MURMURATION

24/05/2022

EARTH MEGAPIXEL

High resolution air quality measurements: from city to neighborhood level, combining Sentinel-5P, CAMS models and in-situ measurements

Tarek Habib

Camille Lainé

TABLE OF CONTENTS

#1 MURMURATION

Satellite data enabling the monitoring of environmental impact

1.1 Murmuration's indicators

WATER
- Water quality
- Aquatic coverage
- Water stress

AIR
- Air quality
- CO_2 emissions
- Carbon sequestration

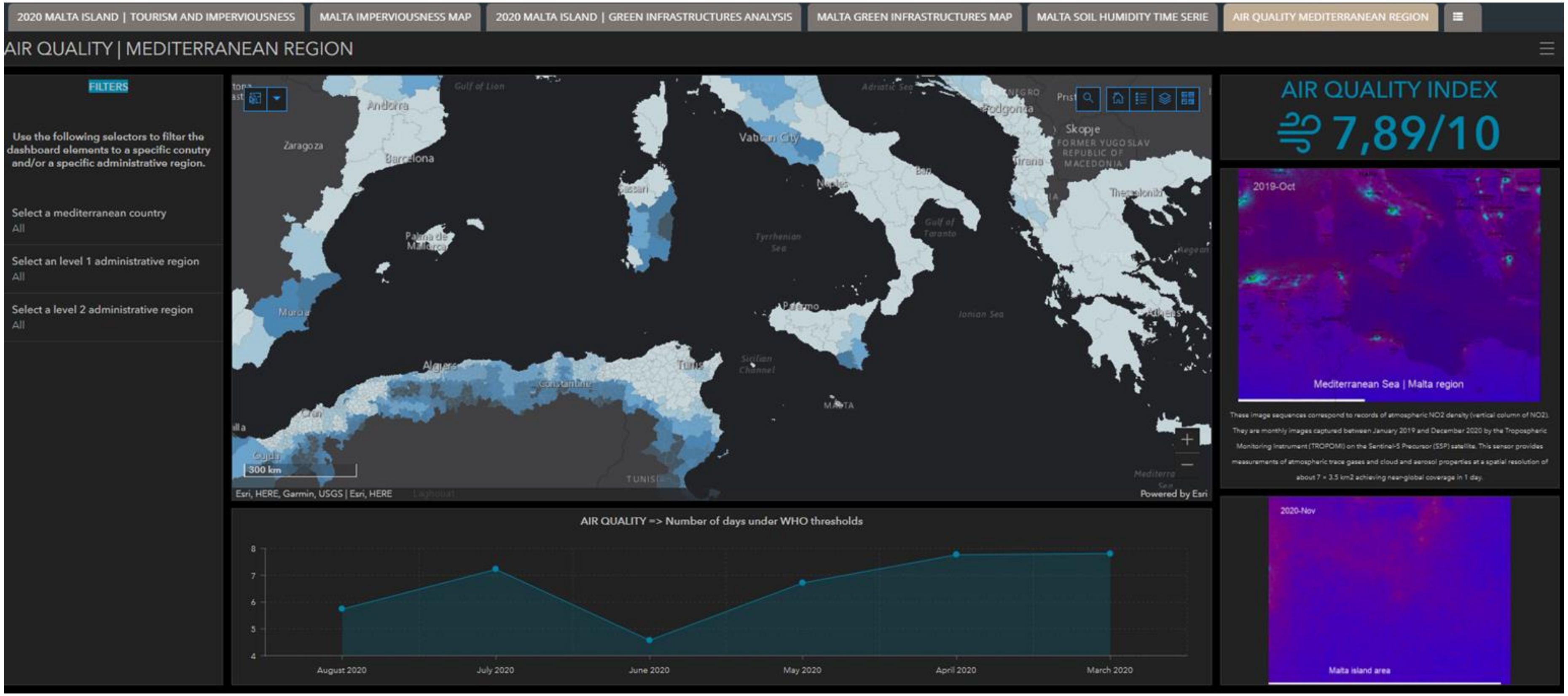
BIODIVERSITY
- Vegetation
- Forests
- Protected areas
- Presence of endangered species

INFRASTRUCTURES
- Land cover
- Urbanization
- Urban heat
- Coastal erosion

TOURISM STATISTICS
- Tourist presence
- Touristic infrastructures

1.2 Air quality monitoring

1.2.1 Dashboards

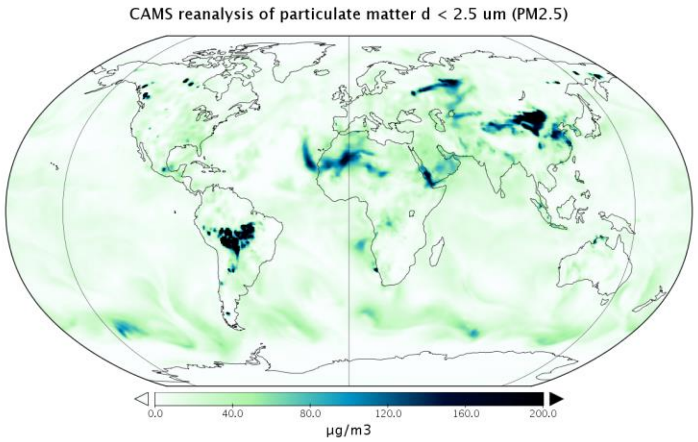


#2 PROBLEM

2.1 Trade-off between spatial coverage and resolution

Satellite data :

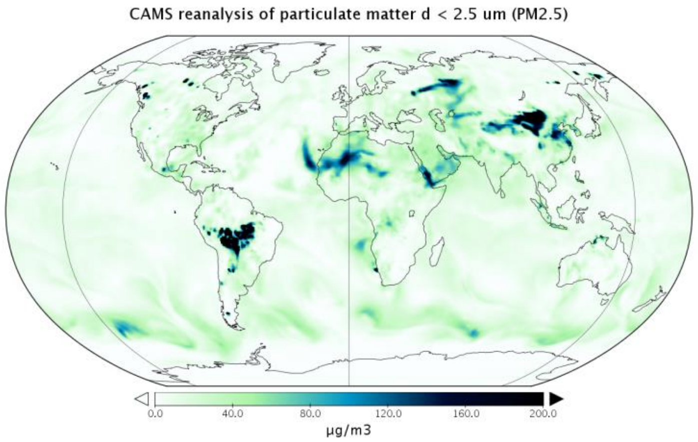
- Coverage
- Resolution



2.1 Trade-off between spatial coverage and resolution

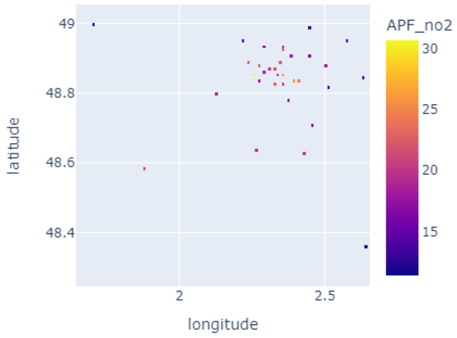
Satellite data :

- (+) Coverage
- (−) Resolution



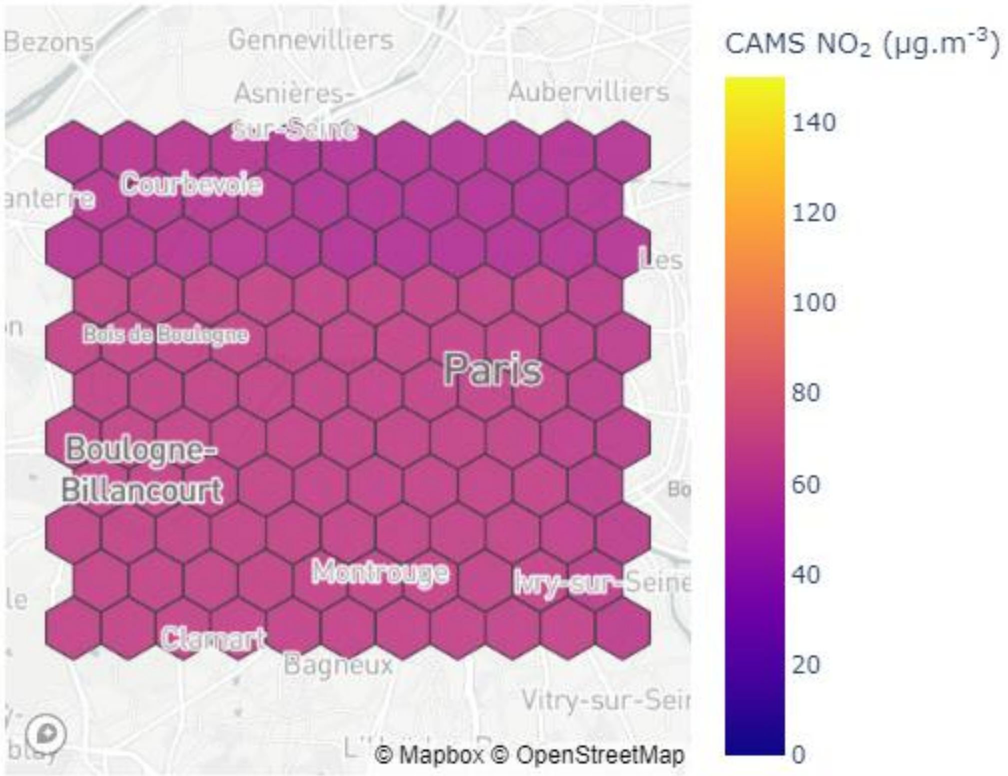
In-situ sensor data :

- (−) Coverage
- (+) Resolution

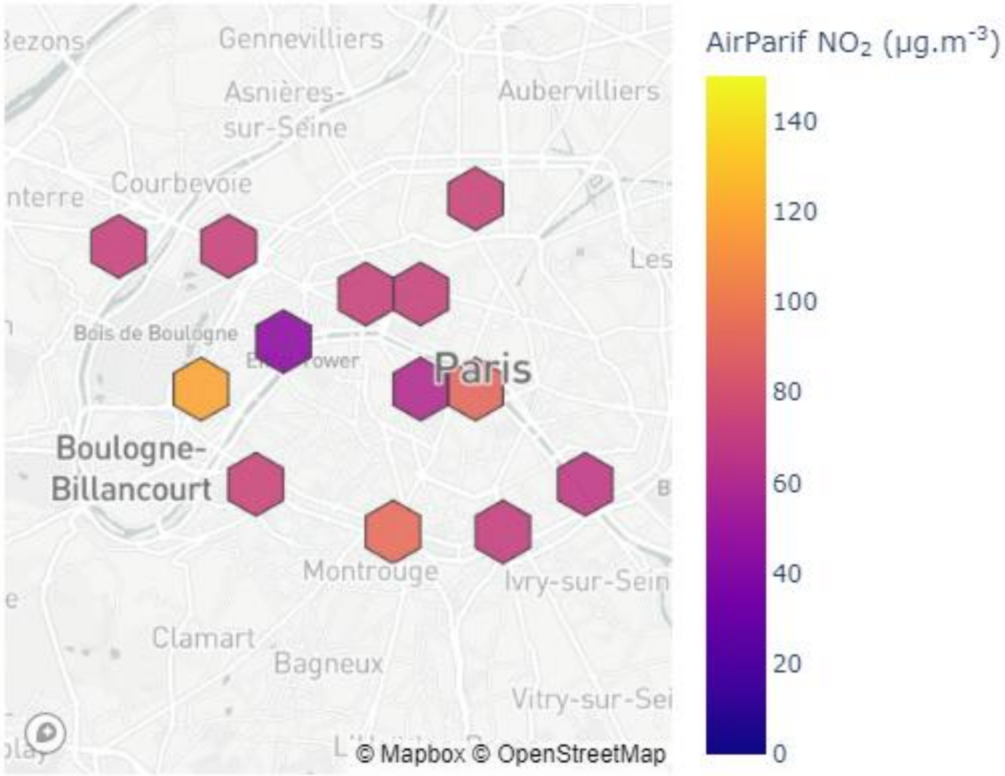


2.1 Trade-off between spatial coverage and resolution

Satellite data



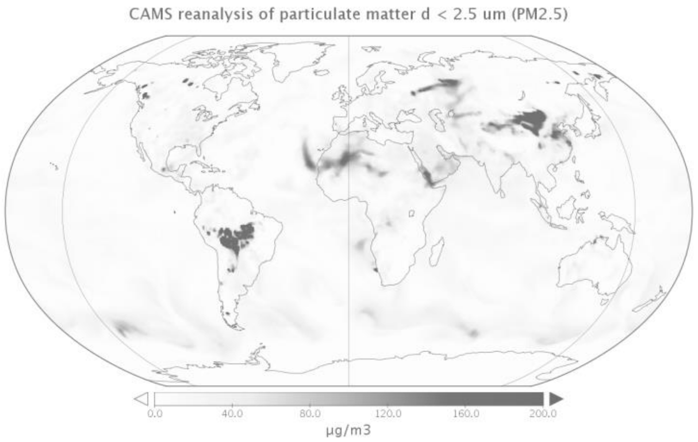
In-situ data



2.1 Trade-off between spatial coverage and resolution

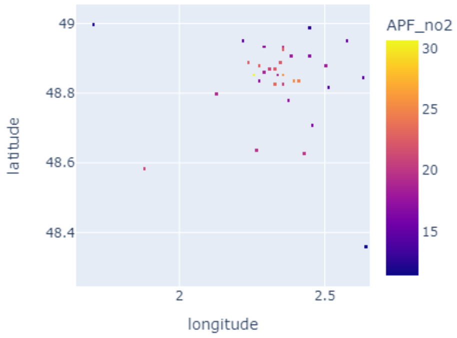
Satellite data :

- (+) Coverage
- (−) Resolution



In-situ sensor data :

- (−) Coverage
- (+) Resolution

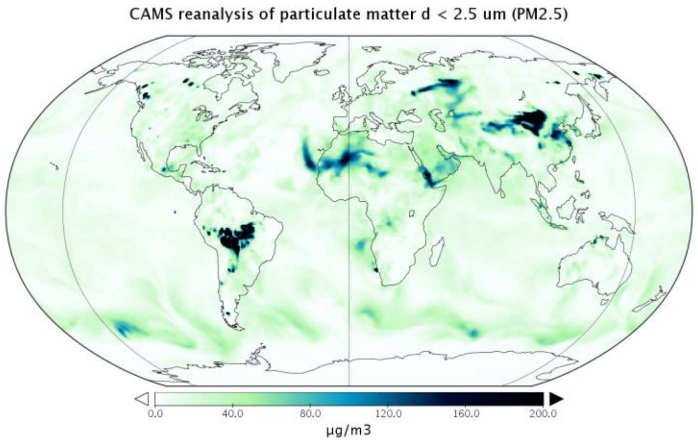


Air quality actors develop regional models from in-situ sensors with good accuracy, but these models are not applicable to regions that are not equipped with in-situ sensors.

2.1 Trade-off between spatial coverage and resolution

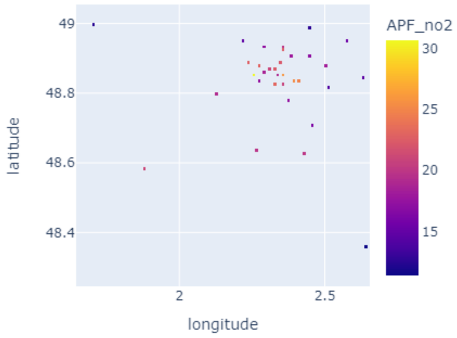
Satellite data :

- Coverage (+)
- Resolution (−)



In-situ sensor data :

- Coverage (−)
- Resolution (+)



Need :
- Global coverage
- High resolution

#3
STATE OF THE ART

3.1 Kim, Brunner and Kuhlman

Remote Sensing of Environment 264 (2021) 112573

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Importance of satellite observations for high-resolution mapping of near-surface NO_2 by machine learning

Minsu Kim *, Dominik Brunner, Gerrit Kuhlmann

Empa, Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland

A R T I C L E I N F O

Edited by: Menghua Wang

Keywords:
Nitrogen dioxides (NO_2)
TROPOspheric Monitoring Instrument (TROPOMI)
Sentinel-5 precursor satellite
High resolution map of NO_2
Machine learning
Extreme gradient boosting (XGBoost)
Shapley additive exPlanations (SHAP)
COVID-19 pandemic lockdown

A B S T R A C T

Nitrogen dioxide (NO_2) is an important air pollutant with negative health effects and a precursor of ozone and particulate matter responsible for photo-chemical smog and wintertime air pollution. To evaluate human exposure to NO_2 for public health assessment, maps of near-surface NO_2 concentrations at a high resolution of 100 m are desirable. In this study, we report hourly maps of gridded near-surface NO_2 concentrations that are produced using an extreme gradient-boosted tree ensemble for an Alpine domain (Switzerland and northern Italy) spanning two years, from June 2018 to May 2020. To estimate the NO_2 distribution at ground level, we used satellite observations of NO_2 vertical column density, land use data, meteorological fields and topographical information to train models with *in situ* NO_2 ground measurements. The best model with this approach captured up to 59% of hourly NO_2 variation for 40 test stations in the domain with a mean absolute error of $7.69\mu g/m^3$, performing especially well for urban regions with dense sampling. We present the first hourly maps of NO_2 concentrations that reveal previously unresolved spatio-temporal variations. Local interpretations of the machine learning model demonstrate that TROPOMI NO_2 satellite observations make a strong contribution to the information content of the near-surface NO_2 maps besides their relatively coarse resolution (3.5×5.5 km^2) and the fact that they are only available once a day under cloud-free conditions. The COVID-19 pandemic lockdown presents a case study that offers new insights into the importance of satellite data that can partially re-mediate statistical models' unsusceptibility to unusual events (like changes due to political intervention) with regard to model training.

3.1.1 Methodology

Model:
XGBoost Regressor

Data:
- TROPOMI NO_2
- In situ NO_2 measurements
- Meteorological data
- Land use data
- Road length density
- Average traffic volume
- Industrial emissions
- Digital Elevation Model

3.1.2 Area of validation
North of Italy and Switzerland

3.2 Murmuration's implementation

3.2.1 Methodology (same basis)

Model:
XGBoost Regressor

Differences in data choice:
- CAMS NO_2 instead of TROPOMI NO_2
- Tomtom data for traffic statistics
- Population density data instead of Land use data

3.2.2 Area of validation
Île-De-France (region of Paris)

Goals

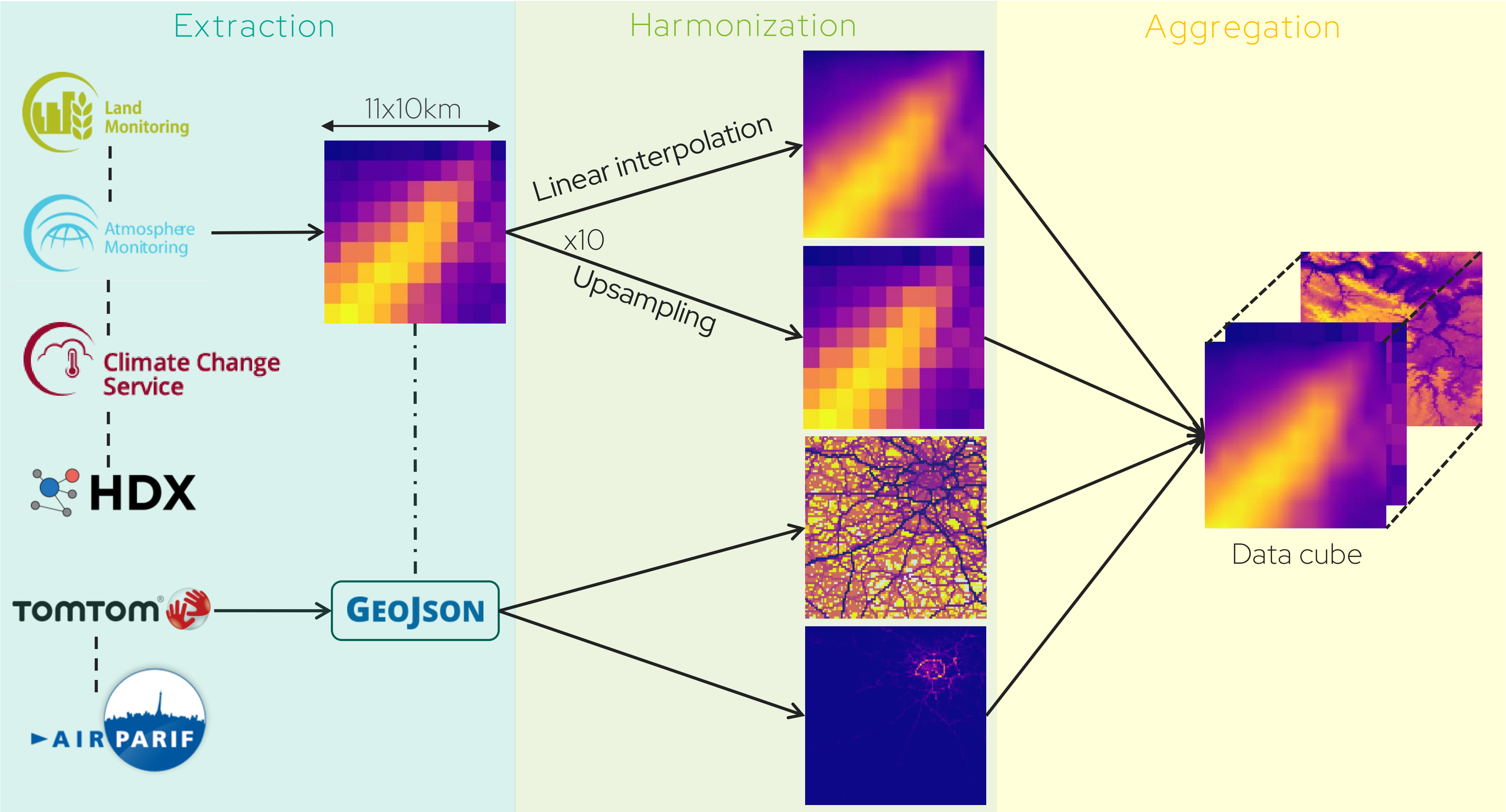
- The ability to generalize the algorithm on other locations
- The possibility to industrialize all over the world with accessible data

#4 DATA

4.1 Input data

	Data source	Variables	Temporal coverage	Spatial coverage	Temporal resolution	Spatial resolution	Size (1 month of data)
Model source data	**CAMS**	Satellite NO_2	2019 - 2022	Europe	Hourly	0.1° (~10 km)	61 MB
	ERA5	Precipitation Temperature Wind Boundary layer height	1950 - 2022	Global	Hourly	0.1° (~10 km)	138 MB
	Copernicus Land	Topography	2011	Europe	-	25 m	51 kB
	Humanitarian Data Exchange	Population density	2020	Global	-	1 km	48 KB
	Tomtom Statistics	Road traffic Road classes	2021/01	France	Hourly	In-situ	300 GB -> 47 MB
Model target data	**Airparif**	In-situ NO_2	2019 - 2021	Ile de France	Hourly	In-situ	55 MB

4.2 Data processing



#5 MODEL

5.1 XGBoost Regressor

XGBoost

5.1 XGBoost Regressor

Input

CAMS NO_2
Road traffic
Road class
Precipitations
Temperature
Boundary Layer Height
Wind
Weekday
Hour
Population density

744 samples (1 month)

→ XGBoost

Parameters :
Test size = 30%

5.1 XGBoost Regressor

Input

CAMS NO_2
Road traffic
Road class
Precipitations
Temperature
Boundary Layer Height
Wind
Weekday
Hour
Population density

744 samples (1 month)

→ XGBoost →

Parameters :
Test size = 30%

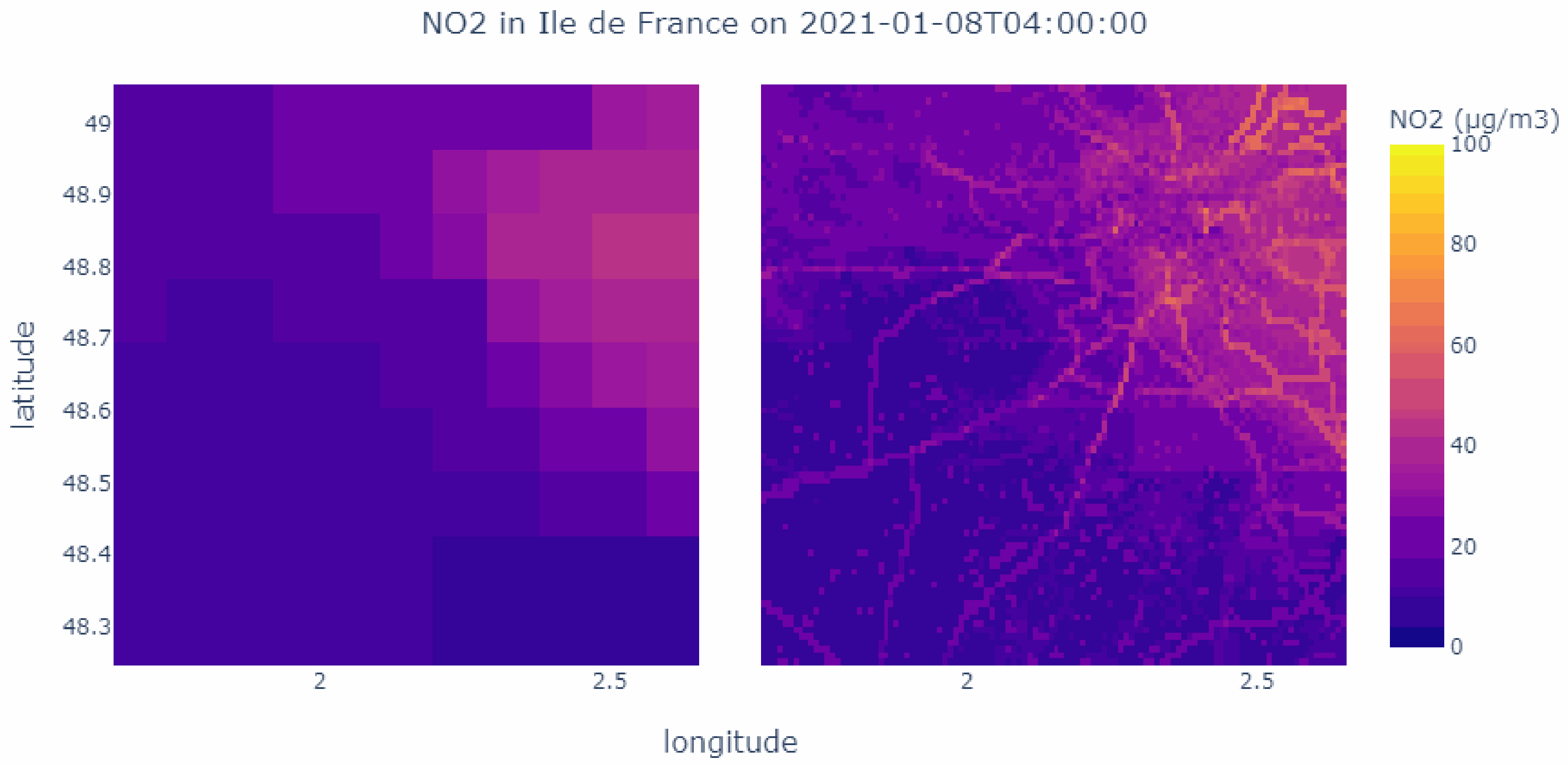
Output

NO_2 prediction

#6 RESULTS

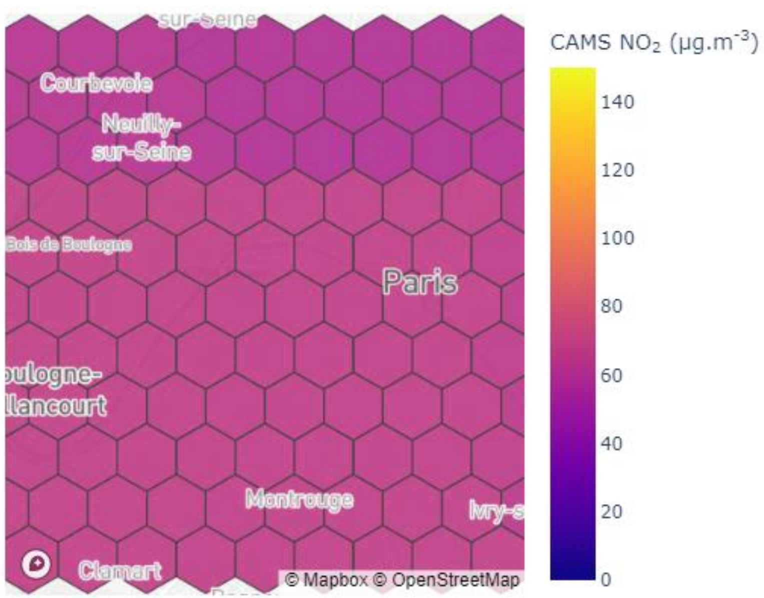
Predictions, evaluation and explainability

6.1 Predictions

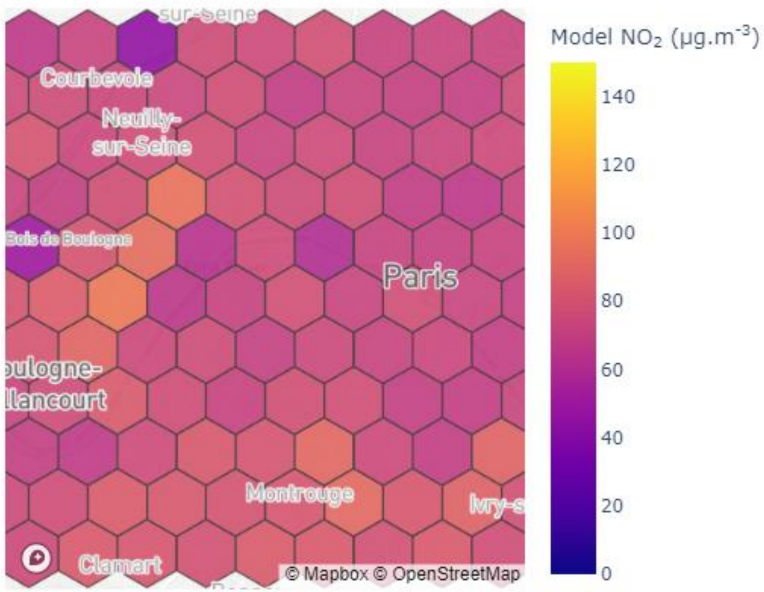


6.1 Predictions

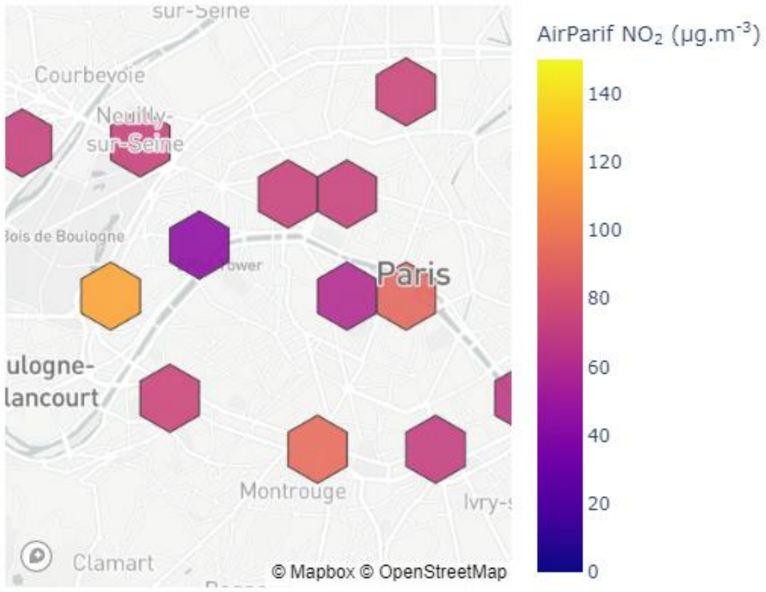
Satellite data



Model prediction

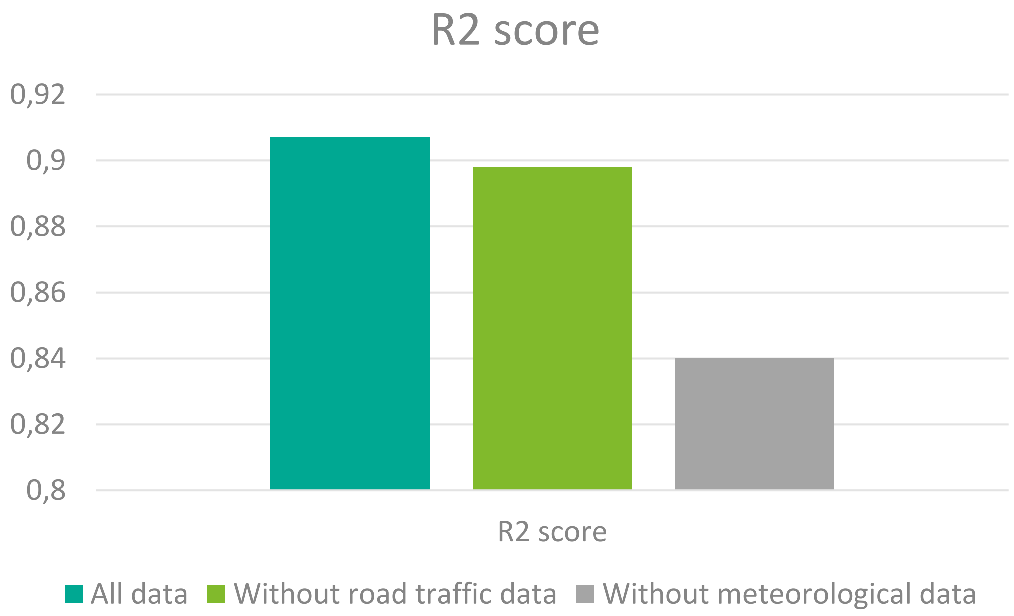


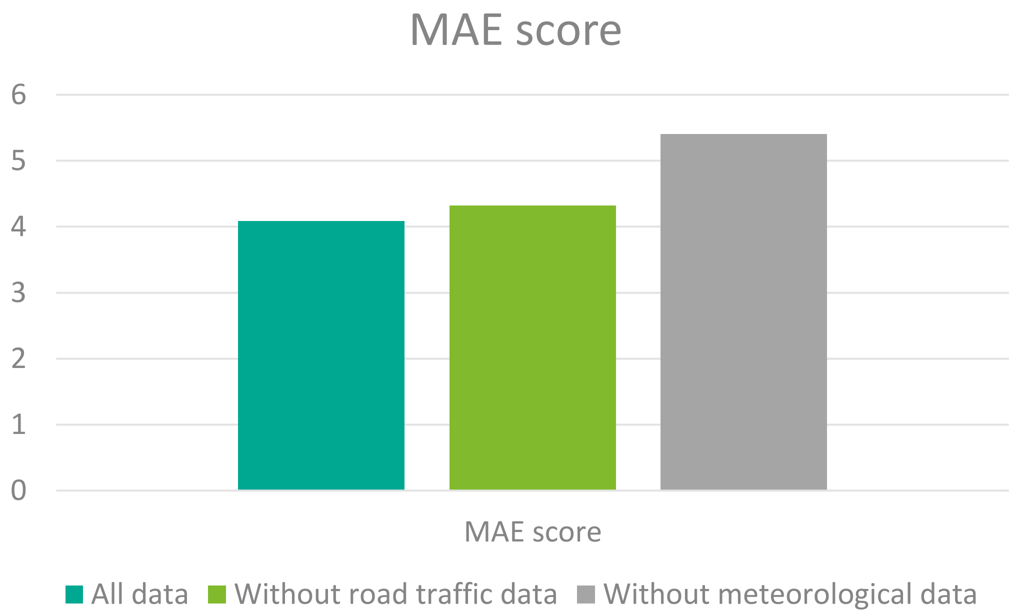
In-situ data



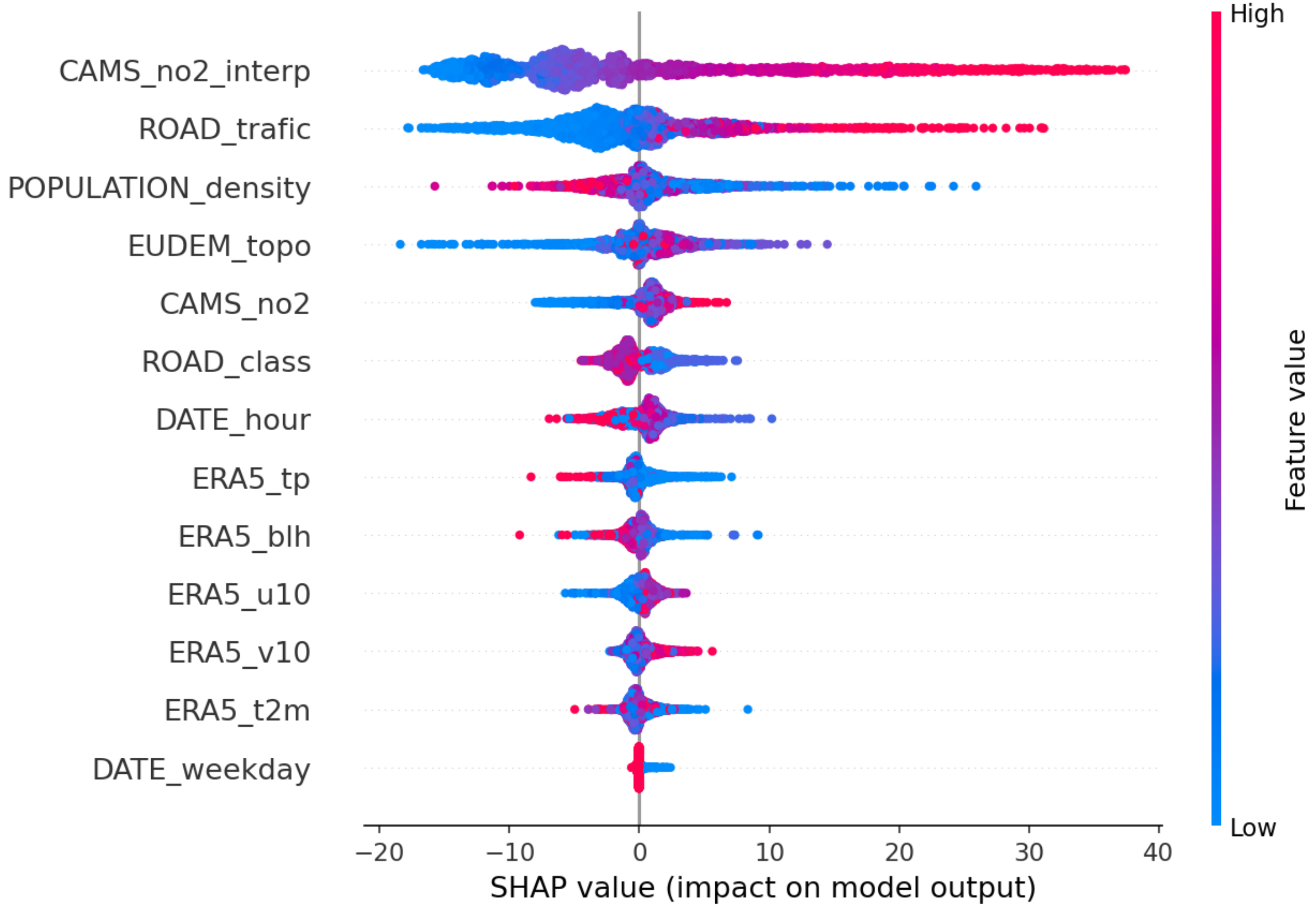
6.2 Model evaluation

$$\text{MAE} = 4.082\ \mu g.m^{-3}$$
$$R^2 = 0.907$$





6.3 Explainability: SHAP



#7
CONCLUSION

7.1 Conclusion and perspectives

Done

- High resolution air quality prediction on new location with accessible data

Next steps

- Add more road traffic data and check the scores
- Validate on other locations : Atmo data
- Create high resolution maps in areas without in-situ sensors

Fédération des associations de surveillance de la qualité de l'air
Atmo France

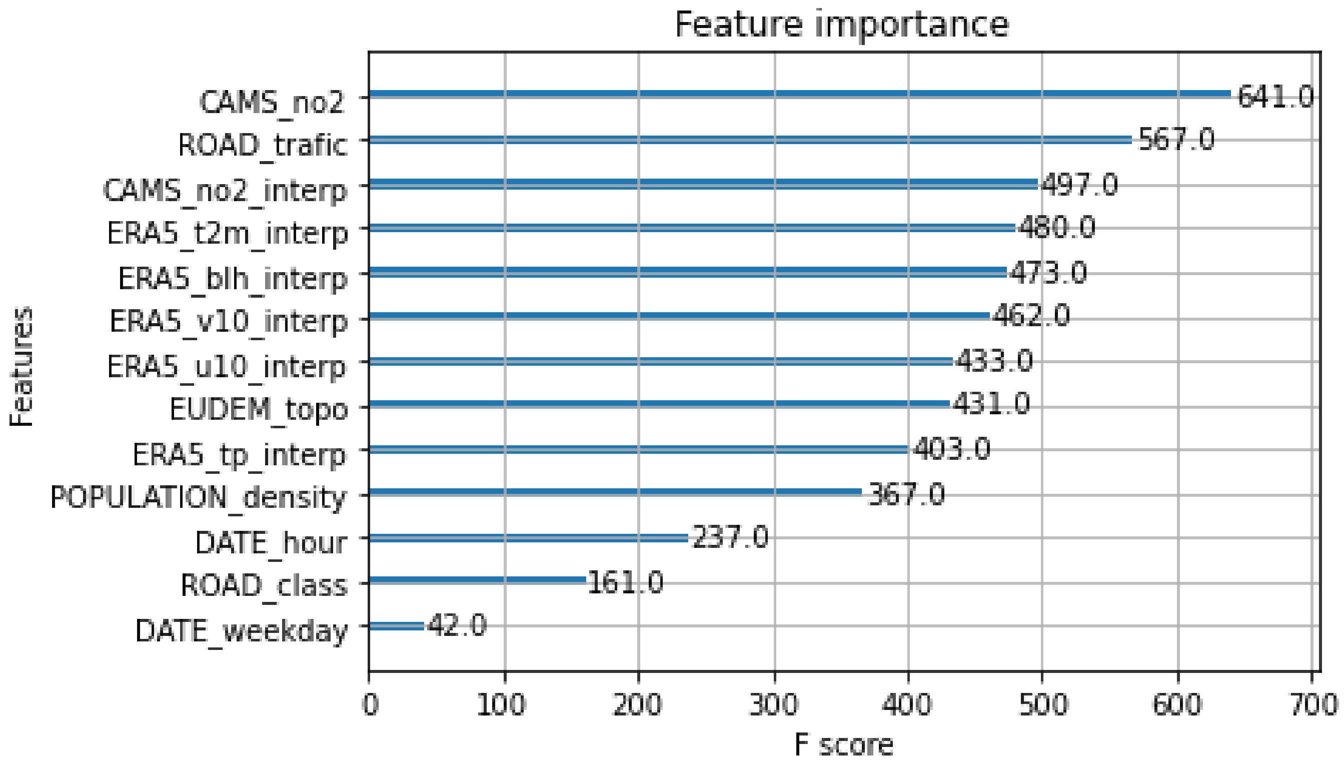
Contact us :

Camille Lainé
camille.laine@murmuration-sas.com

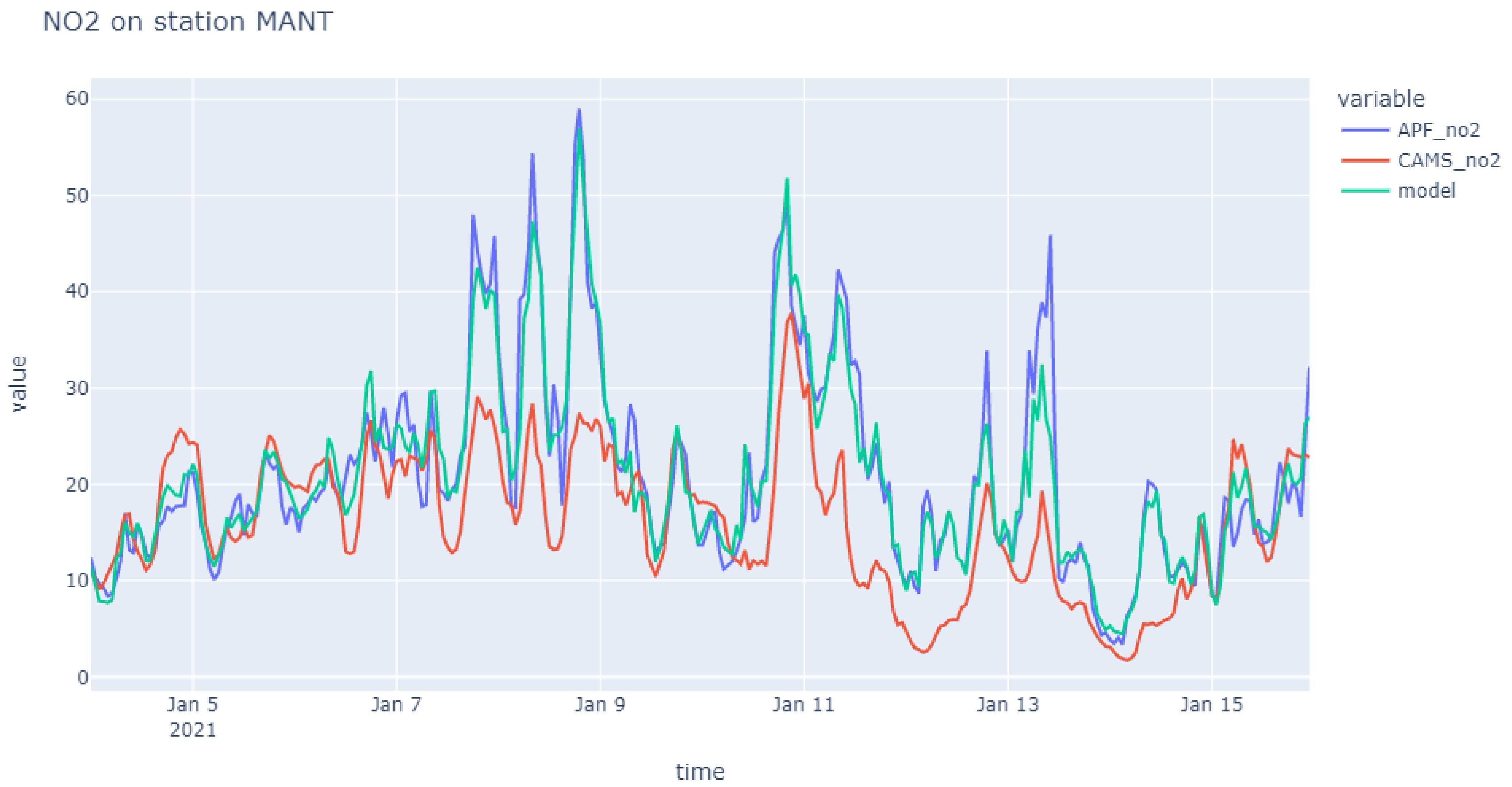
Tarek Habib
tarek.habib@murmuration-sas.com

MURMURATION – SAS
15 Rue Victor Hugo 31150 Bruguière
N° de SIRET : 848934497200029
N° de TVA : FR88848934972 - Toulouse

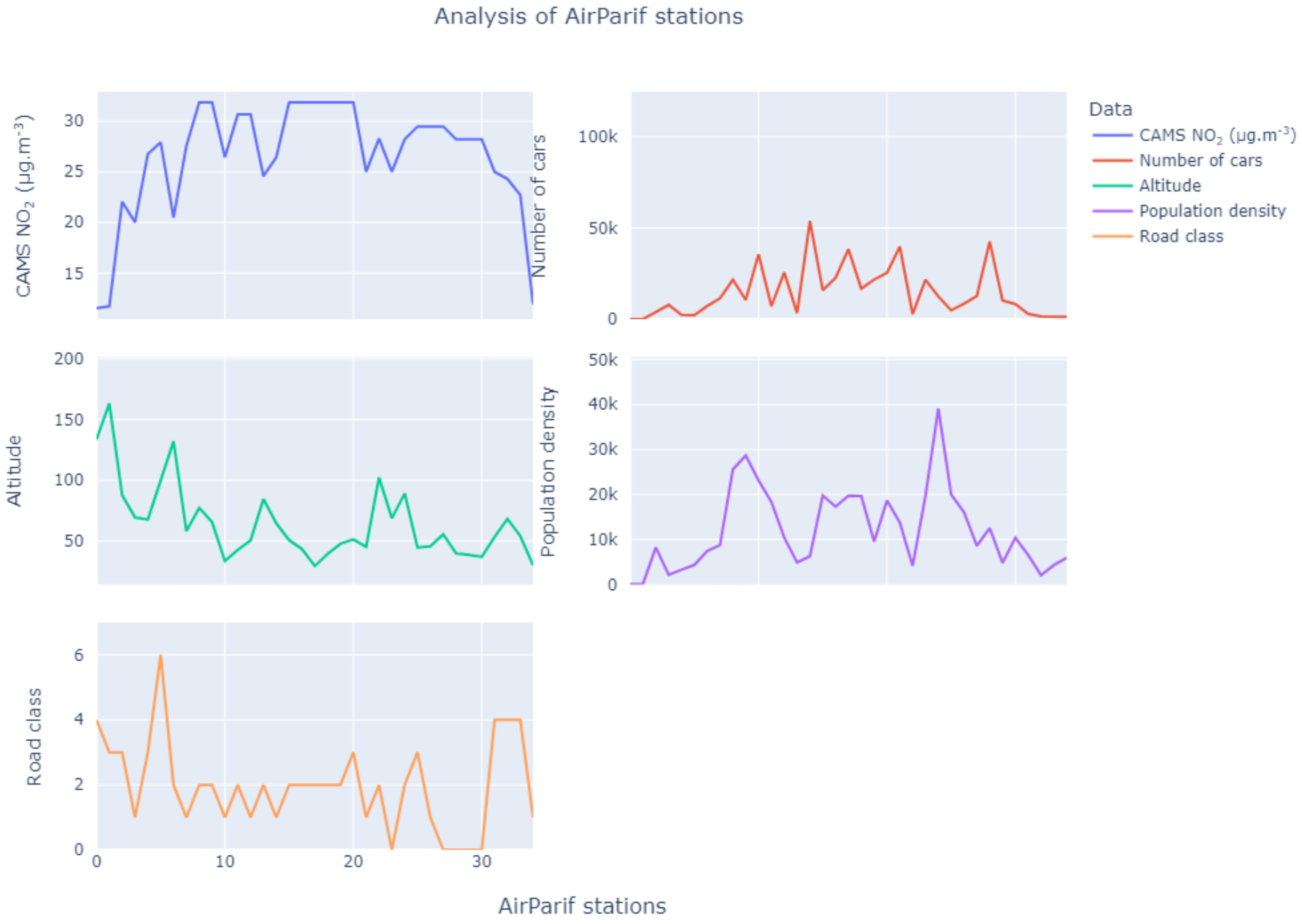
8.1 XGBoost feature importance



8.2 Prediction comparison in Airparif station



8.3 Analysis of Airparif stations' situation



8.4 Tomtom road classes

TOMTOM

Tomtom Functional Road Classes

0 : Motorway
1 : International road
2 : Major road
3 : Secondary road
4 : Connecting road
5 : Major local road
6 : Local road
7 : Minor local road