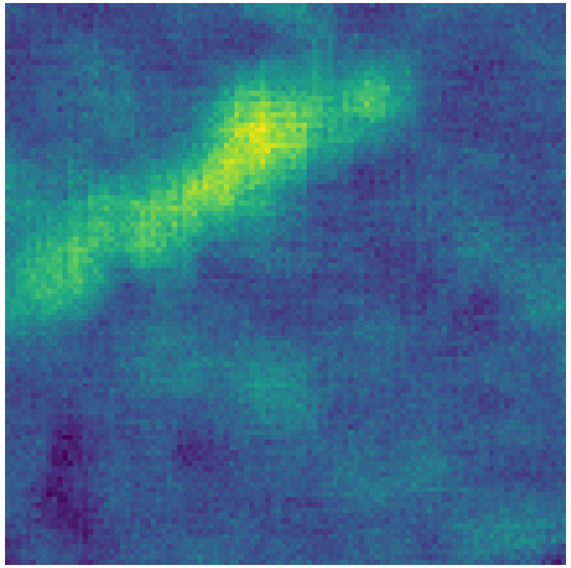
Universität St.Gallen

Estimation of Air Pollution with Remote Sensing Data: Revealing Greenhouse Gas Emissions from Space

Linus Scheibenreif
Michael Mommert
Damian Borth

University of St. Gallen



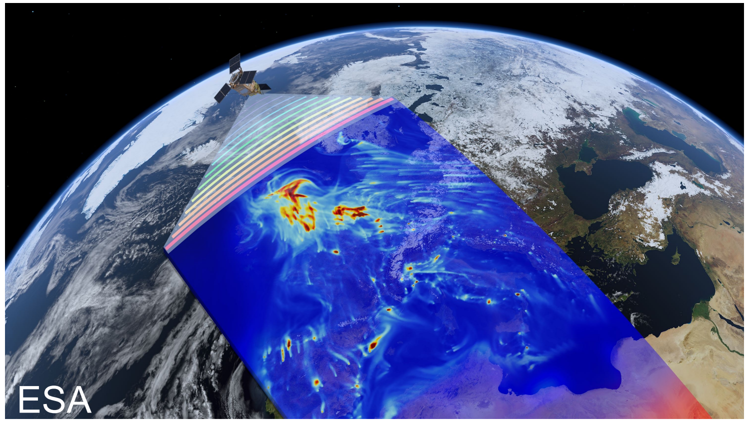
Tackling Climate Change with Machine Learning workshop at ICML 2021
July 23, 2021

"From insight to impact"

Universität St.Gallen

Introduction

Air pollution and the emission of **GHGs** are the main cause of climate change. Anthropogenic GHG emissions from the combustion of **fossil fuels** in industrial plants or vehicles are harmful to the environment and contribute to global warming trends. Besides the primary GHG, **CO_2**, the burning of fossil fuels also emits molecules like **NO_2** and CO, which can be used as **proxies** for the estimation of CO_2 emissions [Konovalov, 2016].



Continual data on air pollution concentrations in the atmosphere are primarily collected with two approaches:

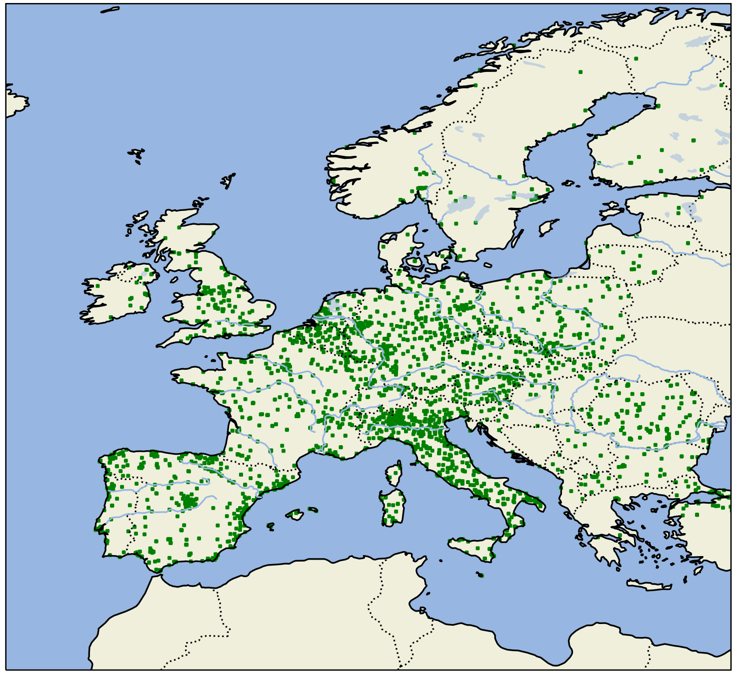
- Networks of **air quality stations** on the ground recording pollutant concentrations at select locations
- **Satellites** with spectrometers measuring atmospheric column densities of pollutants

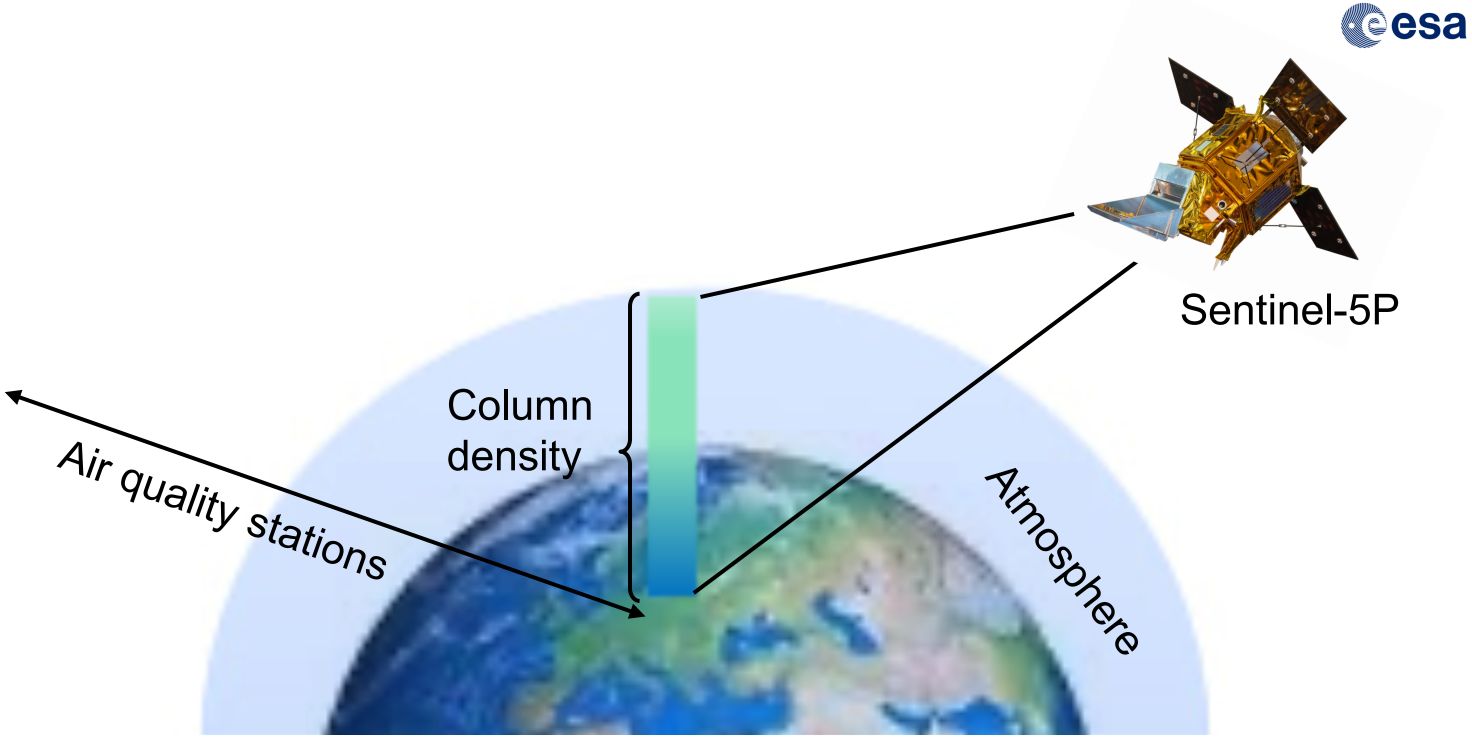
Data Limitations

Air pollution measurements from ground stations provide frequent measurements but **lack spatial coverage**. Satellites provide large spatial coverage but **low spatial resolution** and little information about a pollutant's vertical distribution.

Estimation of pollutant concentrations **near the surface**, where they originate, is a **non-trivial** task.

European Environment Agency





Universität St.Gallen

Data & Objectives

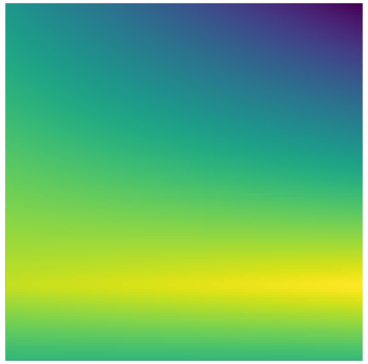
In previous work, approaches like **land-use-regression** (LUR) and geostatistical interpolation methods like **kriging** have been used to derive detailed information about the spatial distribution of air-borne pollutants at the **surface level**. These techniques are limited by the availability of a dense network of air quality stations for interpolation, or large datasets of auxiliary variables such as population statistics or road network data (see [Hoek, 2008] for a review).

European Environment Agency

This work utilizes temporal surface NO_2 measurements from 3000+ **air quality stations** across Europe, averaged for the 2018-2020 timeframe. Additionally, multi-band **remote sensing** data from the ESA Sentinel-2 satellite as well as tropospheric NO_2 column density values from Sentinel-5P are collected at the locations of air quality stations.

By leveraging globally available remote sensing data and deep learning in lieu of detailed, country specific input datasets, we strive to enable the estimation of surface level air pollutants at **high spatial resolution** for **any location on Earth**.

Sentinel-2



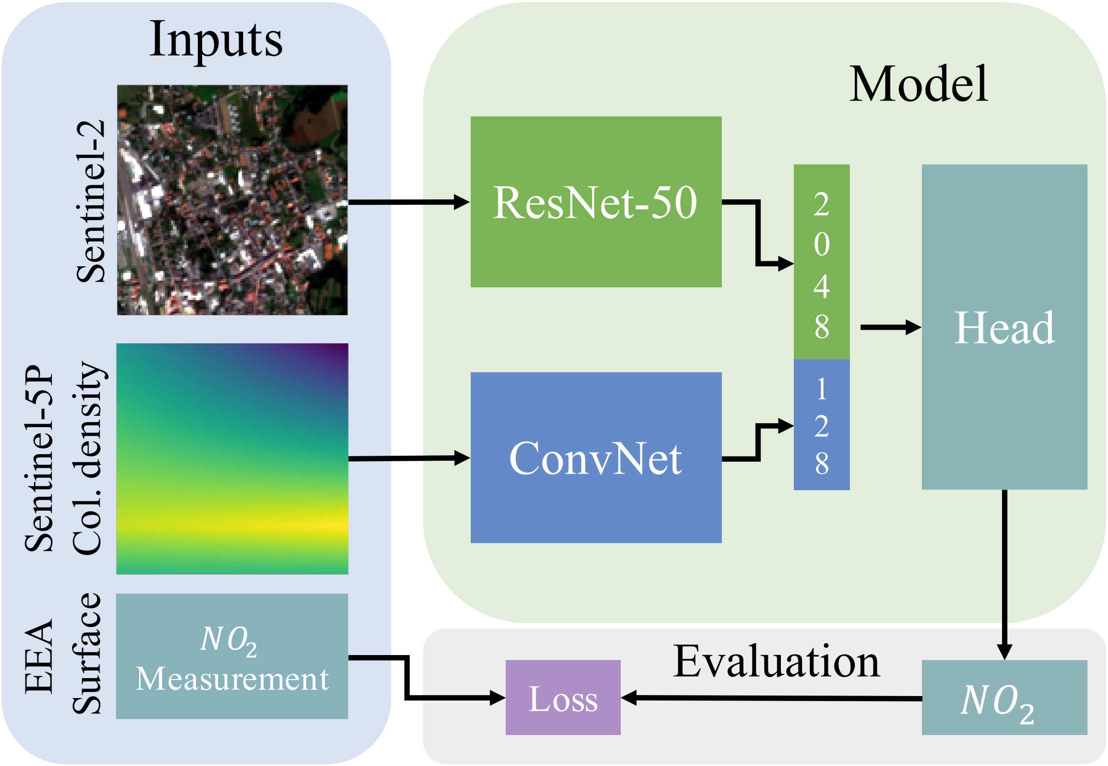
Sentinel-5P

Methods

We present a supervised **deep learning** approach for the prediction of surface NO2 concentrations from Sentinel-2 and Sentinel-5P data.

Features of the **Sentinel-2** image are extracted through a **ResNet-50** [He, 2016], with input layer adapted to the 13-band input data and **pretrained** on a land-use-classification task on the BigEarthNet dataset [Sumbul, 2019].

To account for the lower native resolution and single band nature of **Sentinel-5P** data, this input is separately processed through a **small CNN** before fusing with the Sentinel-2 image features.



The final prediction is produced by the “head”, a stack of 2 fully connected layers

Universität St.Gallen

Experiments

We experiment with different combinations of input data sources and aggregation frequencies. The best performance is obtained from the **combination of Sentinel-2 and Sentinel-5P** inputs and the target aggregated across the entire timeframe of our dataset, reaching an average **R2-score of** 0.54 ± 0.04 across 10 runs.

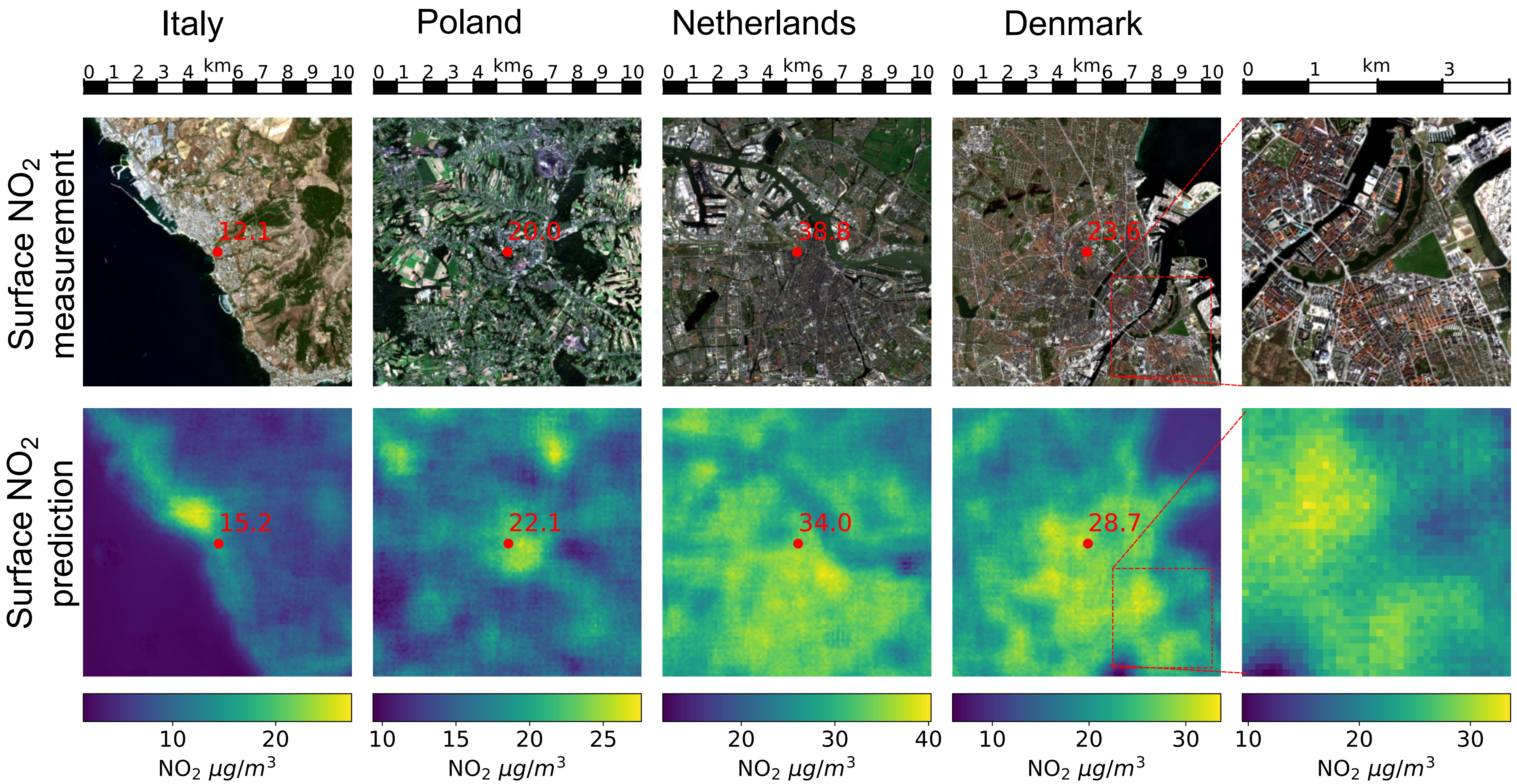
Interestingly, the model almost maintains this level of accuracy when predicting NO_2 concentrations at vastly **increased temporal frequency**. For **monthly** predictions, **an R2-score of** 0.51 ± 0.01 is achieved. This is partly explained by the increased number of available training observations at monthly frequency.

DATA	TIME	N-OBS.	PT	R2	R2-T10	MAE	MAE-T10	MSE	MSE-T10
SEN.-2	2018-20	3.2K	×	0.25±0.05	0.28	8.06±0.49	7.31	105.7±10.29	91.72
SEN.-2	2018-20	3.2K	✓	0.45±0.03	0.49	6.62±0.17	6.23	77.03±3.64	65.81
SEN.-2,5P	2018-20	3.1K	×	0.38±0.03	0.43	7.06±0.35	6.68	83.72±4.14	78.4
SEN.-2,5P	2018-20	3.1K	✓	**0.54±0.04**	**0.59**	**5.92±0.44**	**5.42**	**62.52±5.47**	**56.28**
SEN.-2,5P	QUART.	19.6K	✓	0.52±0.05	0.57	6.24±0.22	5.98	73.1±6.88	66.12
SEN.-2,5P	MONTH.	59.6K	✓	0.51±0.01	0.53	6.54±0.15	6.31	78.96±4.2	73.74

Results

Top row: Sentinel-2 images centered at EEA air quality stations. Average NO_2 measurement in red.

Bottom row: Corresponding NO_2 estimates. Estimated concentration at the air quality station in red.

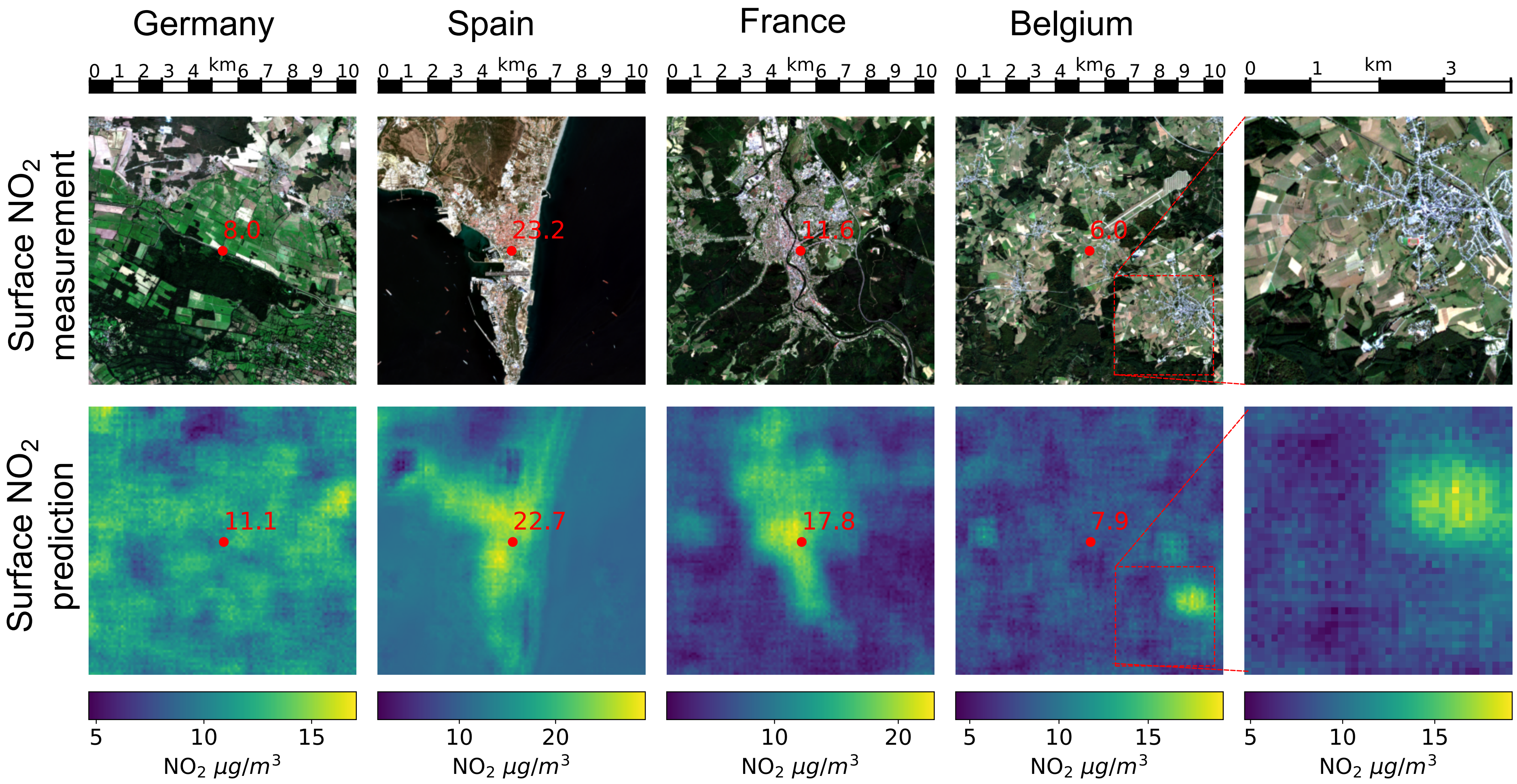


The heatmaps are produced from individual predictions for overlapping tiles of the top image and corresponding Sentinel-5P data.

Results

NO$_2$ estimates correspond well with known differences in surface NO$_2$ concentrations between built up and natural areas.

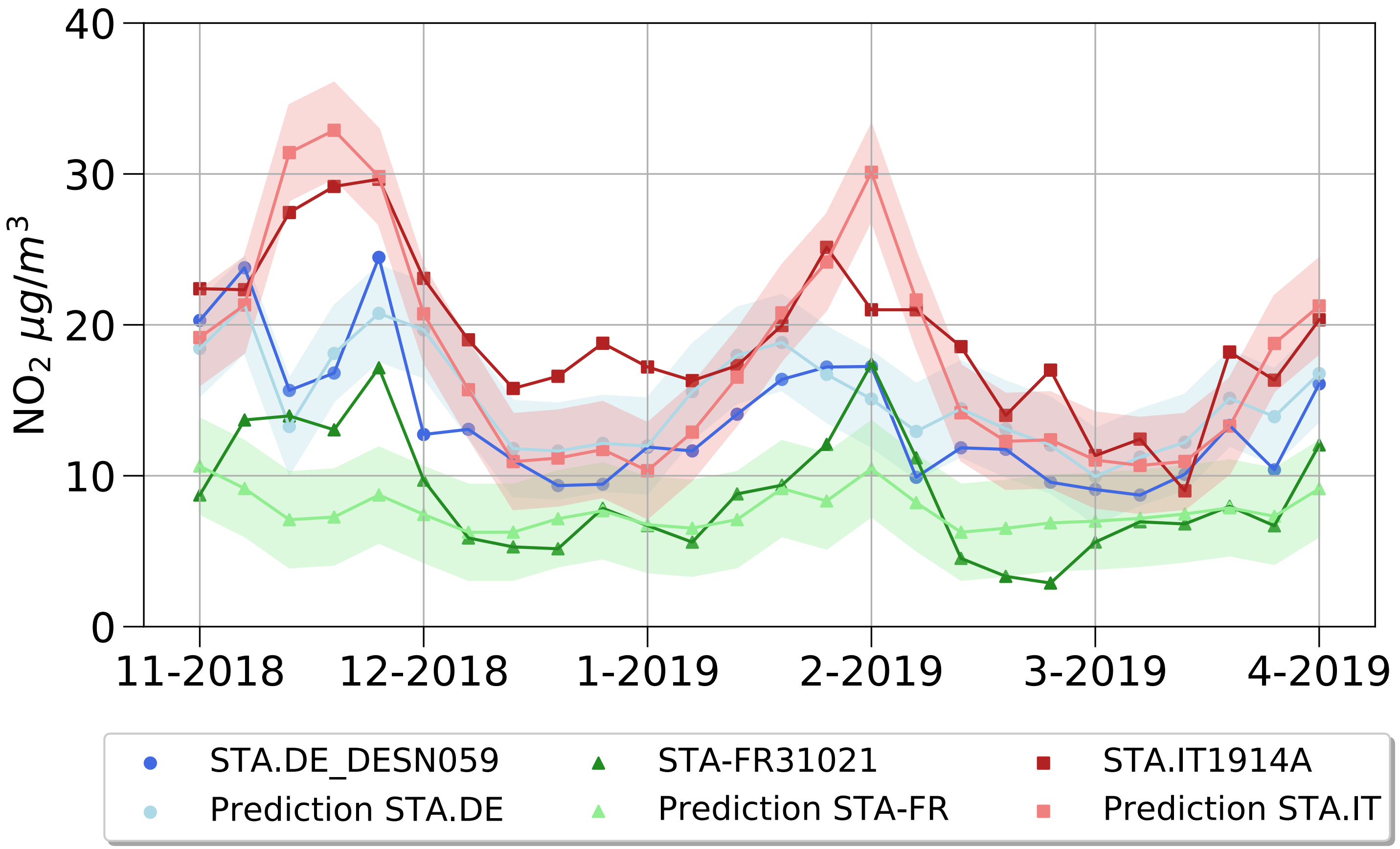
Estimation model is robust in different geographic environments across Europe.



Please see appendix for additional examples.

Temporal Prediction

Monthly average NO_2 **measurements** from three EEA air quality stations in Germany, France, and Italy (dark colors) and monthly NO_2 **predictions** based on our approach at the same locations (not seen during training). The shaded area indicates the model's MAE envelope centered at the nominal predictions.



Conclusions

We present a novel approach for the prediction of ambient **NO_2** concentrations based on **deep learning**, solely from **remote sensing** data.

- Accurate NO_2 estimates (MAE $< 6\mu g/m^3$).
- Applicable at any location on Earth.
- Capable of modelling temporal patterns of surface NO_2 concentration.

Artificial Intelligence
Machine Learning

Universität St.Gallen

linus.scheibenreif@unisg.ch

hsg.ai
Code

References

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Hoek, Gerard, et al. "A review of land-use regression models to assess spatial variation of outdoor air pollution." *Atmospheric environment* 42.33 (2008): 7561-7578.

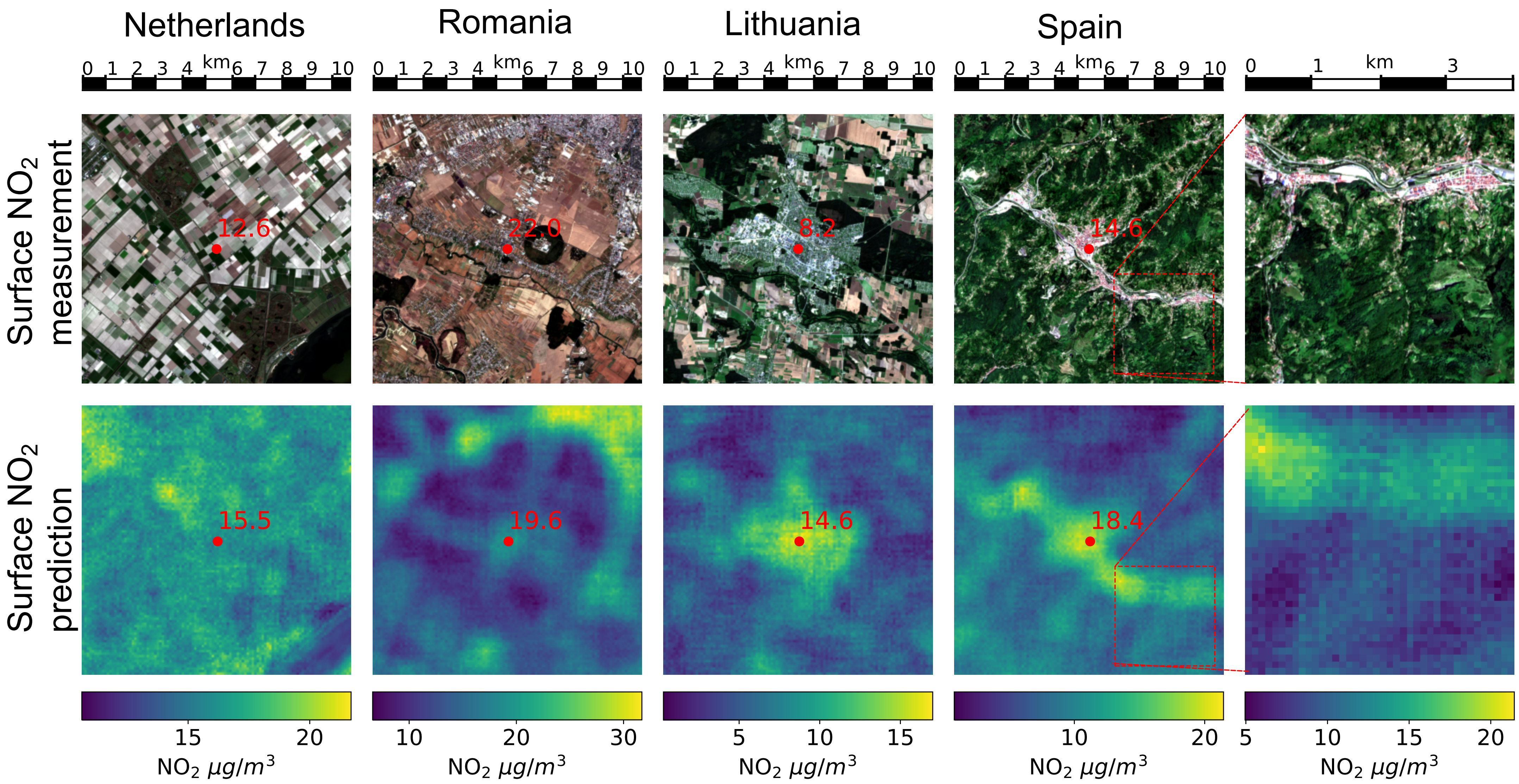
Kefauver, Shawn C., Iolanda Filella, and Josep Peñuelas. "Remote sensing of atmospheric biogenic volatile organic compounds (BVOCs) via satellite-based formaldehyde vertical column assessments." *International journal of remote sensing*35.21 (2014): 7519-7542.

Konovalov, Igor B., et al. "Estimation of fossil-fuel CO 2 emissions using satellite measurements of" proxy" species." *Atmospheric Chemistry and Physics* 16.21 (2016): 13509-13540.

Sumbul, Gencer, et al. "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding." *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019.

Appendix

Results



Netherlands
Romania
Lithuania
Spain
0 1 2 3 4 km 6 7 8 9 10
0 1 km 3
Surface NO_2 measurement
Surface NO_2 prediction
12.6
22.0
8.2
14.6
15.5
19.6
14.6
18.4
15 20
10 20 30
5 10 15
10 20
5 10 15 20
NO_2 $\mu g/m^3$