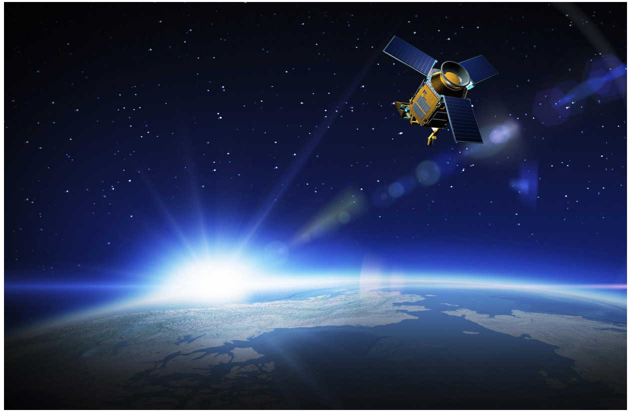
Global air pollution modelling



University of Utrecht, The Netherlands
Global Geo Health Data Center

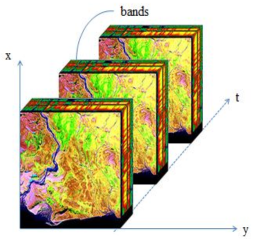
Meng Lu

My background

Remote sensing and information engineering

Spatiotemporal change modelling from multidimensional arrays

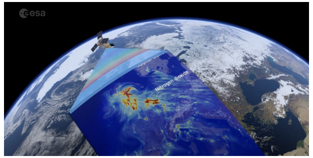
ifgi
Institut für Geoinformatik
Universität Münster



Spatiotemporal data analysis

Universiteit Utrecht

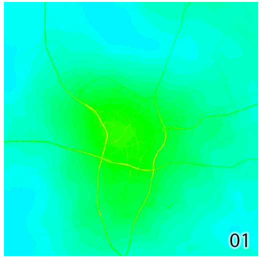
Remote sensing
Image analysis



Exposure
assessment

Agent-based
modeling

Mapping



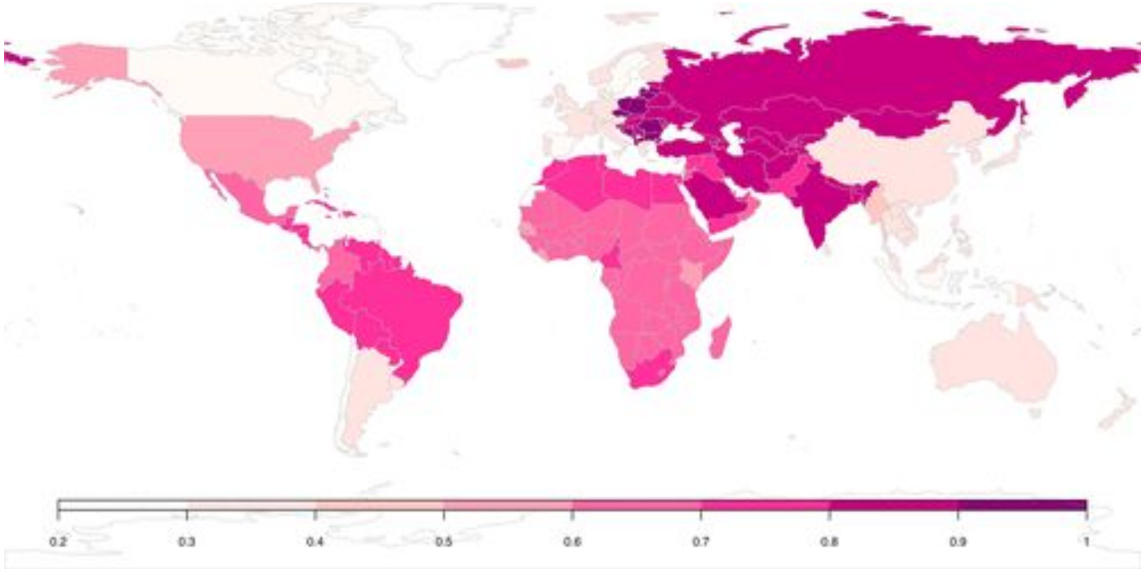
Goal:
Global, high spatiotemporal resolution NO_2 mapping

NO_2: is part of a group of gaseous air pollutants produced as a result of road traffic and other fossil fuel combustion processes.

Why global?

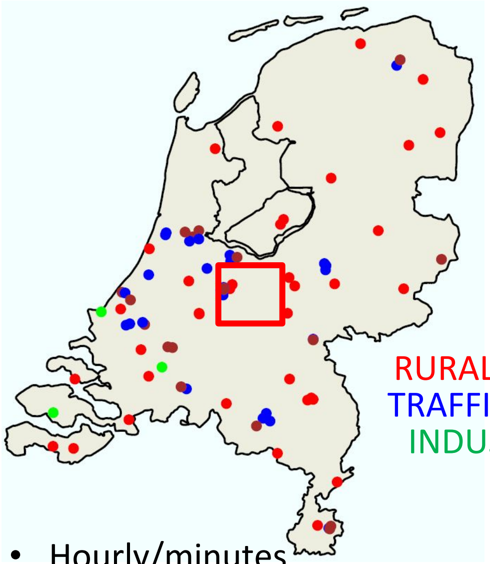
- Lacking of ground monitoring stations in many countries.
- Consistent epidemiological studies world-wide

  Inequality in health and air pollution monitors in middle and low income countries

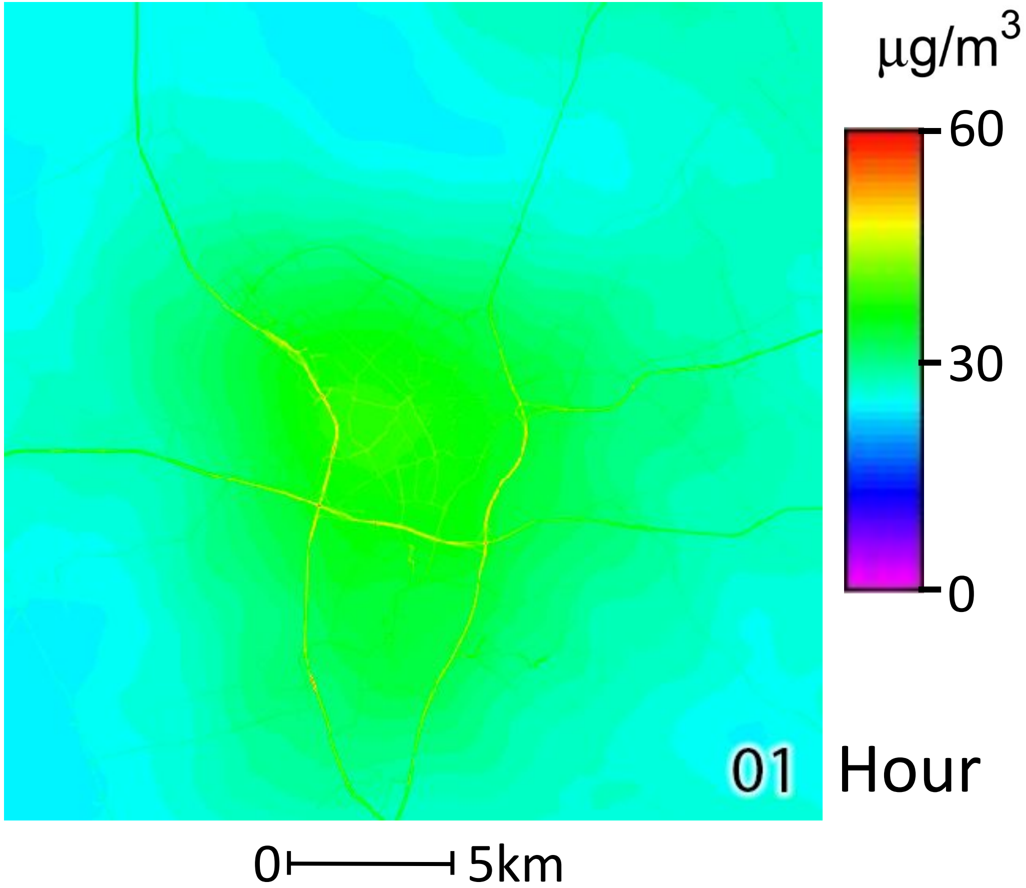


Shaddock et al., 2018

Why high spatial and temporal resolutions?



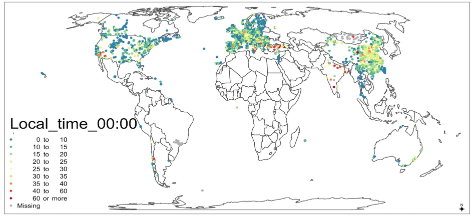
- Hourly/minutes
- LML data (landelijk meetnet luchtkwaliteit)



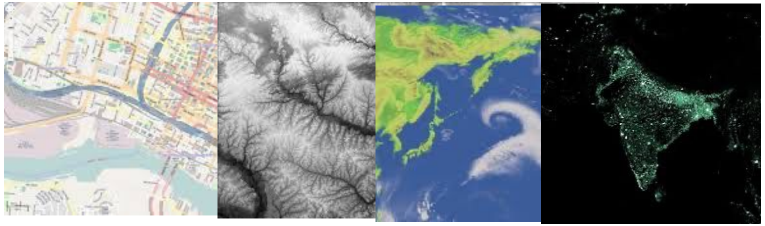
Air pollution modelling methods

- Process-based models:
  Air dispersion models
  Chemical transportation models

- Statistical-based methods:
  Information form predictors, spatiotemporal dependence.

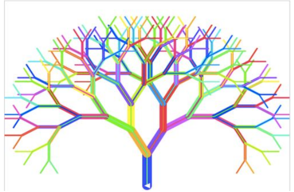


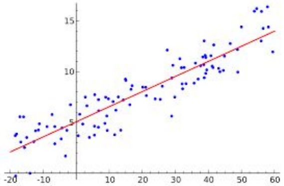
Station measurements



Predictors

Machine learning vs. Linear regression models





Global models Vs. National models

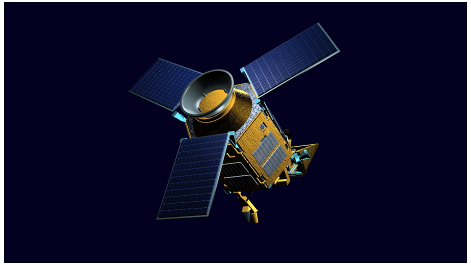
Daytimg models Vs. Nighttime models

Lu, M. Schmitz, O., Van de Hoogh, K., Qin, K., & Karssenberg, D. Evaluation of different methods and data sources to optimise modelling of NO2 at a global scale. Environmental international

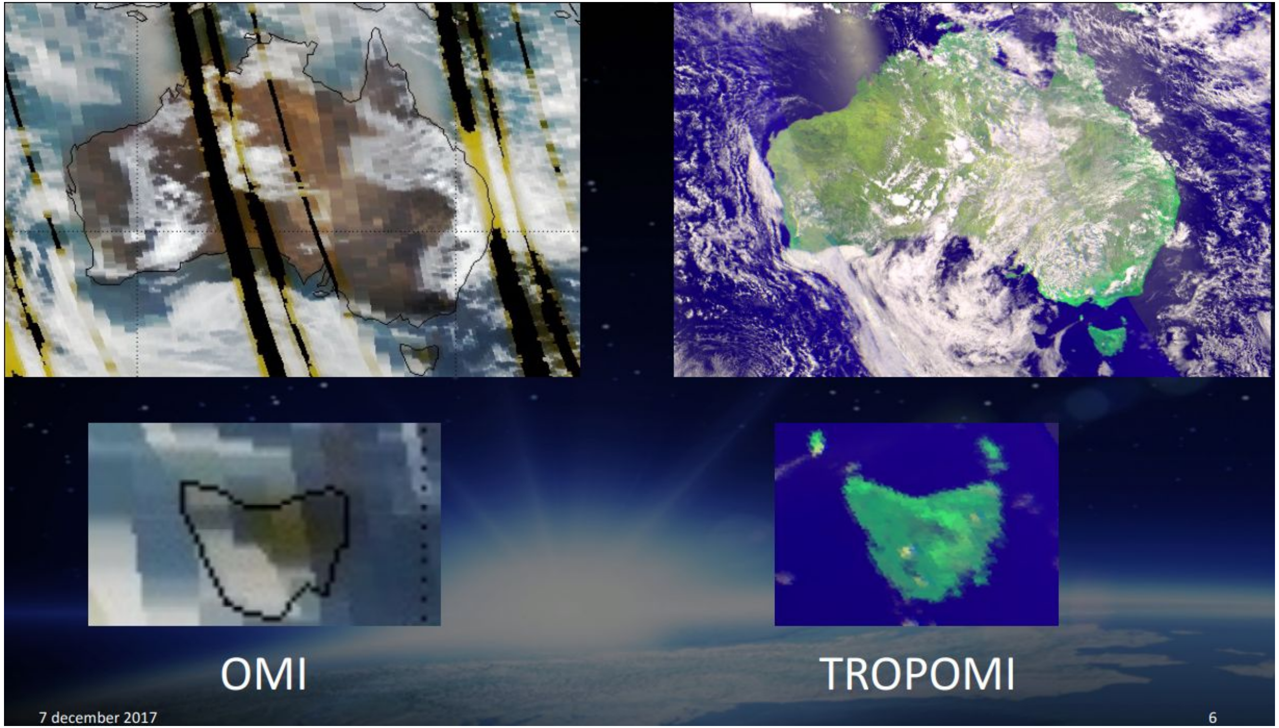
Tropomi: Sentinel 5p (TROPOspheric Monitoring Instrument)

Launched Oct. 2017,

Resolution (NO2)
2018-2019: 7 km
Since Oct. 2019: 5.5 km



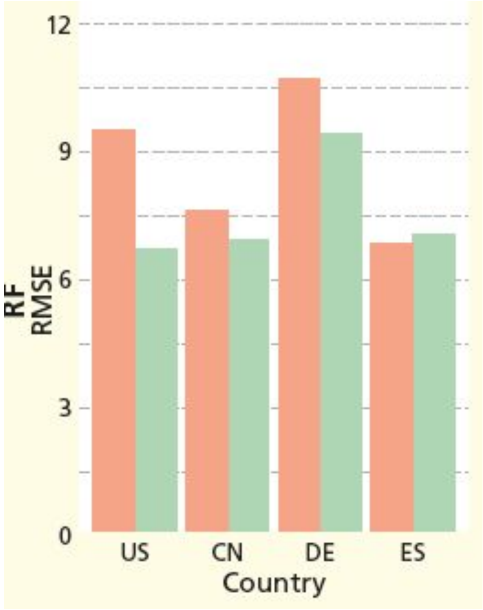
Measurements: NO_2, O_3 (7km × 28km), SO_2, methane and CO



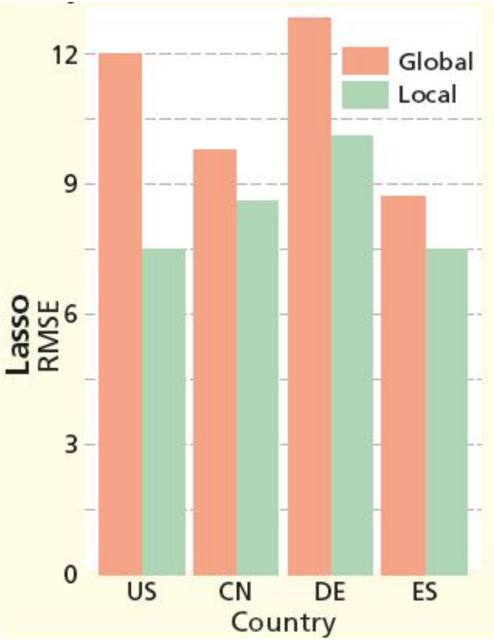
Souce [3]

Global Vs. National models

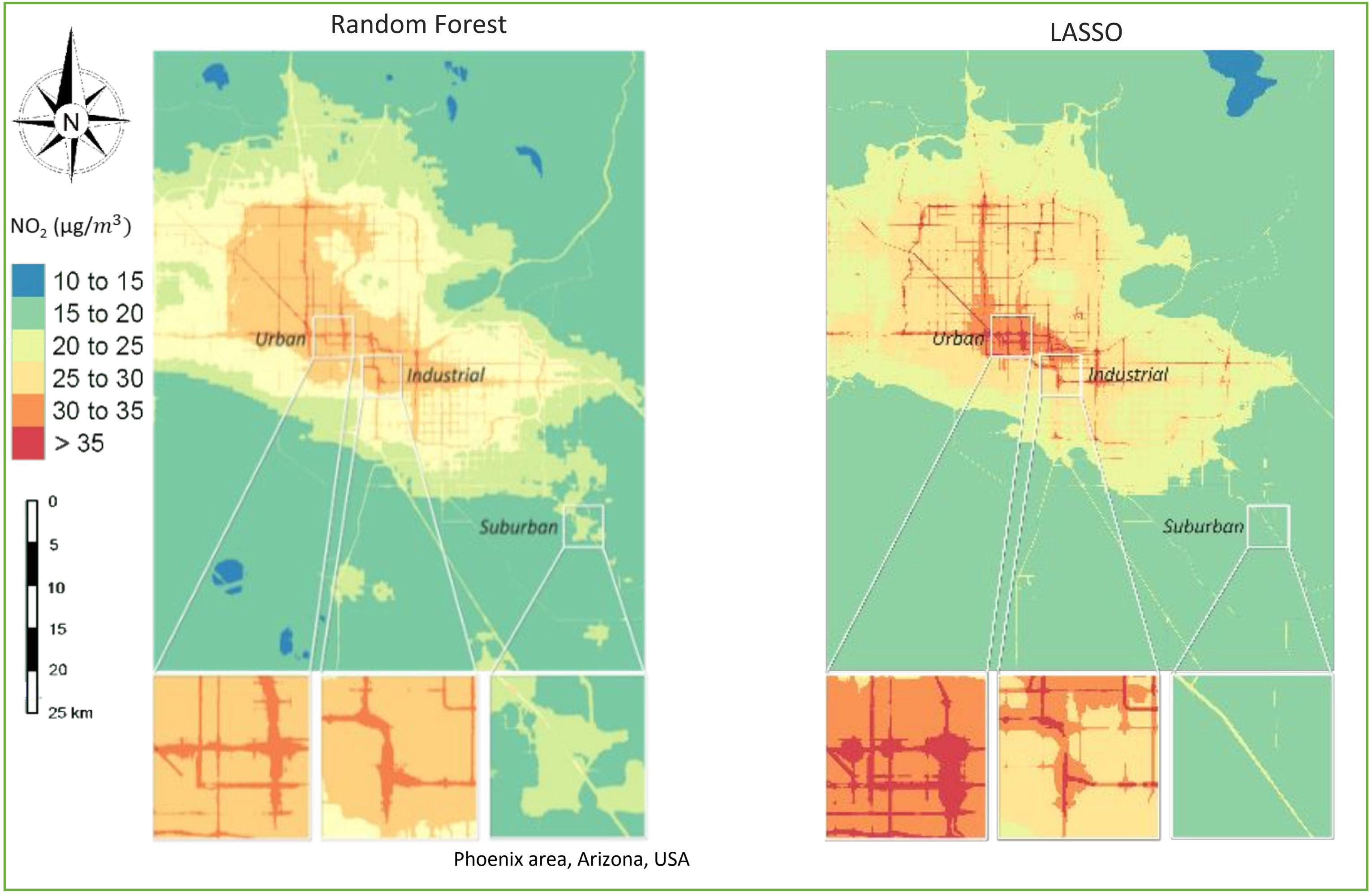
Random Forest

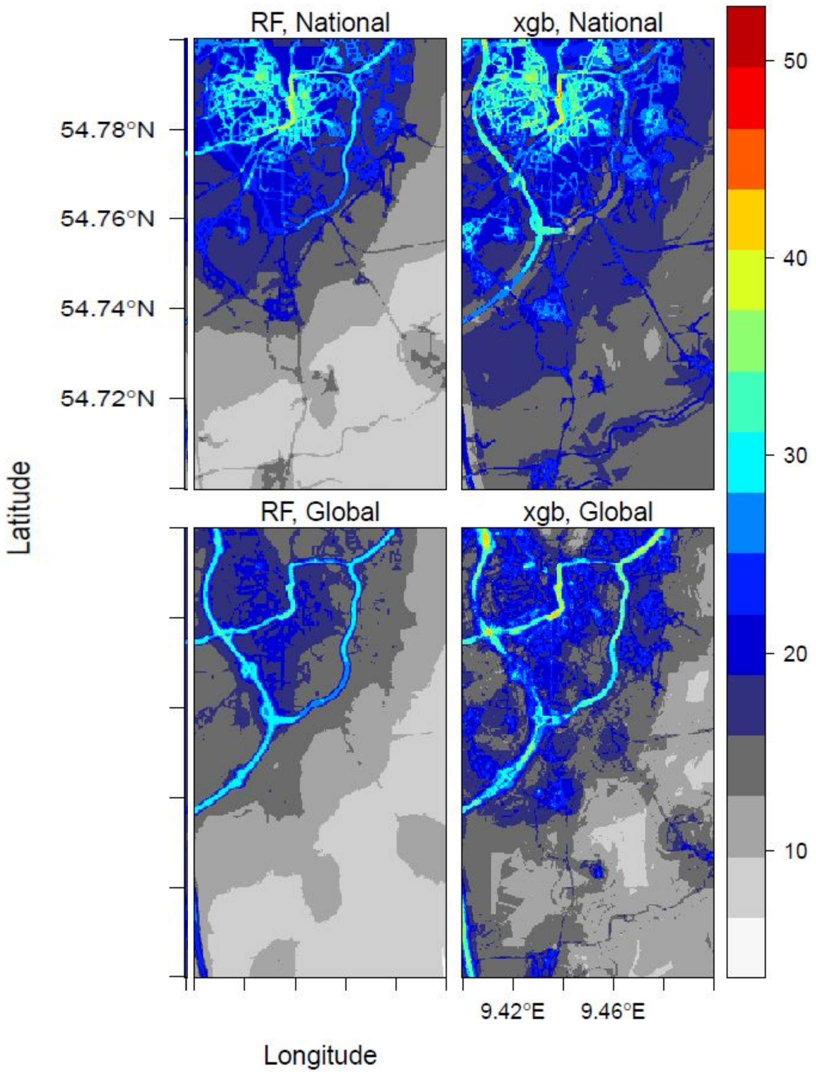


Lasso



RF: random forest
CN: China
DE: Germany
ES: Spain


Random Forest
LASSO
N
NO$_2$ (μg/m^3)
10 to 15
15 to 20
20 to 25
25 to 30
30 to 35
> 35
0
5
10
15
20
25 km
Urban
Industrial
Suburban
Phoenix area, Arizona, USA



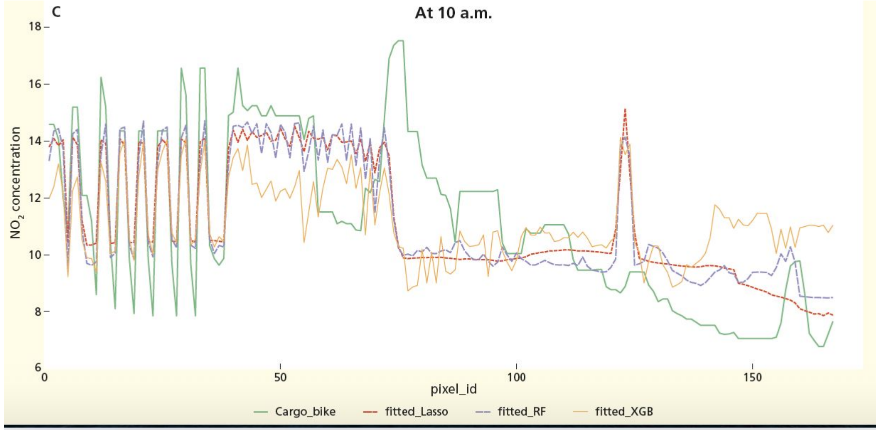
Prediction Patterns

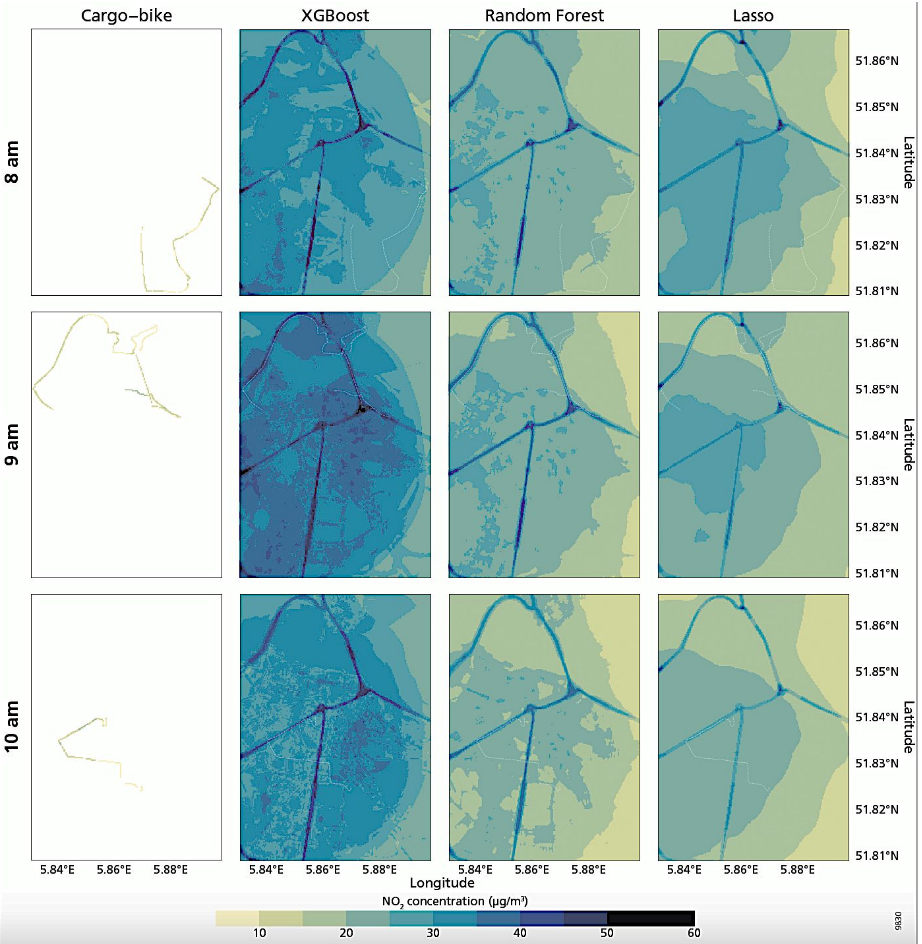
highway

primary

Germany

Further validation: cargo-bike measurements





Statistical challenges in global-scale spatiotemporal regression

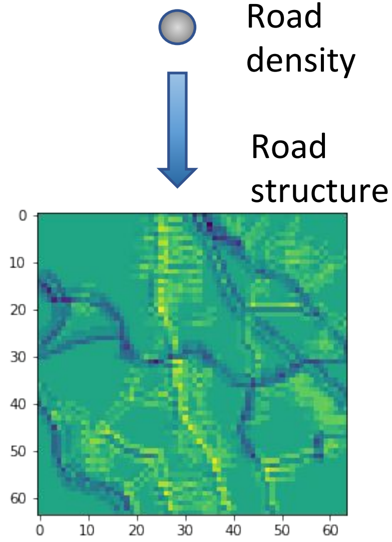
Challenges

- Spatially heterogeneous predictor-response relationships
- Spatiotemporal variability
- **Integrate more spatial information**

On-going

Temporally-resolved and hierarchical modelling:

Mixed-effect models and deep convolutional neural networks



In the workshop

1. Introduction and data (10%)
2. Statistical modelling techniques (70%)

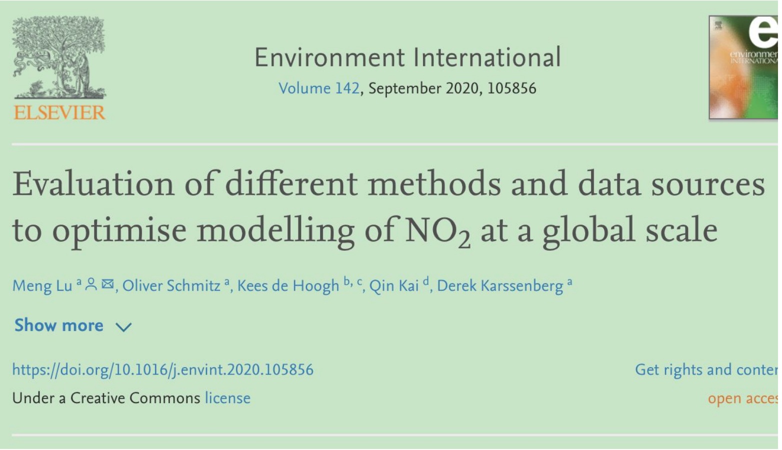
Start:
Ensemble tree-based ML

Deal with overfitting:
Regularization
Post-processing

More spatial information:
Deep convolutional NN

3. Modelling process in practice (20%)

Querying and processing OpenStreetMap,
hyperparameter tuning,
cross-validation, mapping, model comparison


ELSEVIER
Environment International
Volume 142, September 2020, 105856

Evaluation of different methods and data sources to optimise modelling of NO_2 at a global scale

Meng Lu [a], Oliver Schmitz [a], Kees de Hoogh [b, c], Qin Kai [d], Derek Karssenberg [a]

Show more

https://doi.org/10.1016/j.envint.2020.105856

Get rights and conter

Under a Creative Commons license

open acces

Thank you

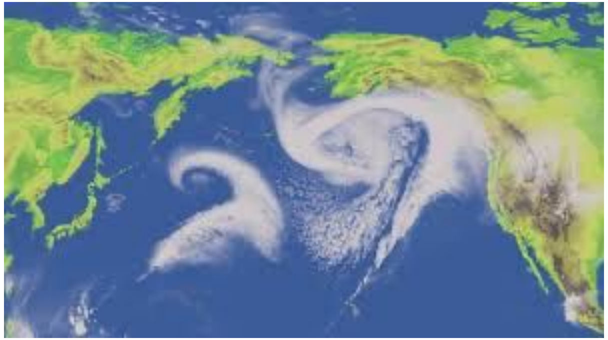
https://github.com/mengluchu/OpenGeoHub2020

Geospatial predictors

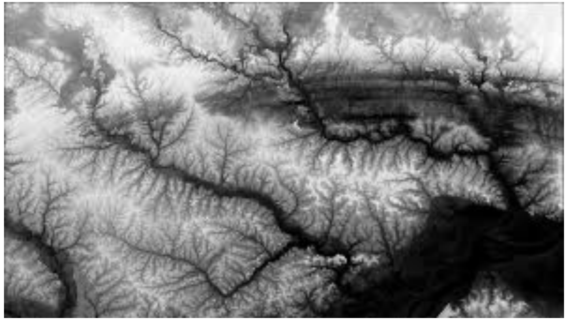
Road and industrial area from OpenStreetMap (vector)

Earth nightlight (500 m)
as a proxy for socio-demographics

Population (100 m)
From human settlement layers



Wind speed and temperature



Elevation: SRTM (90 m)