Global air pollution modelling

University of Utrecht, The Netherlands
Global and geo-health data center

Meng Lu

Overview

- Introduction
  - Spatio-temporal epidemiology
  - Air pollution modelling and exposure assessment for health research

- Global air pollution modelling

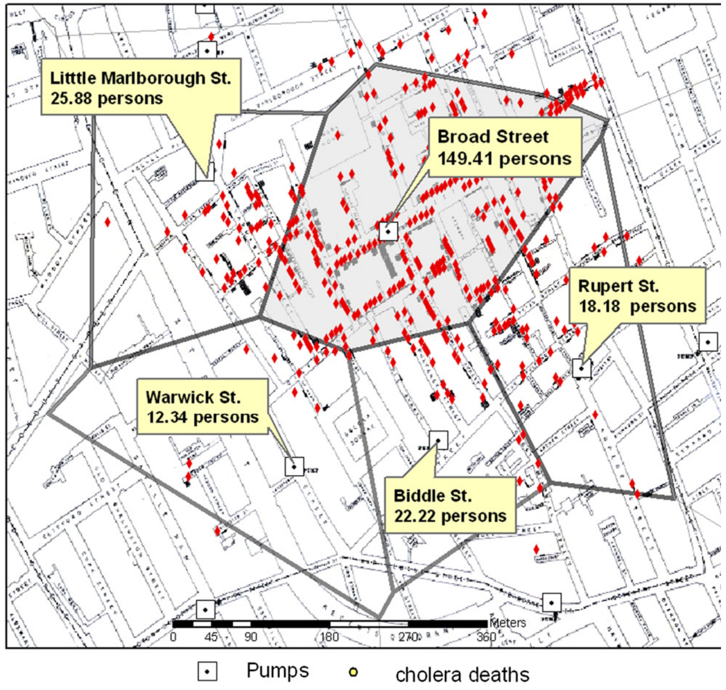
Spatio-temporal Epidemiology

Spatiotemporal epidemiology: The description and analysis of geographical data, specifically health outcome data and factors that may explain variations in these outcome data over space. factors: environmental, demographic, genetic, habits, infectious risk factors.

Environmental epidemiology: Spatiotemporal epidemiology that focuses on how environmental exposures impact human health.

Origin
1854 John Snow,
Identify possible causes of cholera outbreaks.

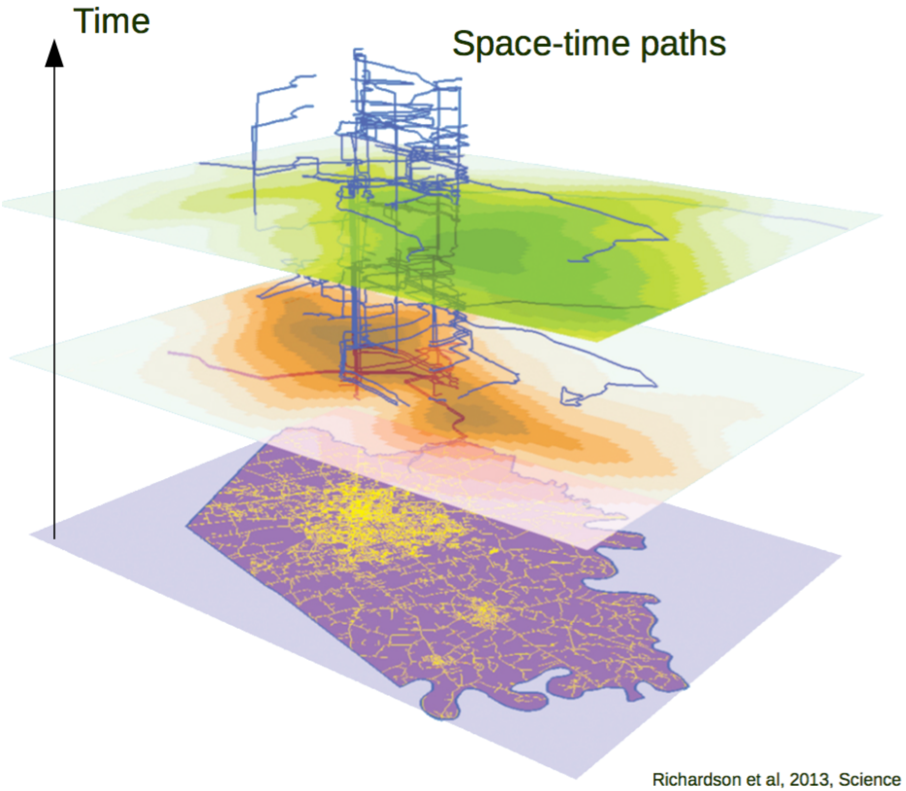
Cholera Mortality per 1,000 persons



Cholera Mortaliy per 1,00 persons for central pump catchments.

Exposome

At its most complete, it encompasses life-course environmental exposures



Richardson et al, 2013, Science

Exposure: aggregating the environmental variables along a person's space-time path

NO_2: traffic related

Air pollution mapping methods

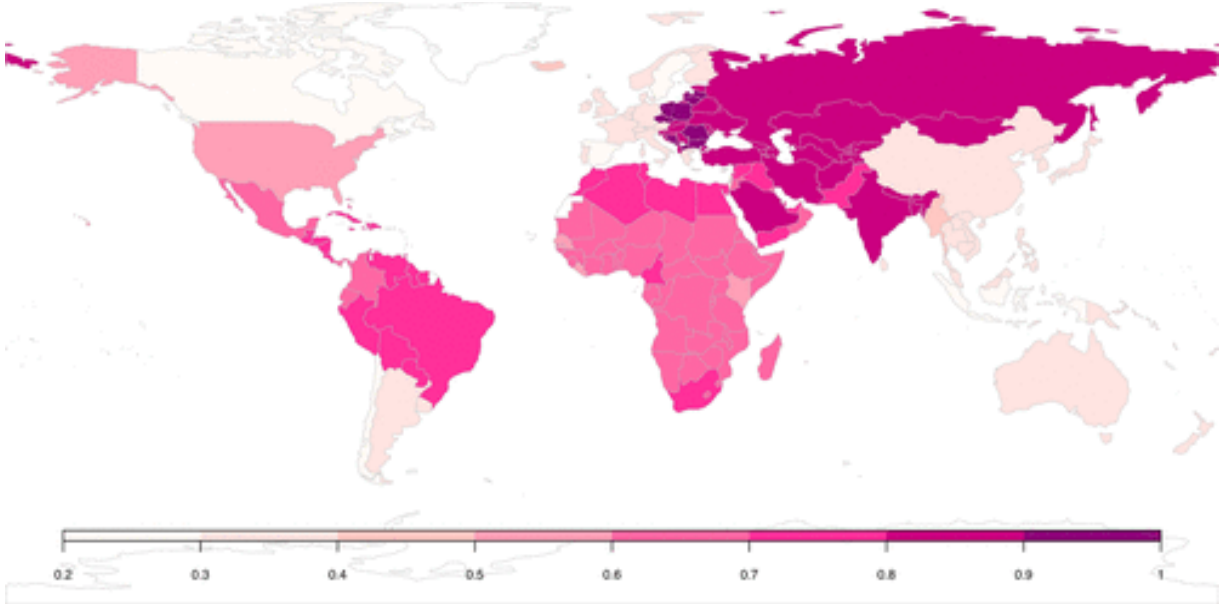
- Chemical transportation models: GEOS-CHEM
- Statistical methods: land use regression, geostatistics

Why is global air pollution mapping and exposure assessment important?

- **Consistent comparison**
- Unequally distributed ground monitors

Station measurements

Shaddick et al., 2018

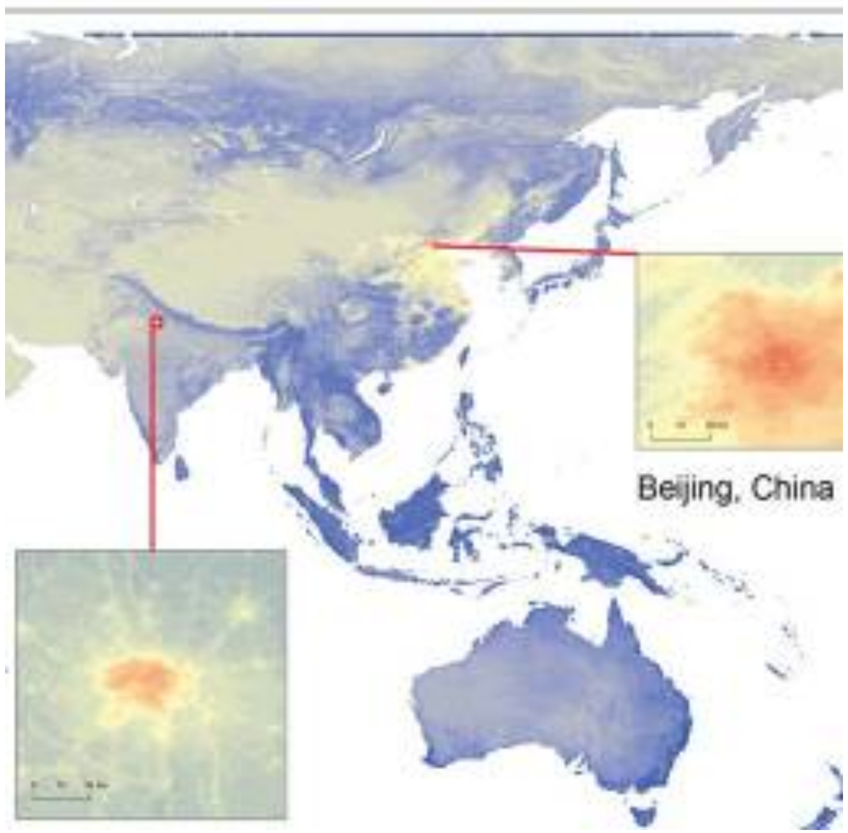


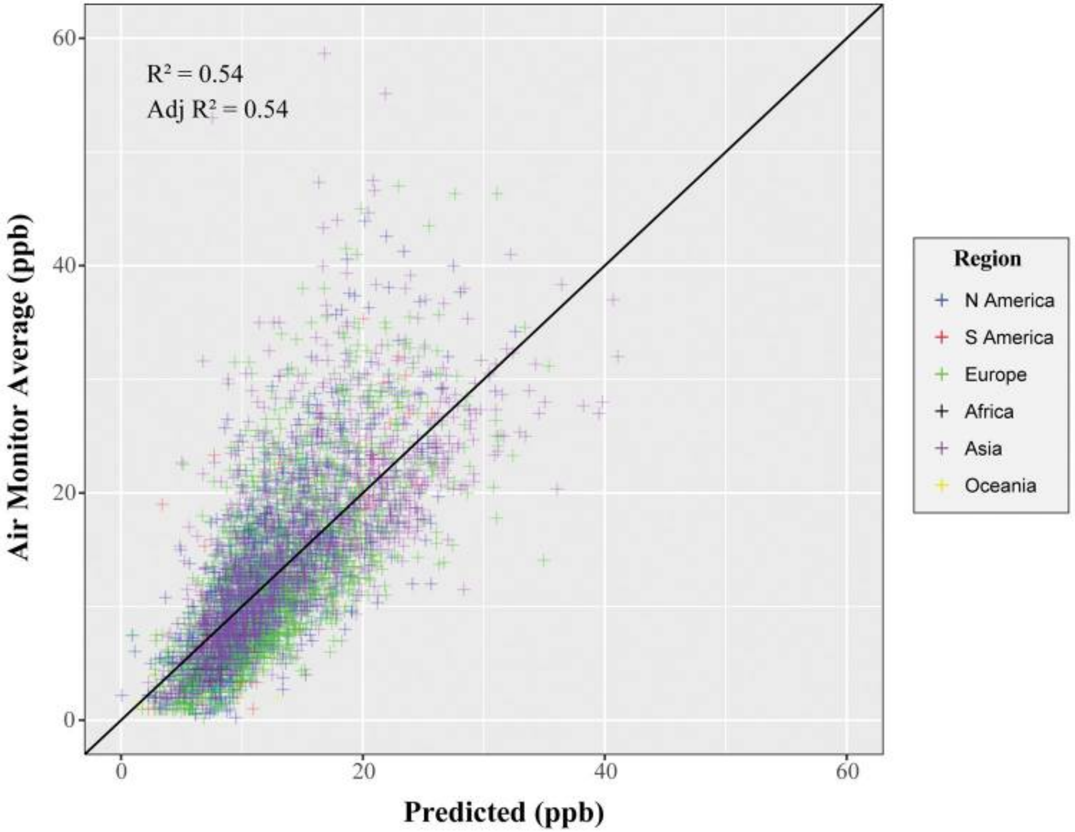
Global NO_2 mapping: Larken et al. 2017 (100m):

Linear regression model
Sensitivity analysis shows the differences between continents

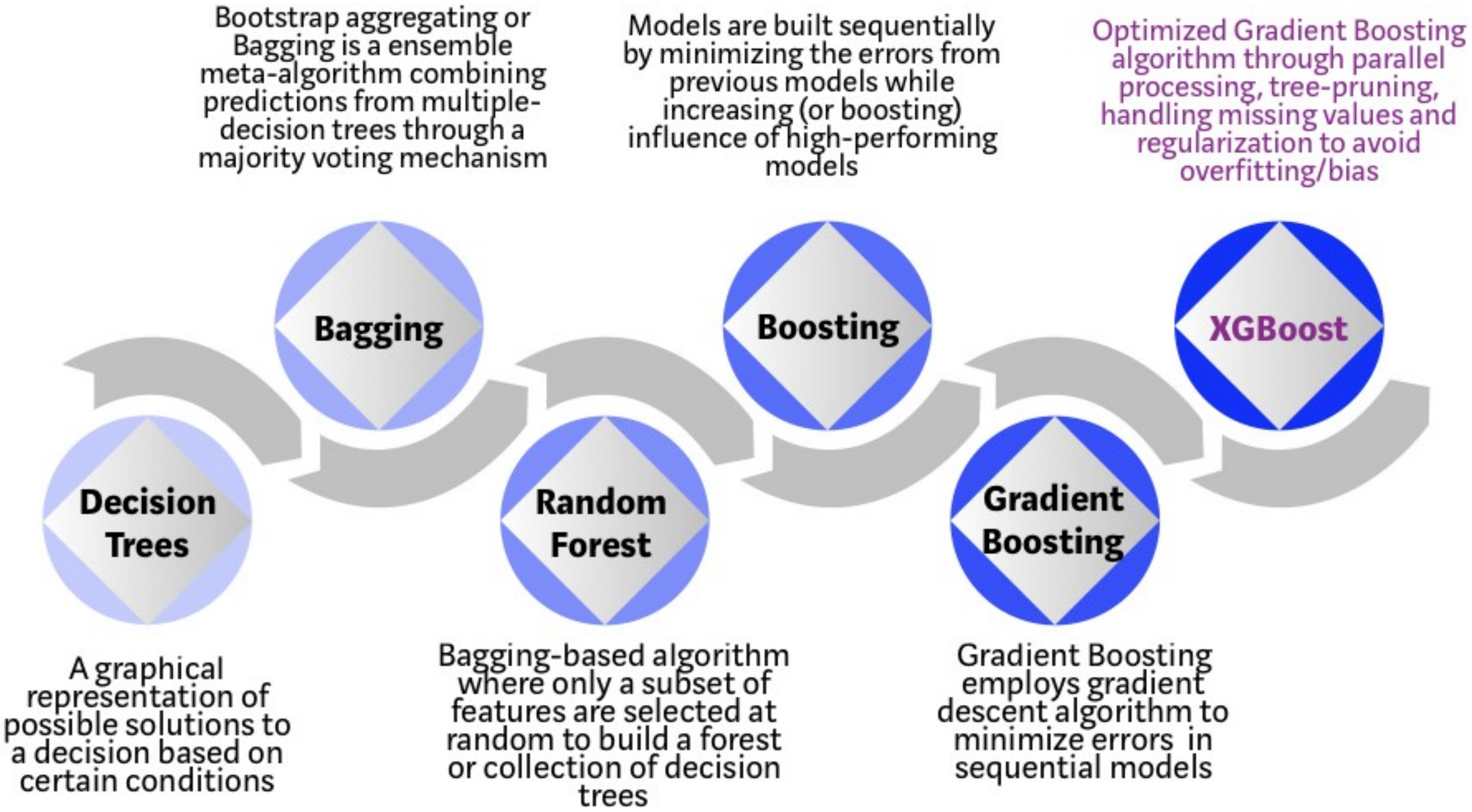
Limitations:

- Linear relationship



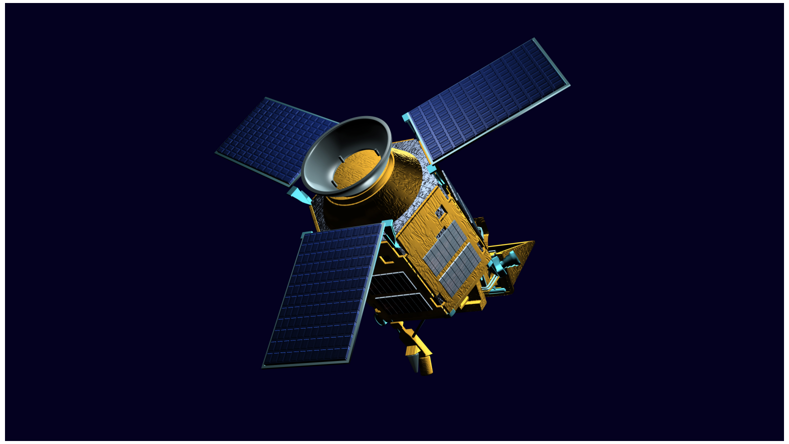


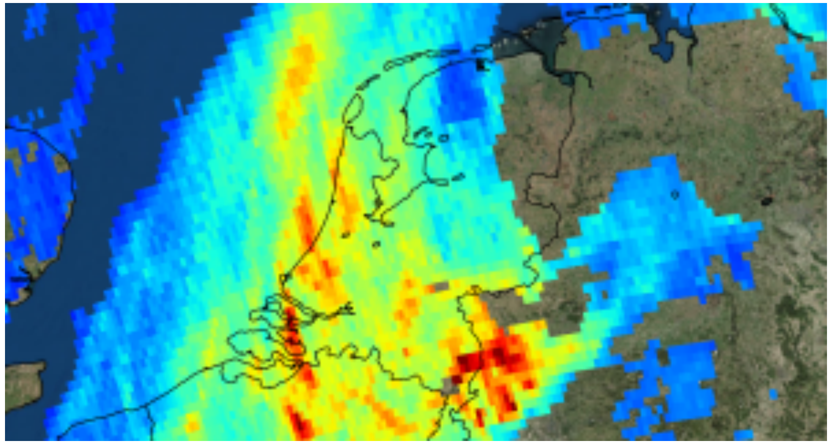
- Can regression tree-based statistical learning methods improve global NO_2 mapping ?
- How do national and global models differ?



Tropomi (TROPOspheric Monitoring Instrument)

The gaseous satellite instrument on board the Sentinel-5 Precursor satellite.
launched 2017, available from Feb 2018

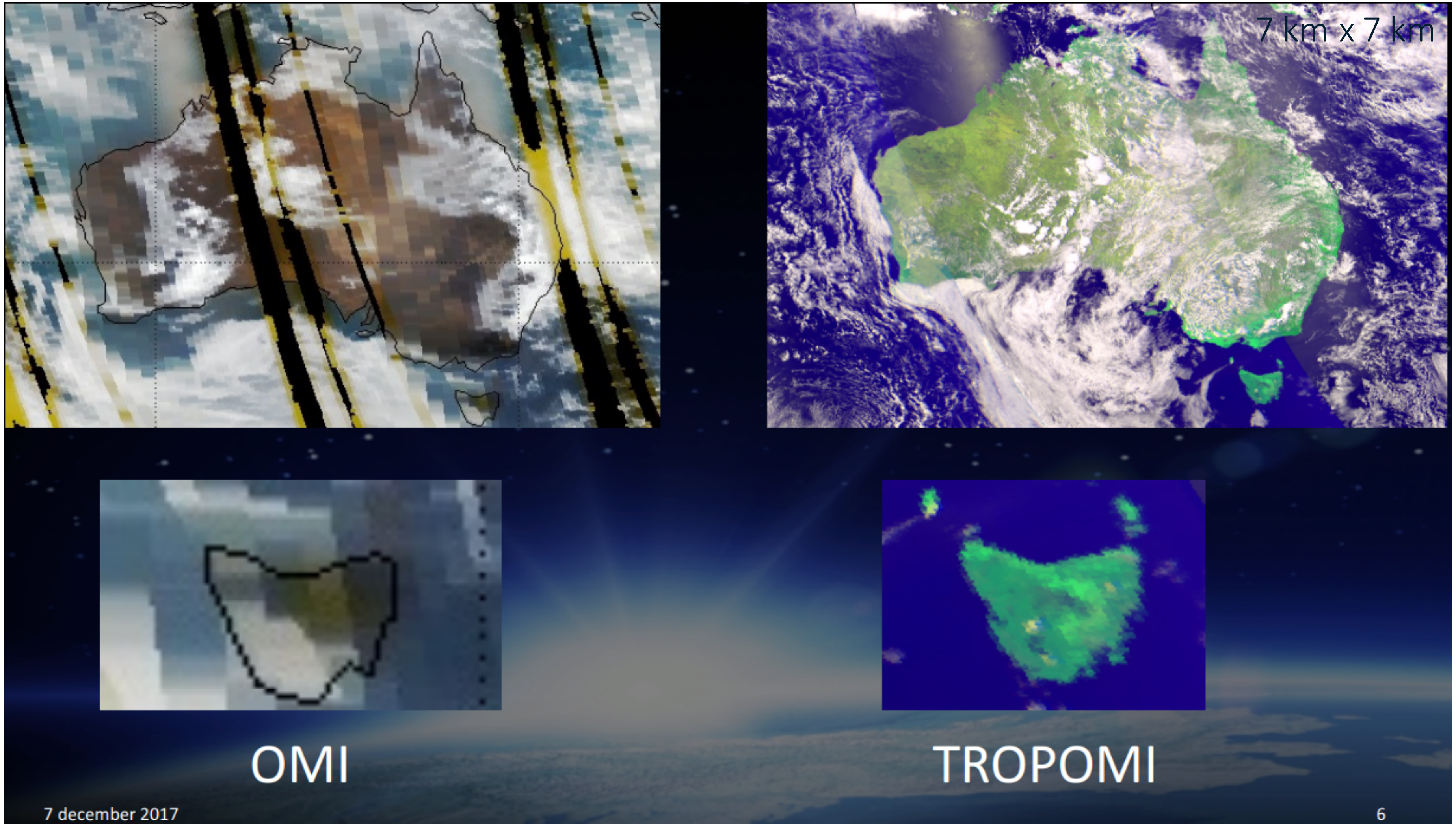




NO_2 (7km x 7 km), O3 (7km × 28km), SO2, methane and CO

Spectral bands:
ultraviolet and visible (270–500 nm), near-infrared (675–775 nm), shortwave infrared (2305–2385 nm) spectral bands.

Tropomi



13 km x 24 km

Souce [3]

Method

Comparing different statistical learning methods

Tree-based

- Random forest
- Stochastic gradient boosting
- Extreme gradient boosting (xgboost)

Regularized regression

- Ridge
- Lasso
- ElasticNet

Land use regression

Compare global and national models

Four national models:

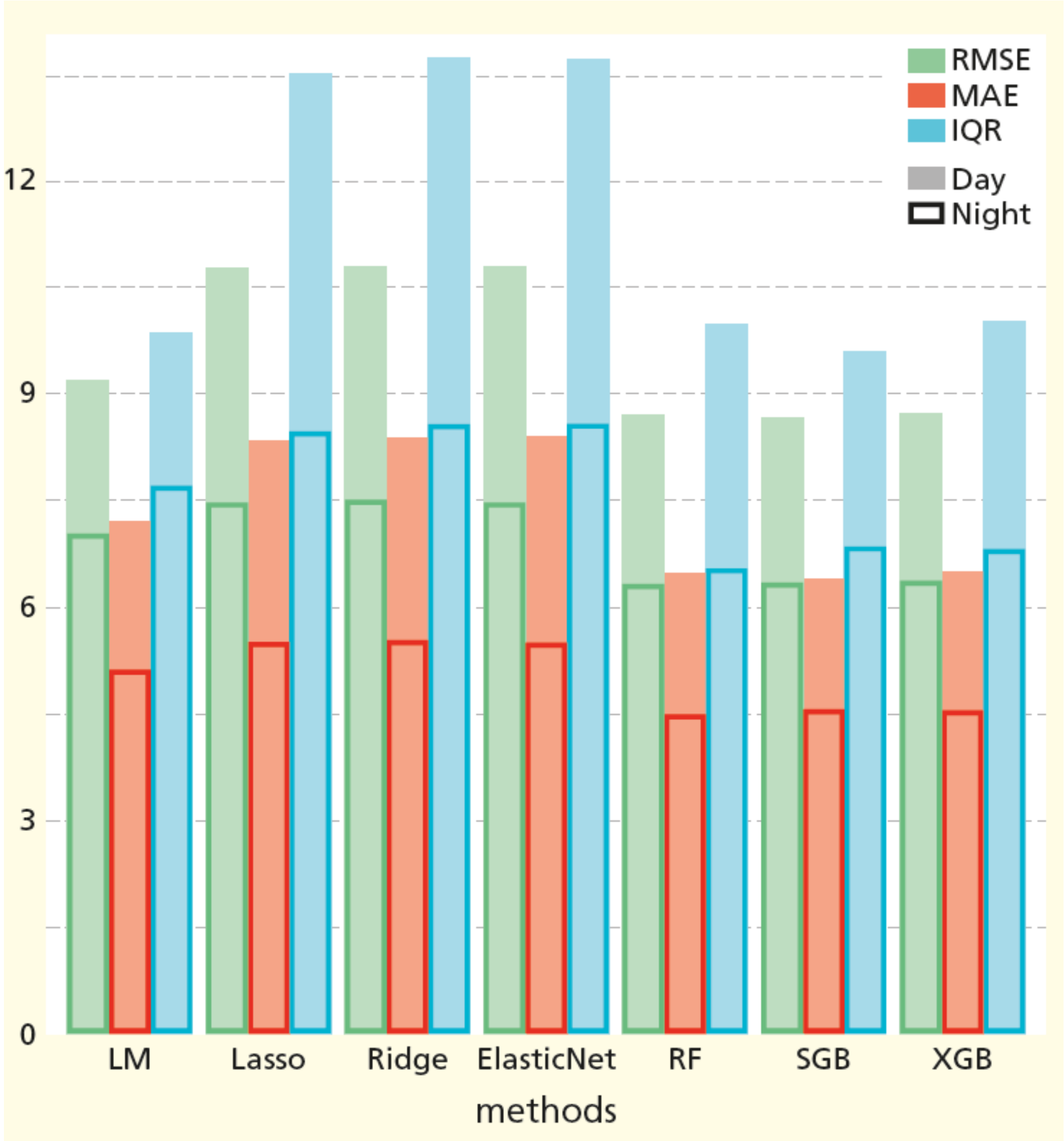
US (100),
China (1296),
Germany (338),
Spain (380)

A global model

(3636)

Contribution of Tropomi
Separating between day and night

Result: global model accuracy

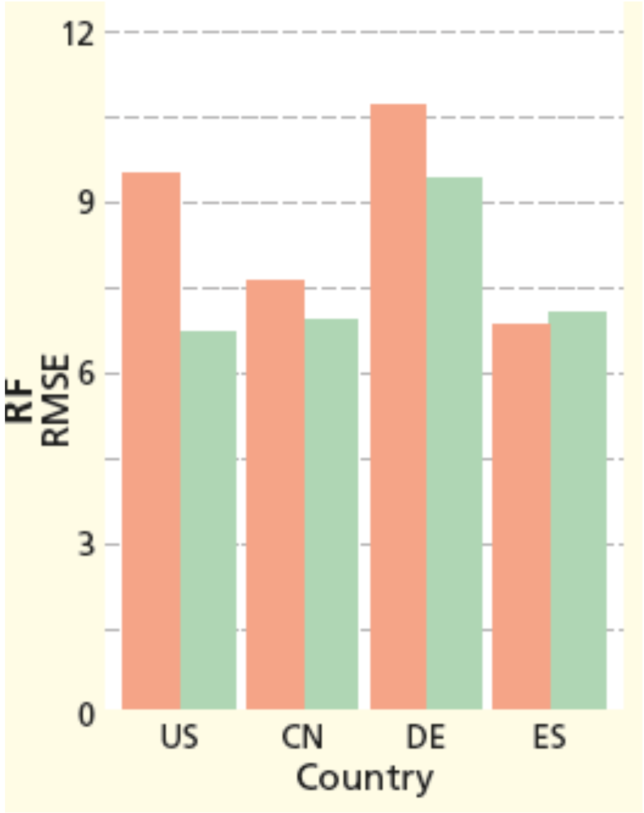


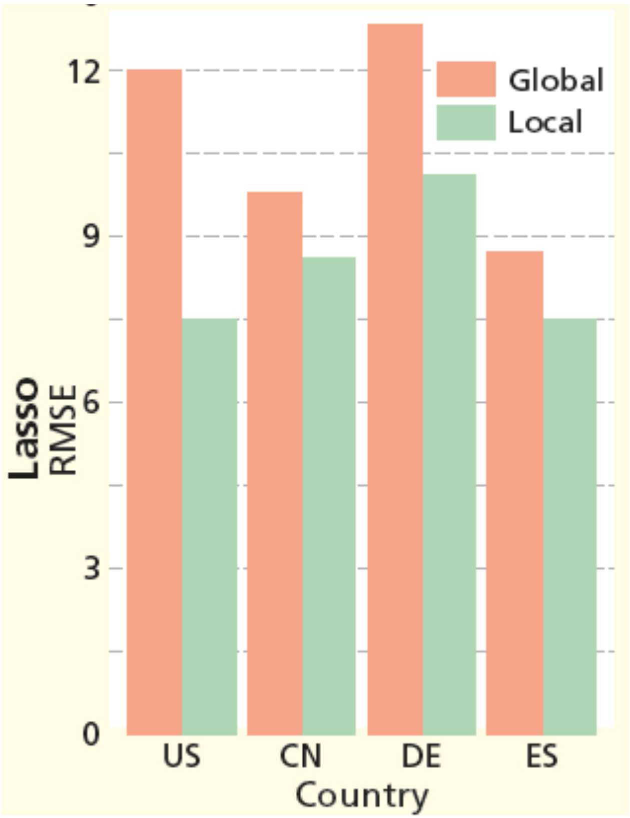
Median of bootstrapped RF OOB R-squared
Day: 0.70
Night: 0.72

RMSE: root mean squared error
MAE: mean absolute error
IQR: interquartile range

LM: Multiple linear regression
RF: random forest
SGB: Stochastic gradient boosting
XGB: xgboost

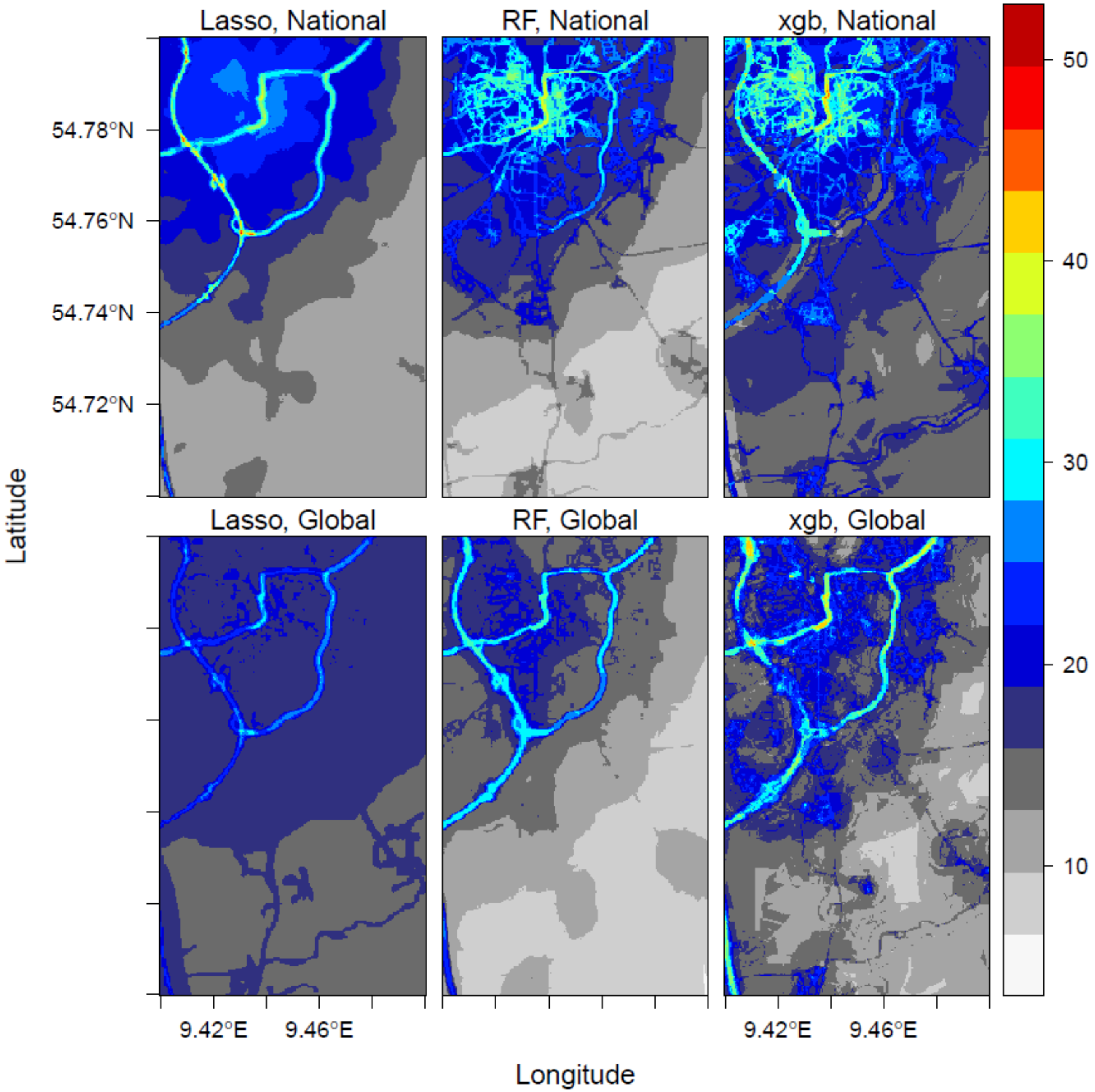
Result: global and national models
RF vs. Lasso





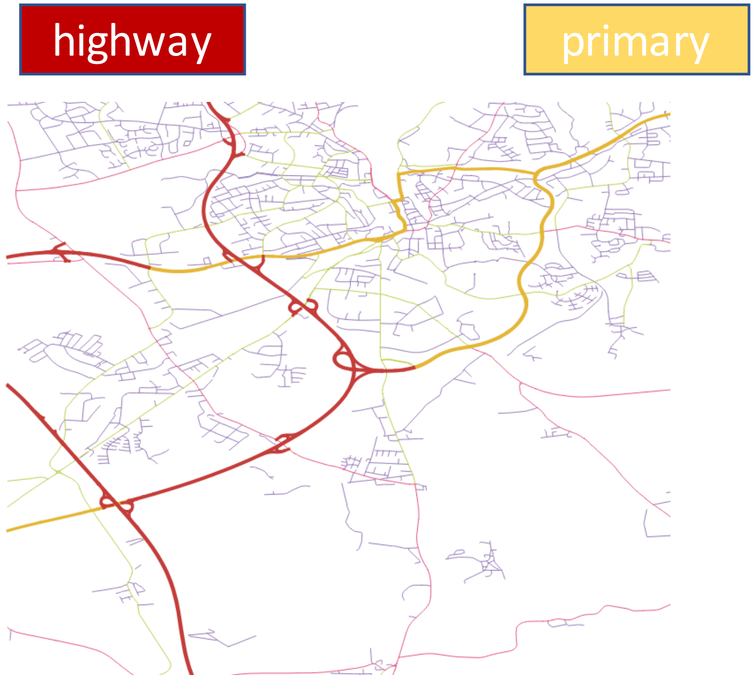
CN: China
DE: Germany
ES: Spain

Conclusion: random forest is more suitable than Lasso for a global NO2 model and using random forest can achieve an accuracy as good as national models.



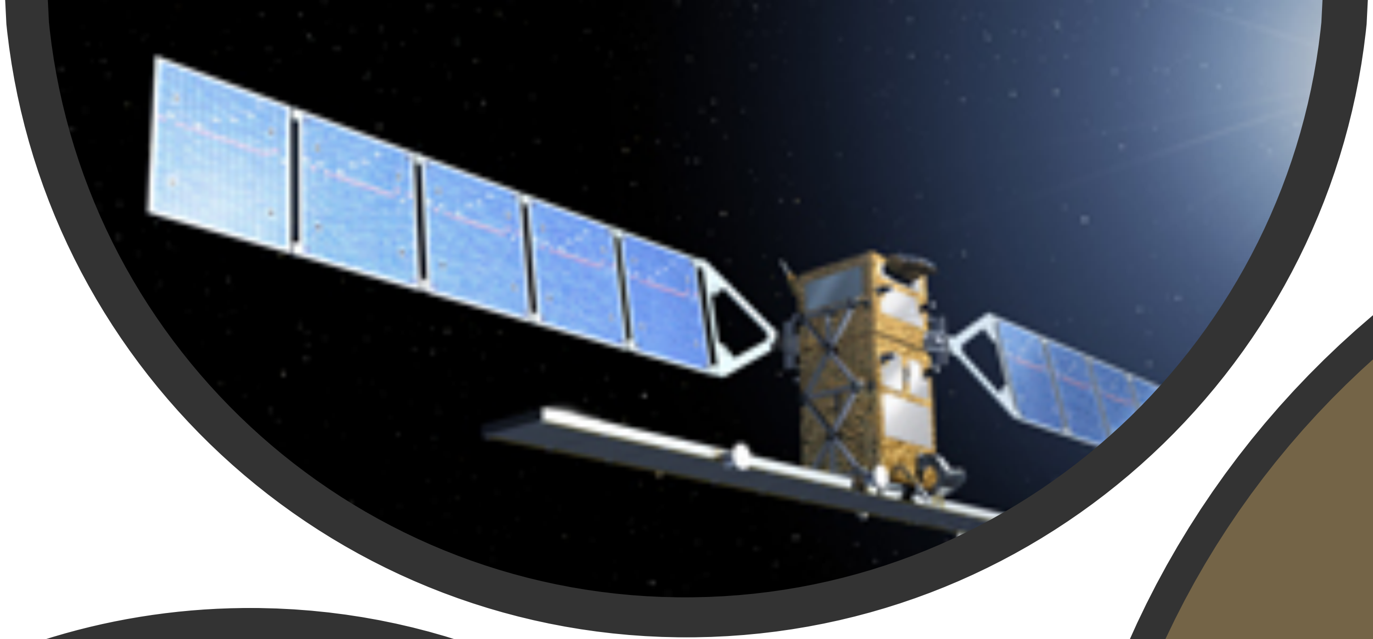
Accuracy obtained by regularized regression and tree-based methods are more similar than for the global model.

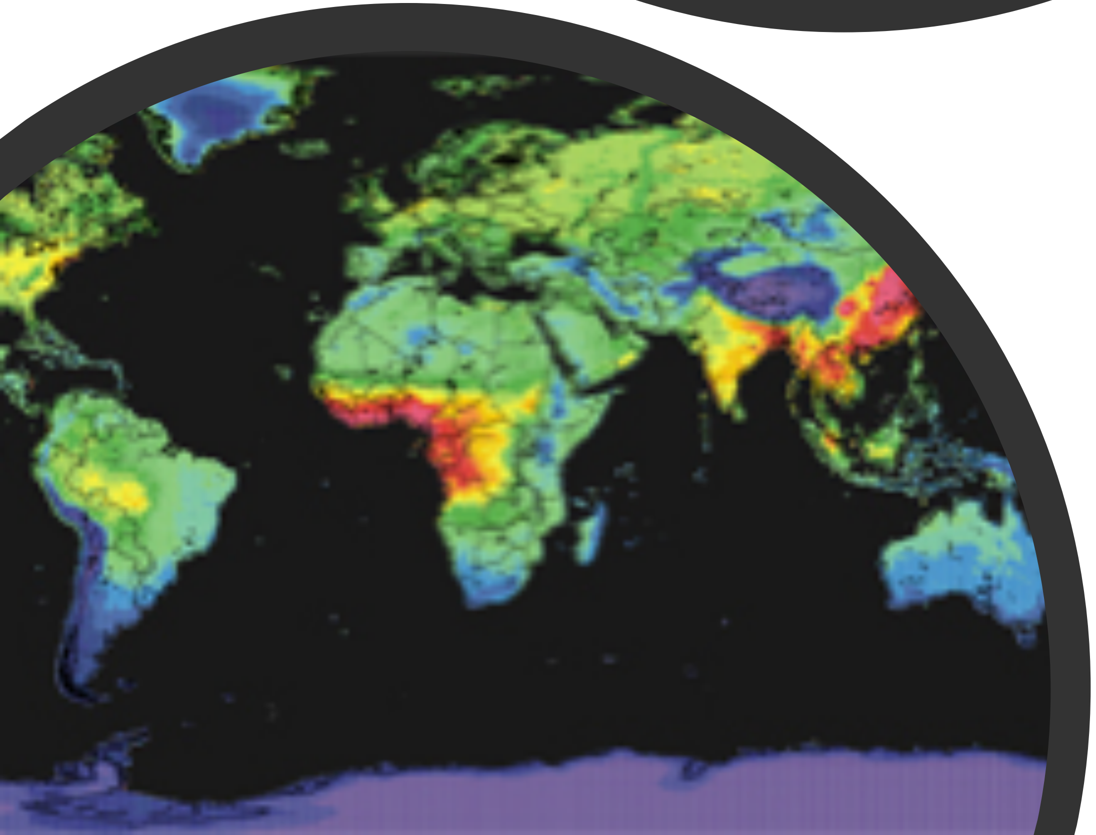
Germany



In the afternoon

- Air pollution modeling and statistical learning methods
- Go in-depth the modelling process, information from the model
- results





Thank you