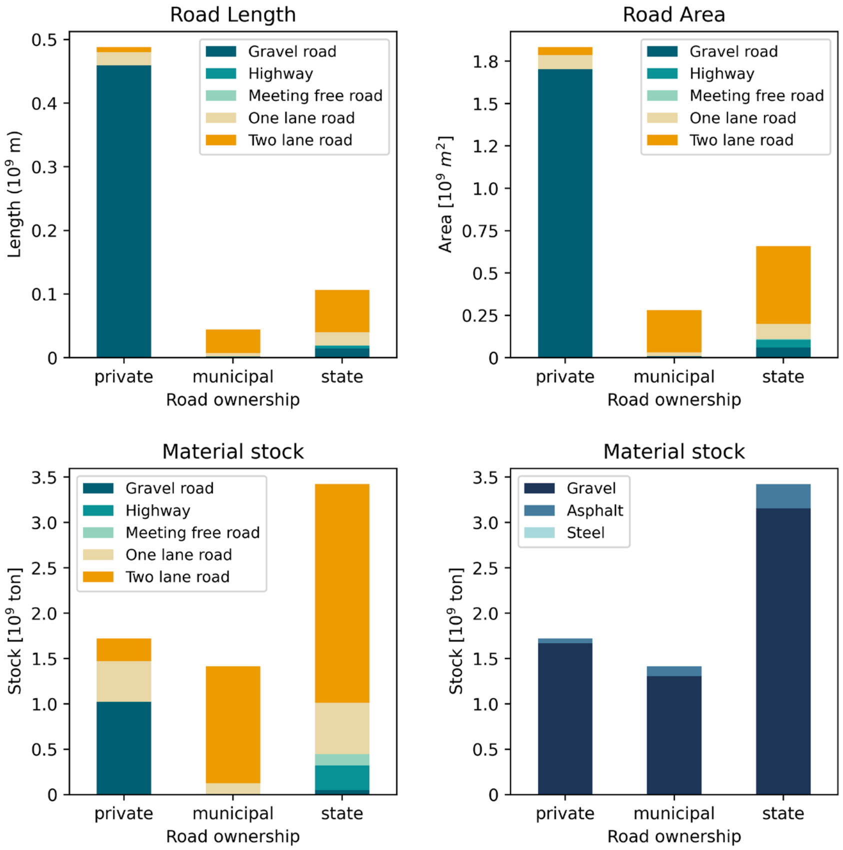
Target	Type of model	R^2
Road width	Regression	0.78

Residential building results

Target	Type of model	Evaluation metrics
Building age	Classification	Accuracy: 89%
Building floor space	Regression	R^2: 0.74

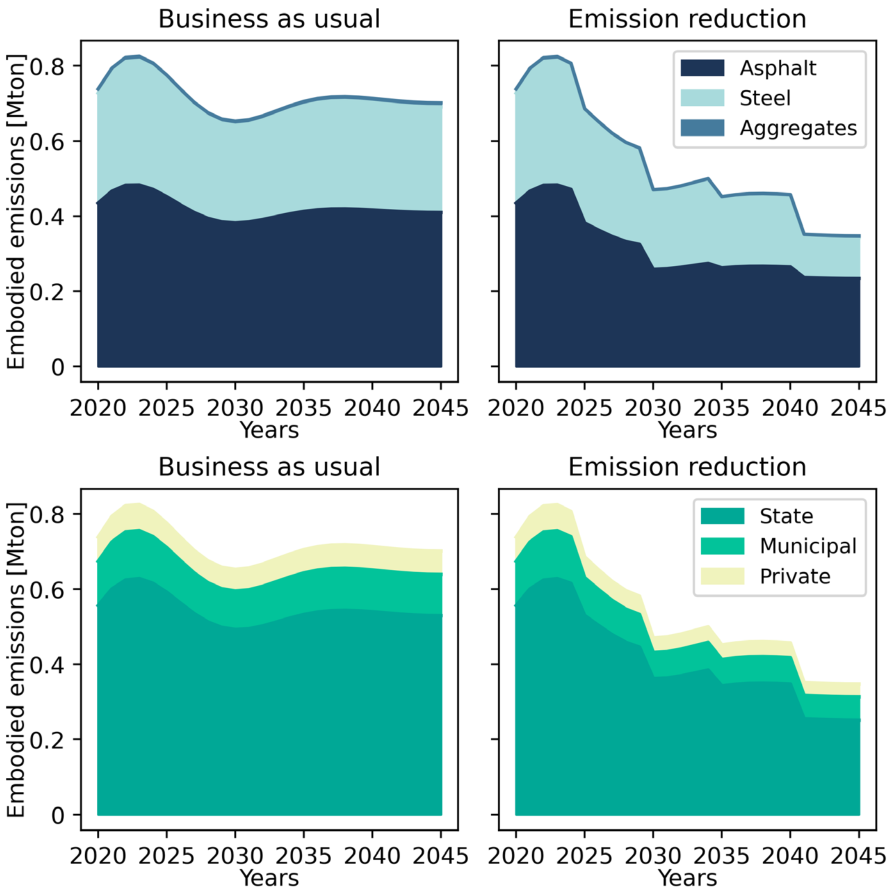
Road stock results

- Private roads have the longest absolute length
- But it mostly consists of gravel roads
- State-owned roads contains the most in-use material stock



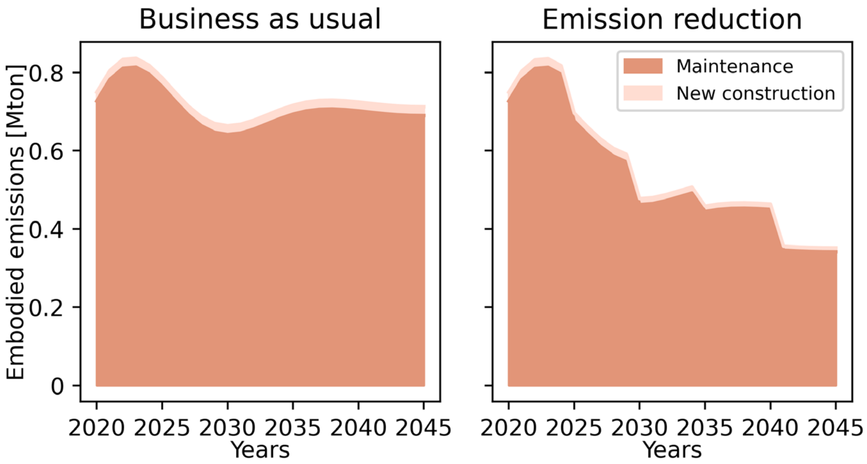
Embodied emission results

- Despite the lower absolute mass, steel still contribute to a significant share of embodied carbon due to its shorter lifetime and higher emission factor
- Municipalities and private road owners still have a role to play



Emissions from new construction

- New construction of roads contributes a very low percentage of the overall yearly embodied carbon
- To decarbonize Swedish roads, supply side technological innovations are crucial



Summary

- Machine learning methods perform relatively well for predicting missing data
  - The result is a hybrid of real and synthetic data
  - Needs to be tested on a case-by-case basis
- Steel contributes a significant percentage of embodied carbon
  - Mainly because the emission factor for steel is magnitudes higher than asphalt
- Reducing or stopping new construction of roads will have limited effect
  - Decarbonizing the material supply is more important
  - More attention could be directed at improving maintenance routines to prolong lifetime
  - Results might change if bridges and tunnels are included due to the use of concrete

Future work

- Quantitatively forecasting future new housing construction through ML-based methods
- Scenario-based embodied carbon analysis of residential buildings
- Non-residential buildings embodied carbon

AVANCEZ

1829

CHALMERS

UNIVERSITY OF TECHNOLOGY