Inferring High-resolution Near-surface NO_2 Concentrations over Belgium through Convolutional Neural Networks

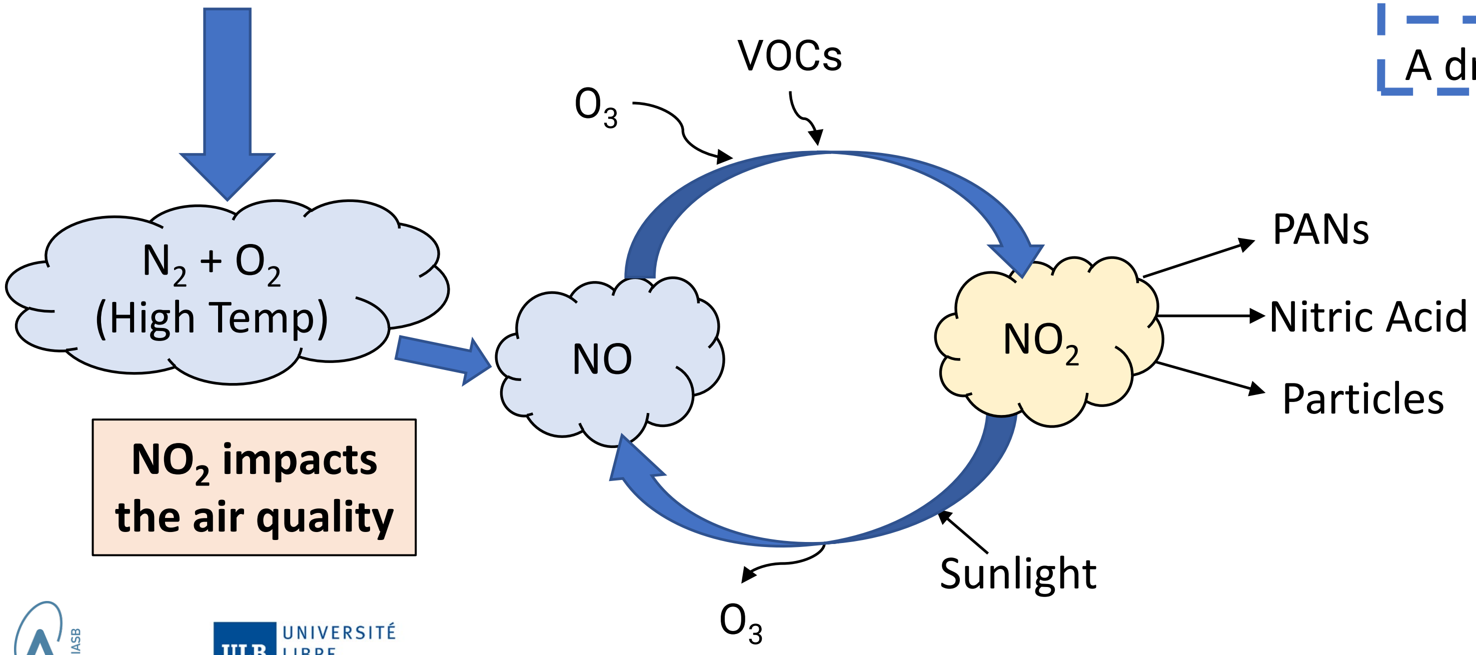
Wenfu Sun[1,2]
wenfu.sun@aeronomie.be

Frederik Tack[1], Lieven Clarisse[2], Rochelle Schneider[3], Michel Van Roozendael[1]

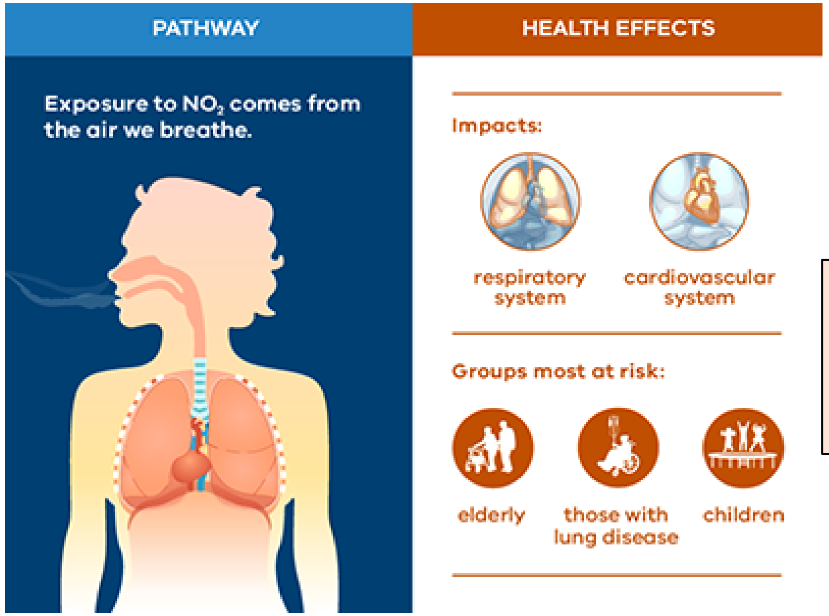
[1]BIRA-IASB, Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
[2]Spectroscopie de l`Atmosphère, Service de Chimie Quantique et Photophysique, Université Libre de Bruxelles, Brussels, Belgium
[3]Φ-Lab, European Space Agency (ESA-ESRIN), Frascati, Italy

2022-11-15

Concerns about NO_2



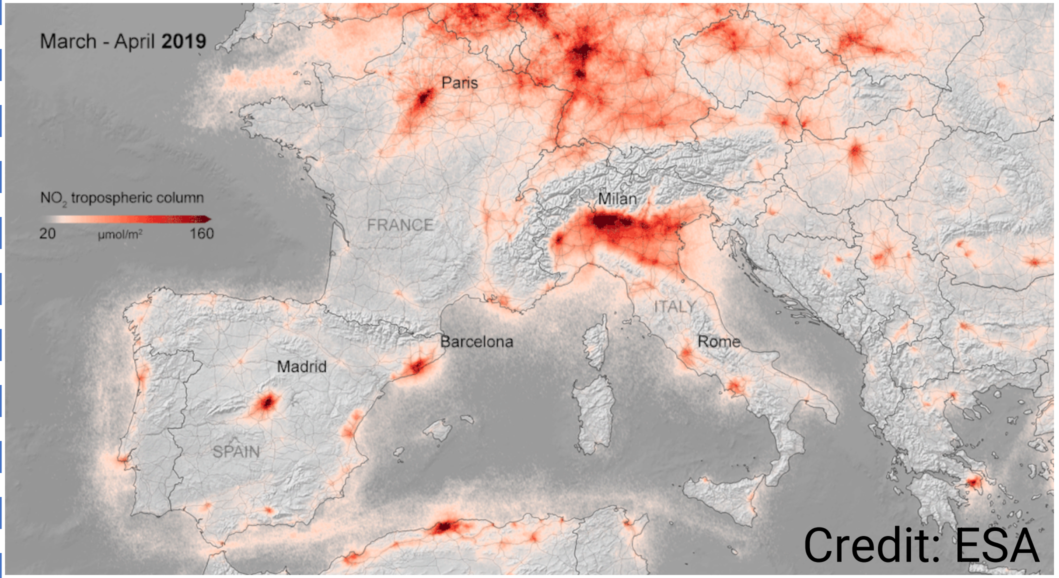
NO_2 impacts the air quality



https://www.epa.vic.gov.au/for-community/environmental-information/air-quality/nitrogen-dioxide-in-the-air

NO_2 health effects

A dramatic decrease in NO_2 during the Covid-19 outbreak



Credit: ESA

Inferring NO_2 near-surface concentrations

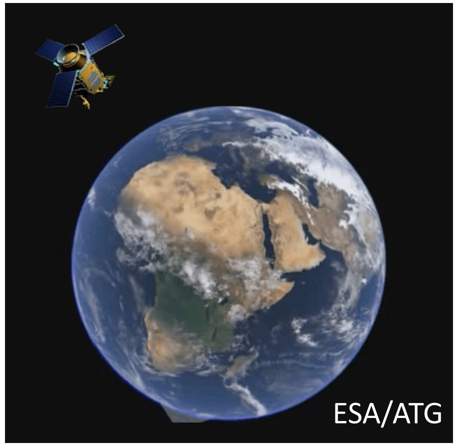
Surface NO_2 is of great concern due to its adverse impacts on air quality and human health

Measurements

- ☺ Accurate and consistent estimations
- ☹ Limited spatial coverage

Satellite observations

- ☺ Large spatial coverage
- 😐 Limited temporal resolution
- ☹ Coarse spatial resolution
- ☹ Low sensitivity to surface NO_2



\+

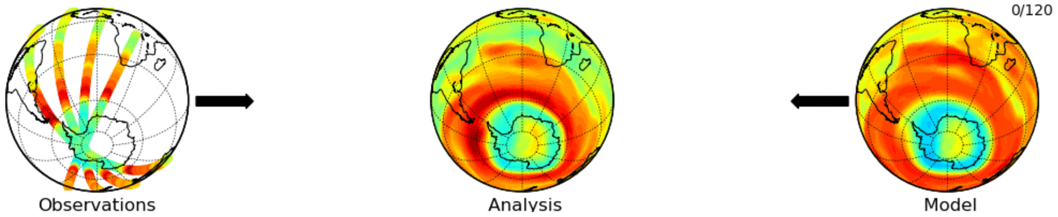
Ancillary datasets

Meteorology Emissions Society
Land information ……

→

Physical method

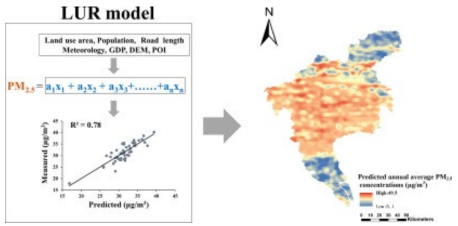
- ☺ Good knowledge (physics and chemistry)
- ☺ Can be constrained by reality (data assimilation)
- ☹ Computationally intensive, coarse spatial resolution
- ☹ Biases caused by mechanism and EMI inventories



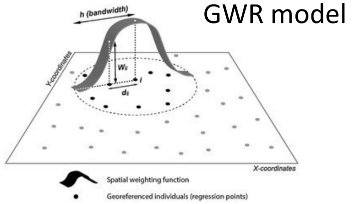
https://www.issibern.ch/teams/dataassimsphere/

Empirical statistical method

- ☺ Flexible modeling and can achieve high resolution
- ☹ Limited ability for very complex nonlinear relationship



Mo *et al.*, *Sci. Total Environ.*, (2021)



Feuillet *et al.*, *Int. J. Health Geogr.*, (2015)

Inferring NO_2 near-surface concentrations by machine learning

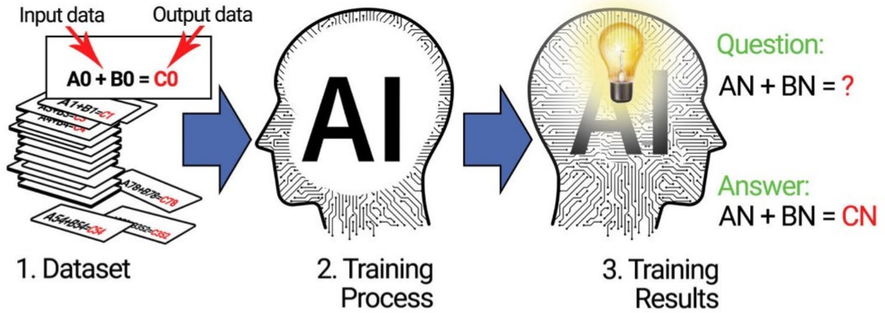
BIRA-IASB aeronomie·be
ULB UNIVERSITÉ LIBRE DE BRUXELLES

Current ML methods

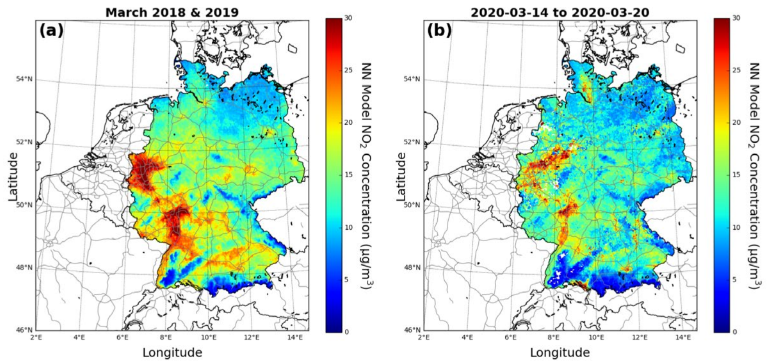
- Random Forest (RF)
- eXtreme Gradient Boosting (XGBoost)
- Neural Network (NN)
- ……

Machine learning method

- 🙂 Can establish a complex nonlinear mapping
- 🙂 High accuracy, efficiency, and fine resolution
- ☹ Risk of manufacturing artifacts when mapping

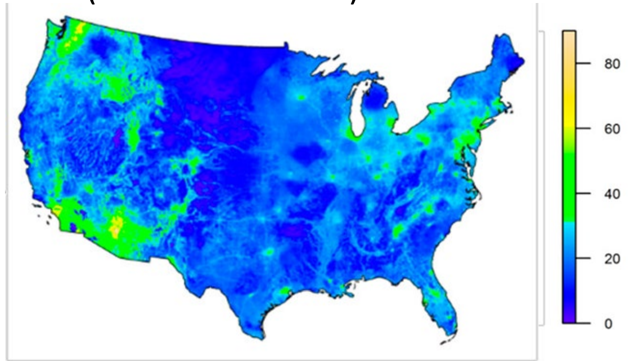


Daily 0.5km x 0.5km (NN)



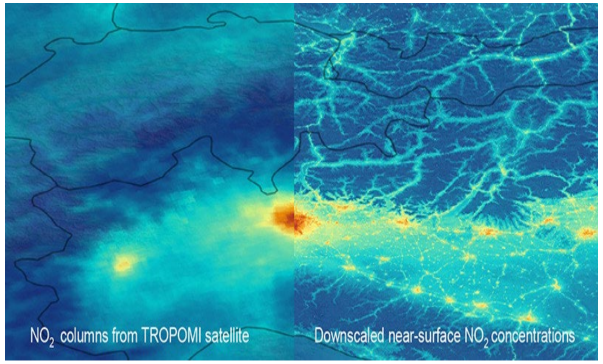
Chan *et al.*, *Remote Sens.*, **2021**

Daily 1km x 1km (NN+XGBoost+RF)



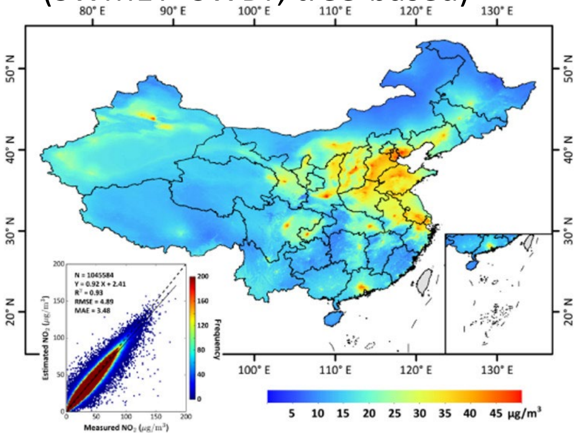
Di *et al.*, *Environ. Sci. Technol.*, **2020**

Hourly 100m x 100m (XGBoost)



Kim *et al.*, *Remote Sens. Environ.*, **2021**

Daily 1km x 1km (SWMET+SWDF, tree-based)



Wei *et al.*, *Environ. Sci. Technol.*, **2022**

Surface NO_2 mapping for Belgium by machine learning

Predictor X (23 inputs in total):

- TROPOMI NO_2 L3 (daily, 1 km)
- Meteorology:
  - Wind, Temperature, Radiation, Boundary layer height, ……

MET

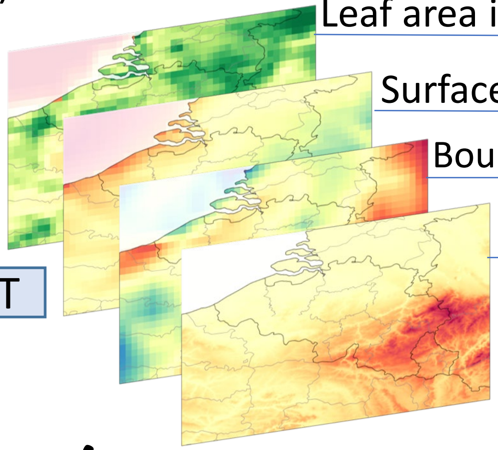
- Leaf area index
- Land Cover
- Elevation

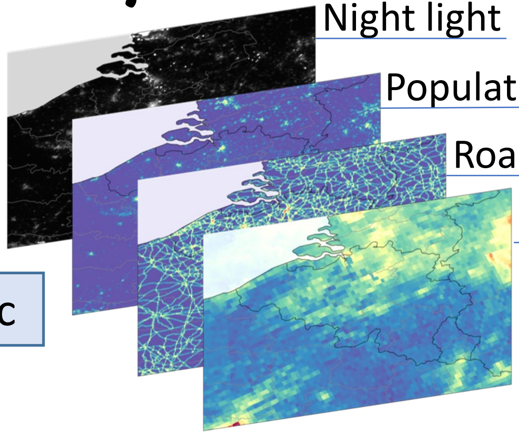
Land

- Emission inventory
- Road density
- Population
- Night light

Anthropogenic

- Day of week, Day of year

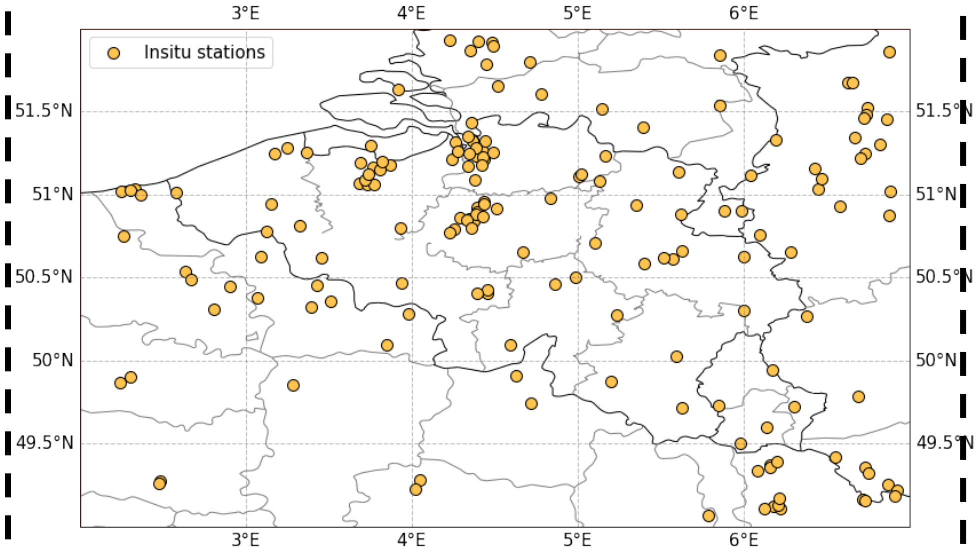




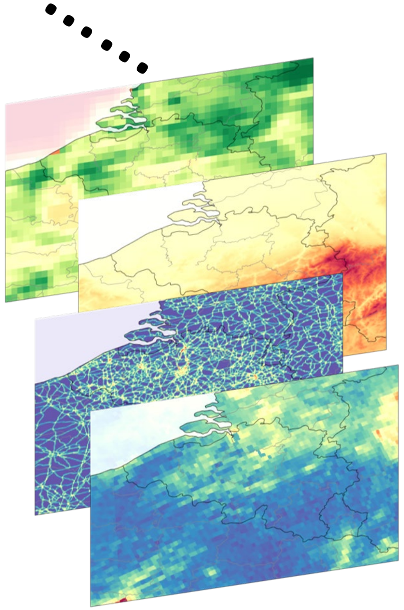
Machine Learning Model

Ground truth Y:

- Ground NO_2 daily measurements (2018-05 ~ 2020-12) (166 stations)



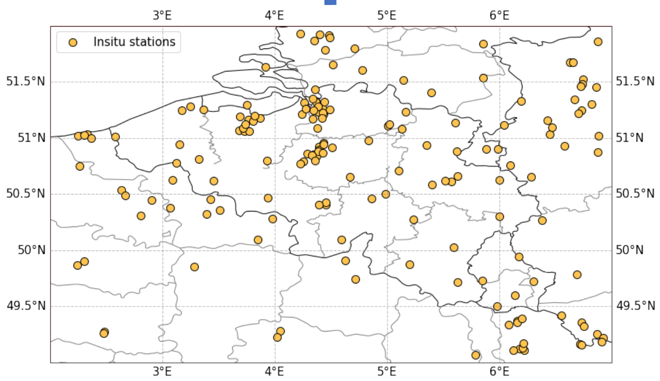
Surface NO_2 mapping for Belgium by machine learning



Predictors selection and preprocess

X: Predictors

Y: Ground NO_2 measurements



Machine Learning Model

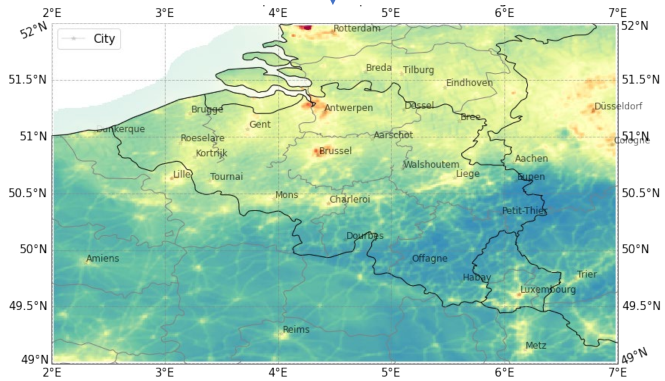
Model hyperparameters auto tunning

Train and Test on 10-folds site-based Cross-validation

X: Predictors

Trained Model

Surface NO_2 mapping (1km, daily)



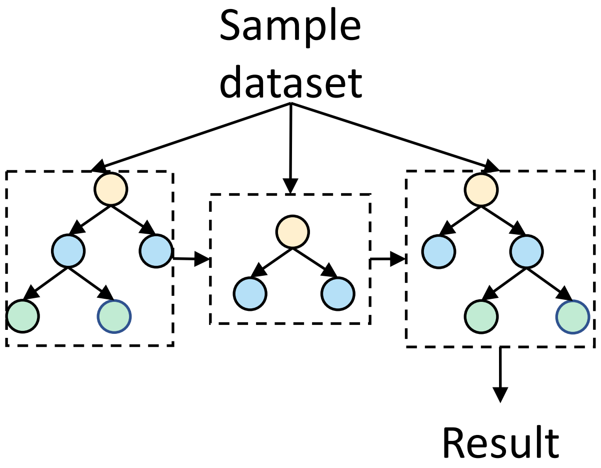
Surface NO_2 mapping for Belgium by machine learning

BIRA-IASB aeronomie.be
ULB UNIVERSITÉ LIBRE DE BRUXELLES

Start with models that used in previous works

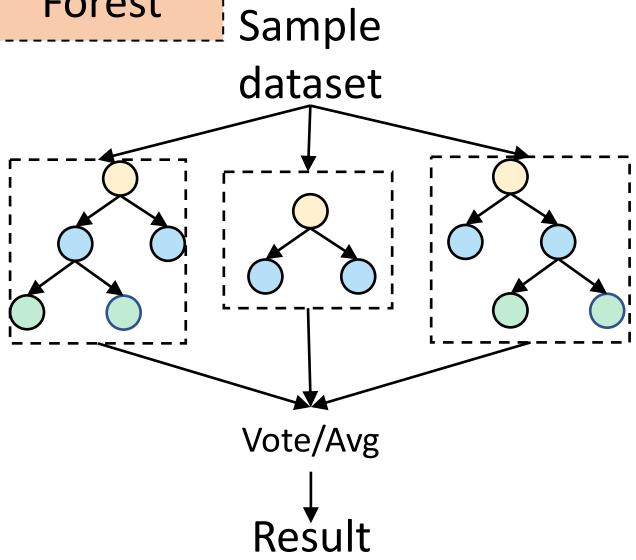
Test models on sites-based 10-folds Cross validation

XGBoost



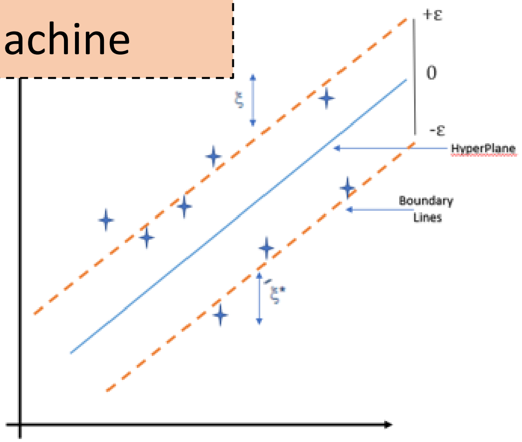
R^2: 0.74
Pearr: 0.86
RMSE: 5.47
MAE: 4.00

Random Forest



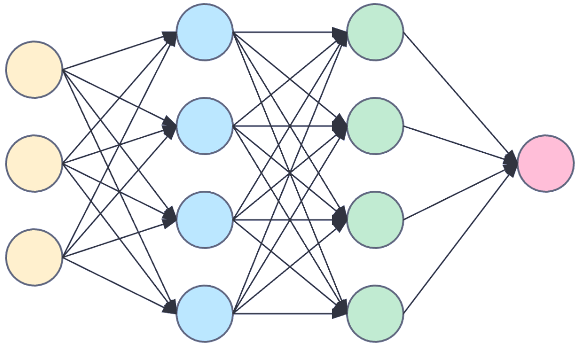
R^2: 0.59
Pearr: 0.78
RMSE: 6.92
MAE: 5.10

Support Vector Machine



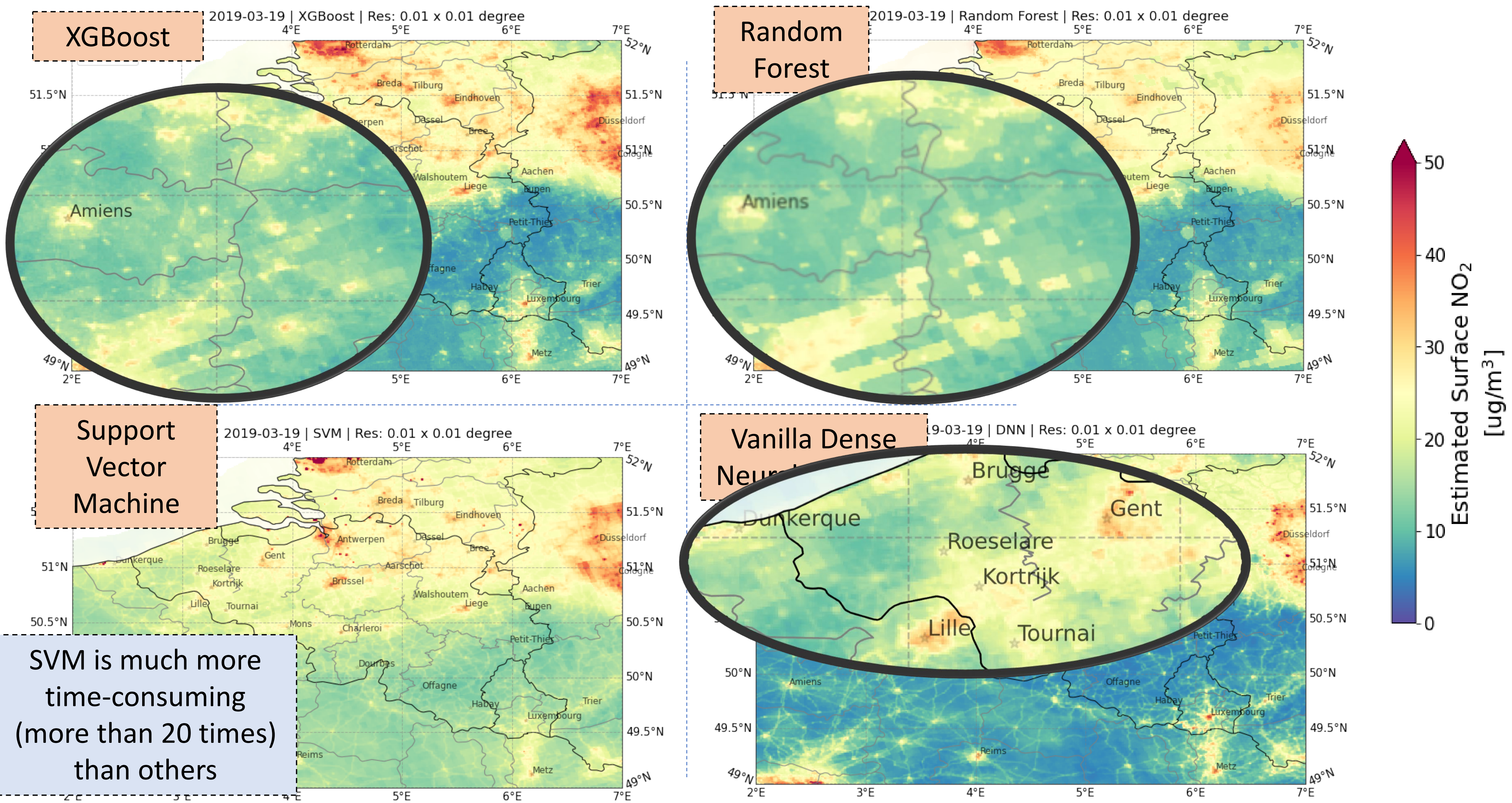
R^2: 0.62
Pearr: 0.79
RMSE: 6.62
MAE: 4.82

Vanilla Dense Neural Network



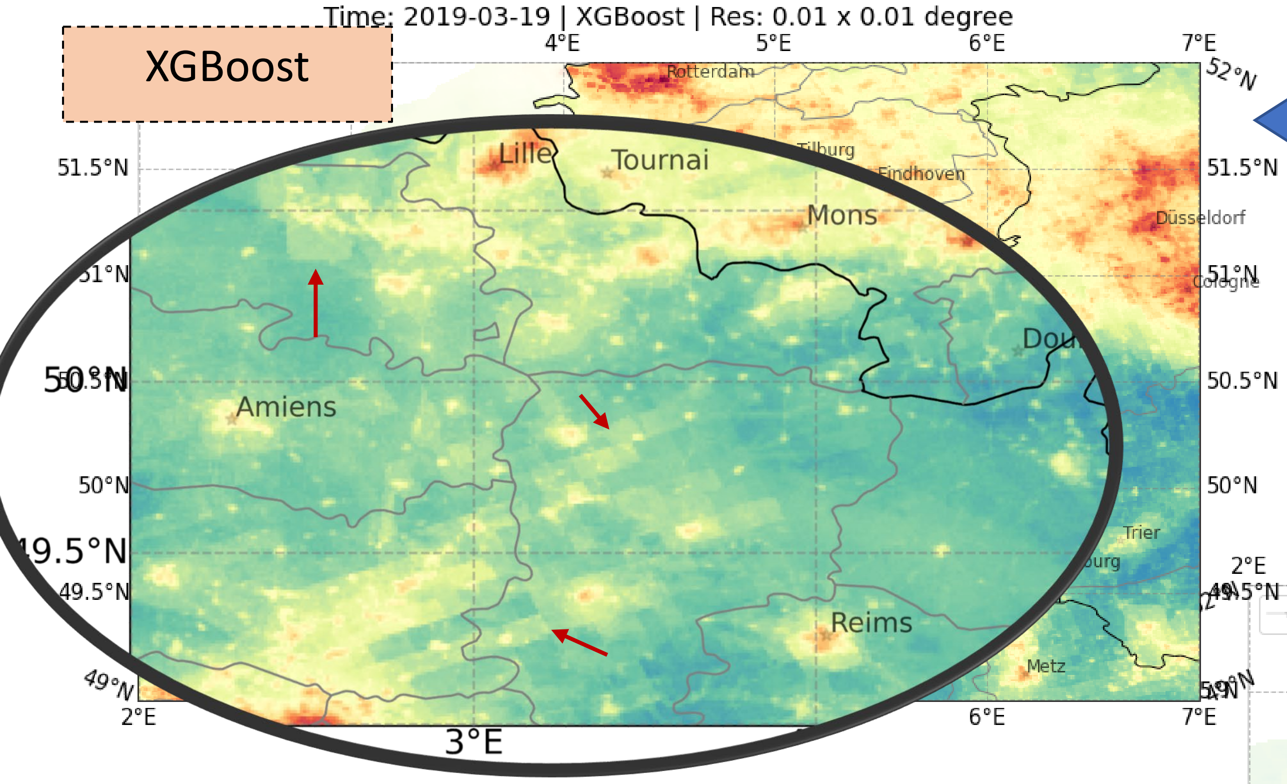
R^2: 0.65
Pearr: 0.81
RMSE: 6.42
MAE: 4.62

Surface NO_2 mapping for Belgium by machine learning



SVM is much more time-consuming (more than 20 times) than others

Surface NO_2 mapping by XGBoost



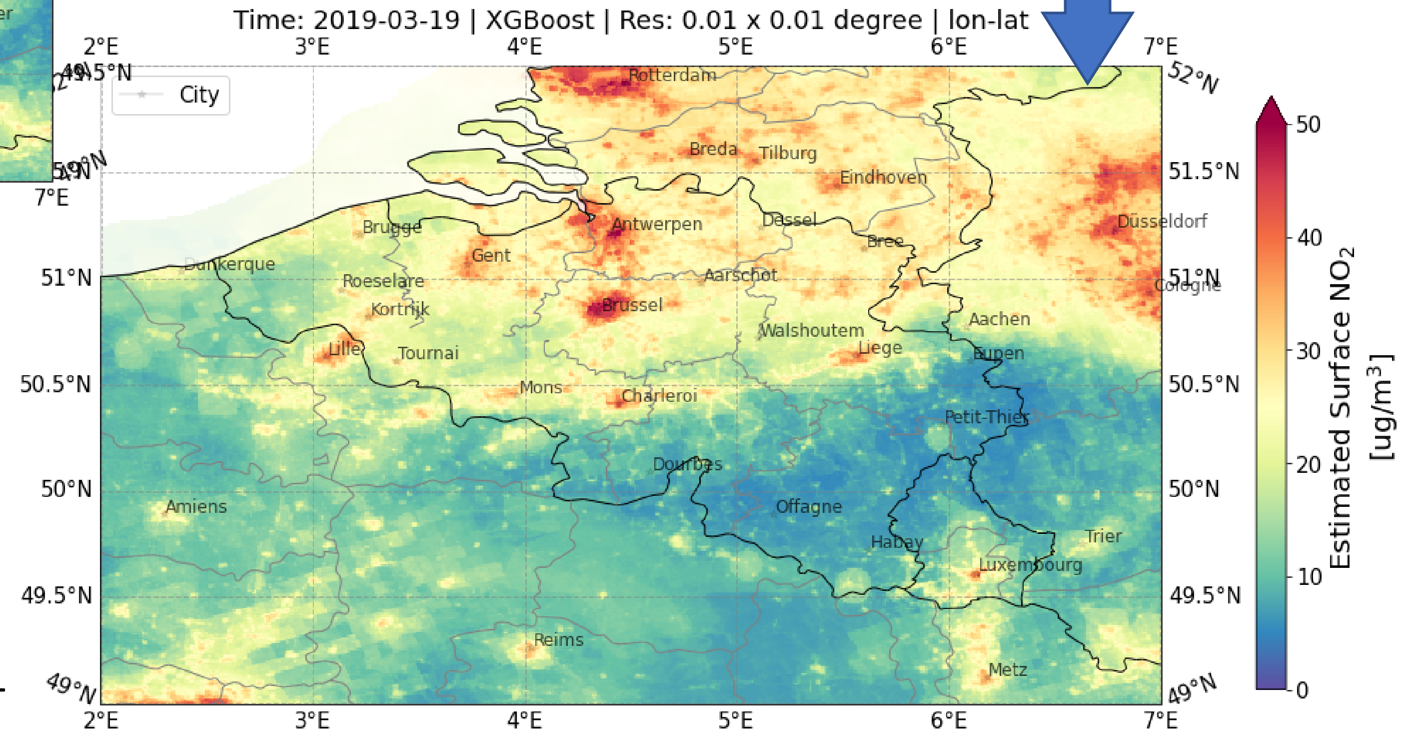
XGBoost achieves superior predictive accuracy but

- Prediction shows some checkerboard pattern of artifacts.

- Incorporating spatial locations (longitude, latitude) as predictors doesn't improve spatial continuity and mitigate artifacts in predictions.

Tree models exhibit **limitations** especially regarding the ability **of producing spatially continuous** maps without step-wise changes. …. Nevertheless, some **artifacts remained**, which were most pronounced at locations with a complex distribution of emission sources.

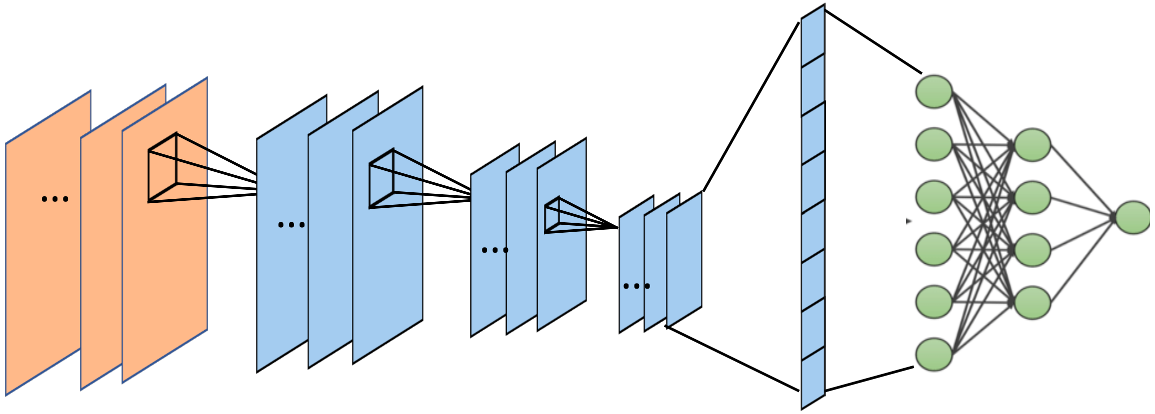
-- Kim et al., *Remote Sens. Environ.*, 2021



Surface NO_2 mapping by 2D-CNN

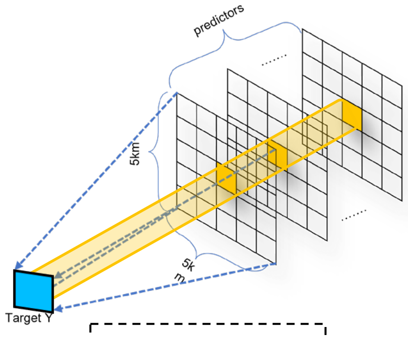
Sun *et al.*, (2022, in preparation)

2D-Convolutional Neural Network



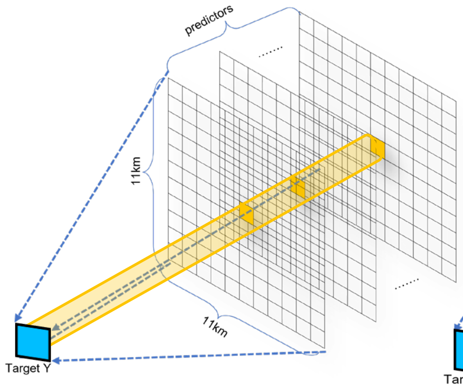
- Given NO_2 has a relationship with surrounding environments.
- 2D-CNN considers neighboring pixels of the target grid.

- Receptive field size is essential to model predictive accuracy and pattern continuity.
- A large receptive field will introduce more predictors' variance in the model
  - The model will benefit from fine-resolution predictors.
  - The model will become conservative when many predictors are downsampled from coarse resolution as variance doesn't increase significantly.



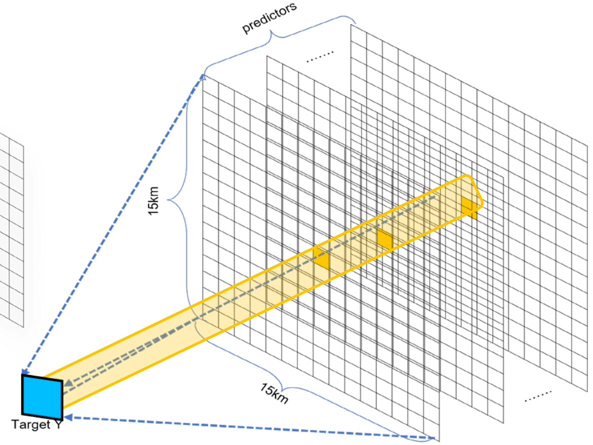
5 Km

R^2: 0.68
Pearr: 0.82
RMSE: 6.14
MAE: 4.47



11 Km

R^2: 0.66
Pearr: 0.82
RMSE: 6.31
MAE: 4.59

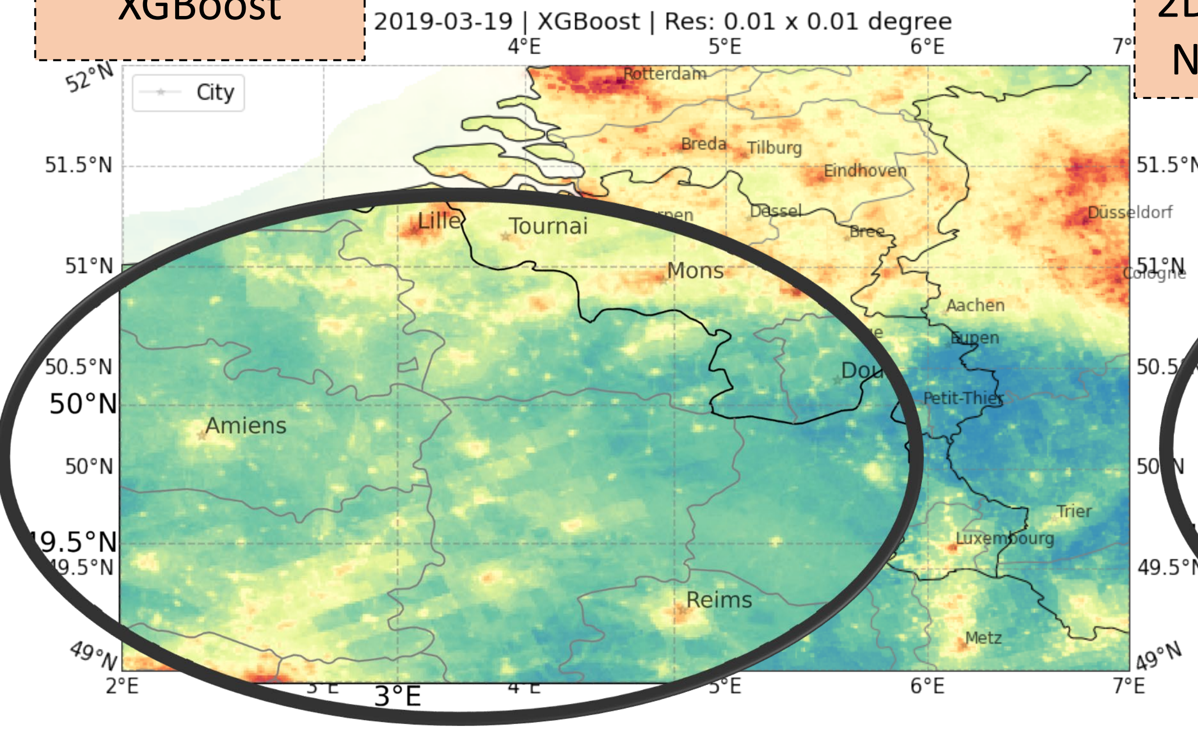


15 Km

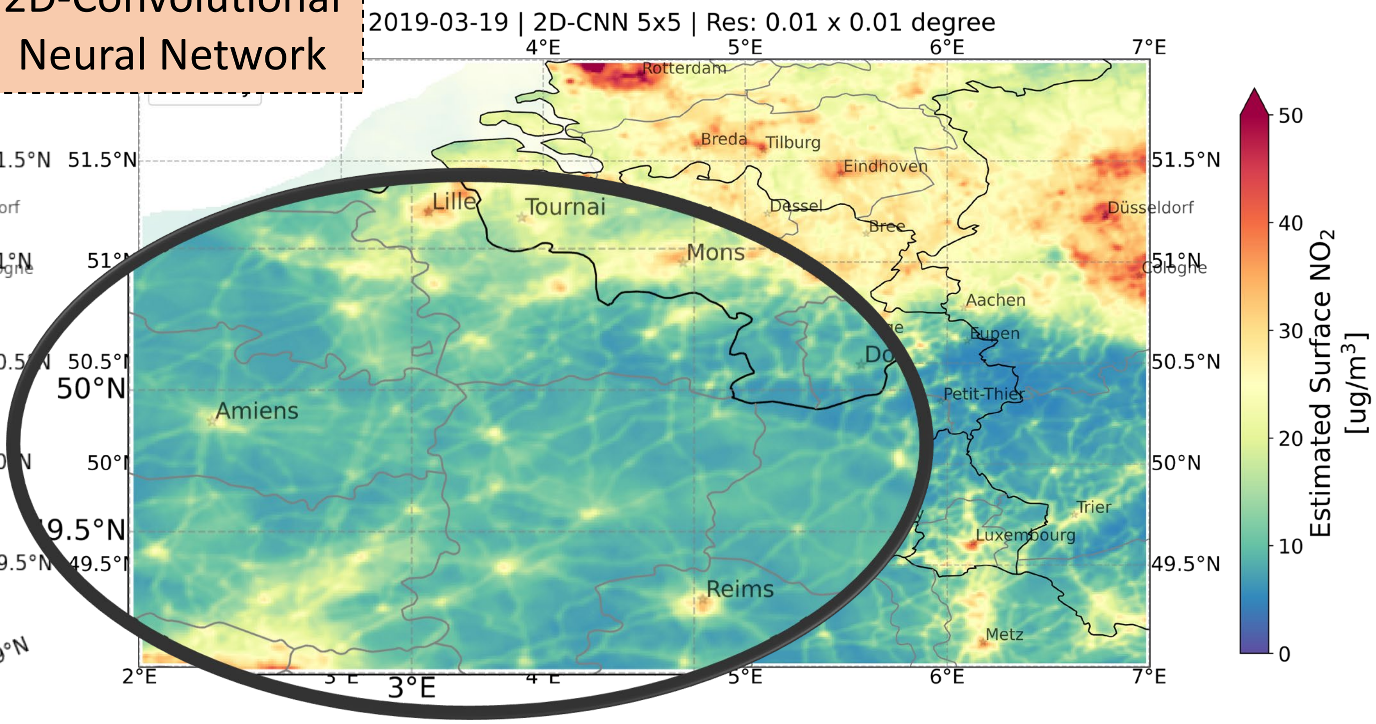
R^2: 0.61
Pearr: 0.82
RMSE: 6.78
MAE: 4.79

Compare XGBoost and 2D-CNN on spatial mapping

XGBoost



2D-Convolutional Neural Network



R^2: 0.74
Pearr: 0.86
RMSE: 5.47
MAE: 4.00

A tree-based XGBoost performs best in statistical tests, but discrete patterns indicate obvious artifacts, especially in local mapping.

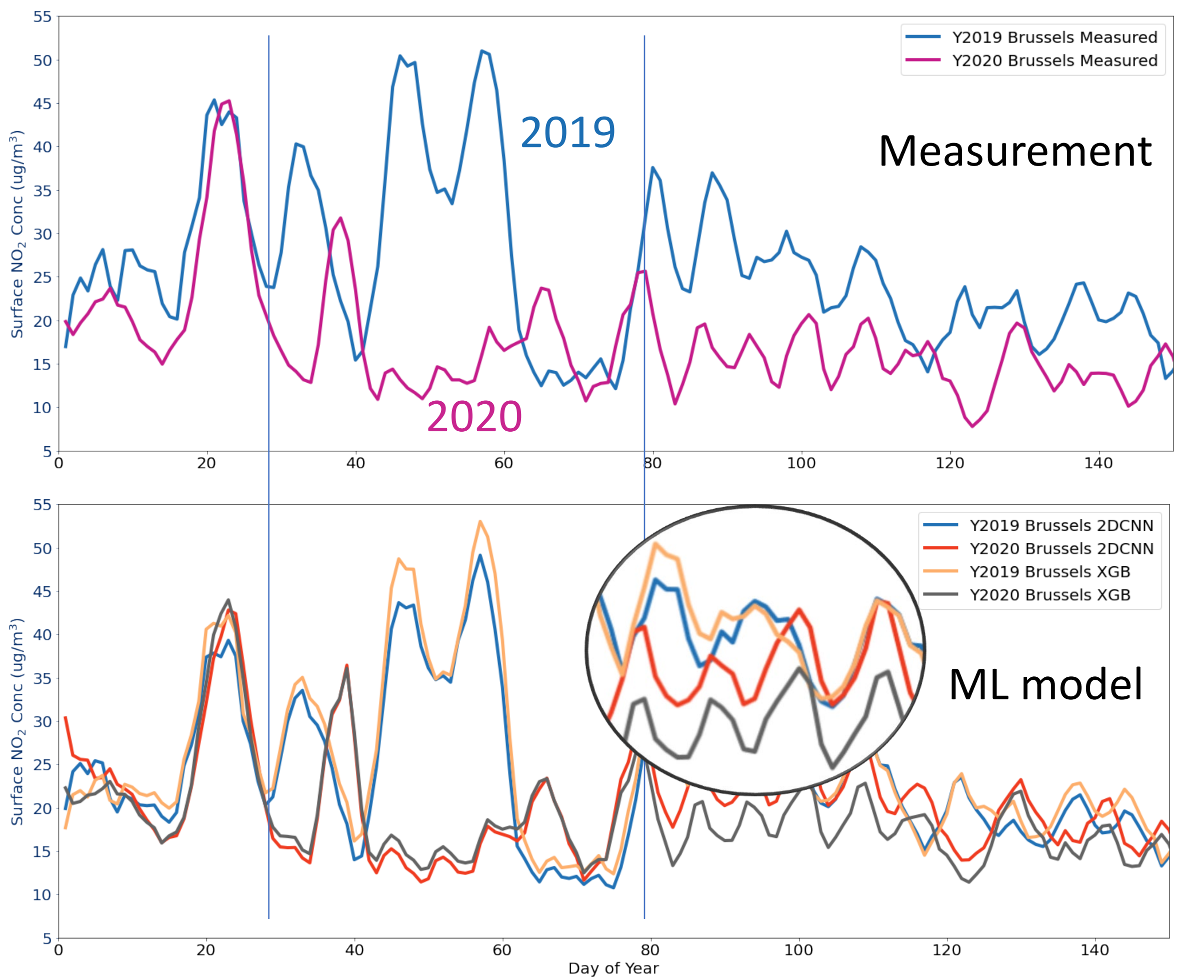
R^2: 0.68
Pearr: 0.82
RMSE: 6.14
MAE: 4.47

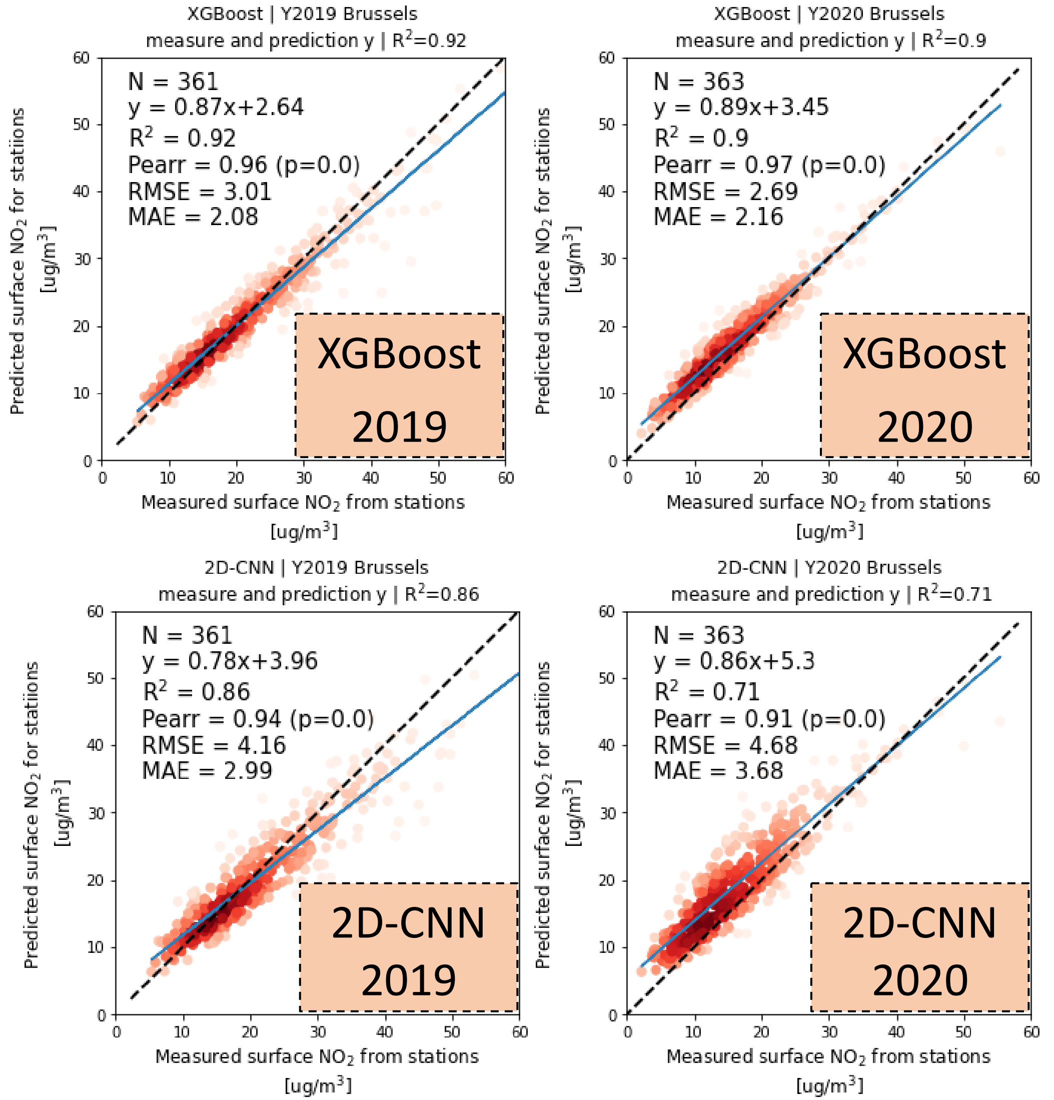
The 2D-CNN considering neighboring pixels within 5 Km provides a continuous NO_2 pattern and mitigates the artifacts to a large extent, but still sacrifices some accuracy.

Sun *et al.*, (2022, in preparation)

Compare XGBoost and 2D-CNN on time-series

Sun *et al.*, (2022, in preparation)





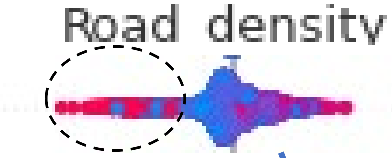
- 2D-CNN and XGBoost can identify the dramatic change in NO_2 during the Covid-outbreak.
- XGBoost performs better in time series since 2D-CNN becomes conservative after Covid and cannot well capture the difference between two years.

BIRA-IASB aeronomie.be ULB UNIVERSITÉ LIBRE DE BRUXELLES

Issue in NO_2 mapping by 2D-CNN

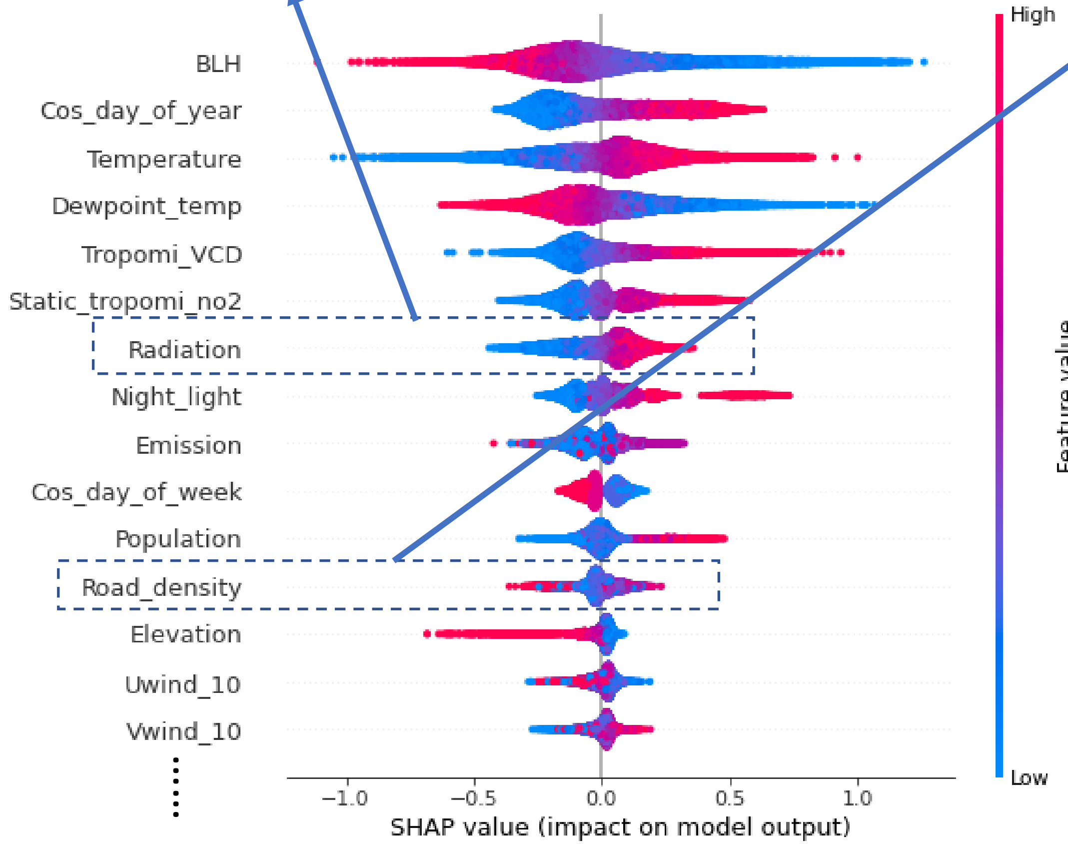
- Involving as many influential predictors as possible can improve the model's accuracy.
- However, the auto-correlation between variables may cause the model to capture a wrong relationship.

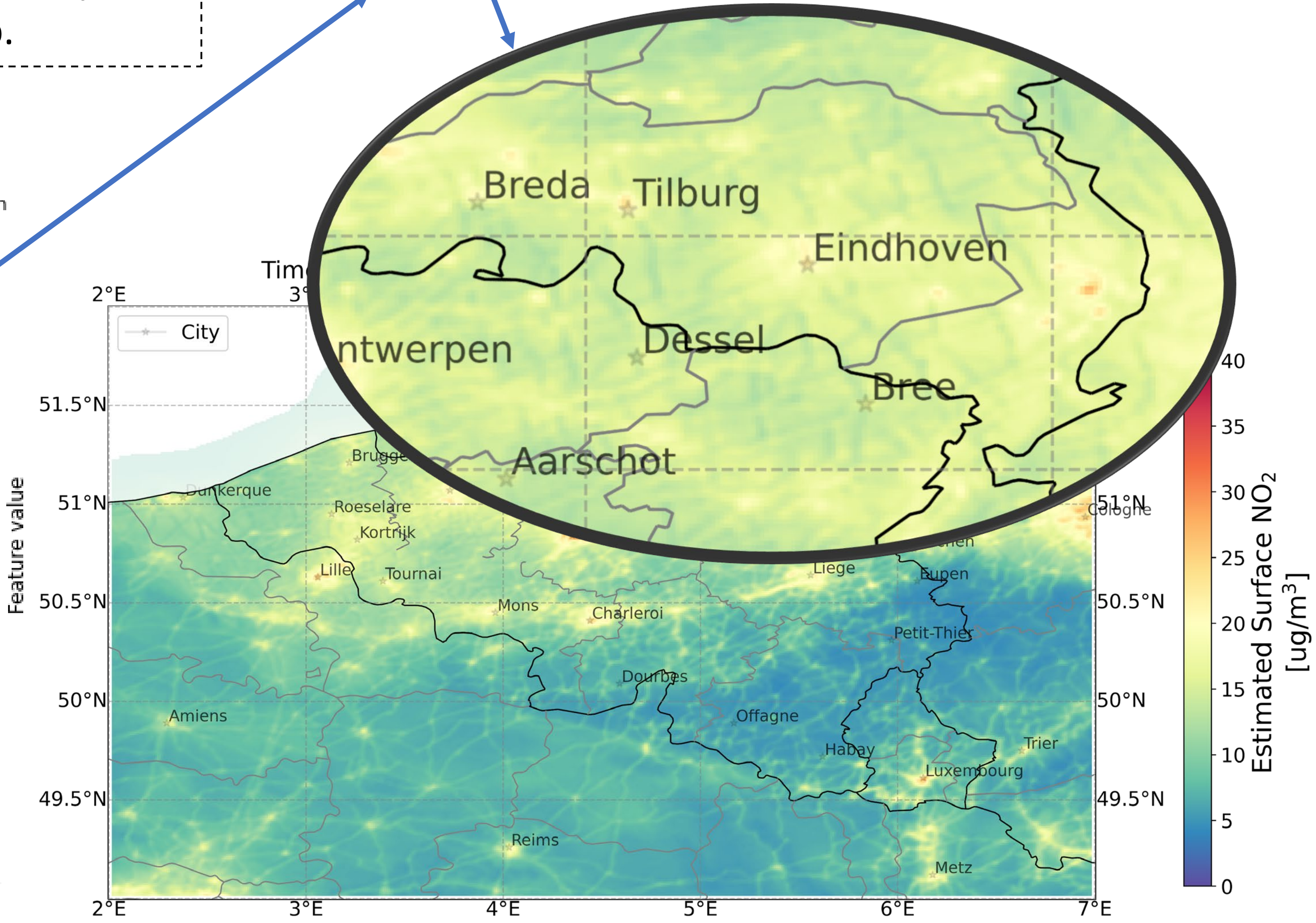
Sometimes high road density reduces NO_2 ?



Traffic volume might need to be included

High radiation increases NO_2





Sun *et al.*, (2022, in preparation)

Conclusion

- Use machine learning models to map surface NO_2 should consider both accuracy and spatial pattern.
- The 2D-CNN model provides surface NO_2 mapping with continuous patterns and fewer artifacts.
- In comparison to XGBoost, 2D-CNN performs conservatively and loses some accuracy.
- Auto-correlation within variables may cause the model to capture the wrong relationship and manufacture other possible artifacts.
- Given the trade-off between predictive accuracy and plausible spatial pattern, the ensemble of different models could be an option.

Contact: Wenfu Sun
wenfu.sun@aeronomie.be

Thank You