# MindEarth

# Downscaling Sentinel-5P $NO_2$ Data Using Gradient-Boosted Trees for High-Resolution Urban Air Quality Mapping

**Alessandro Austoni**, Francesco Asaro, Alessandra Feliciotti, Emanuele Strano, Mattia Marconcini

Living Planet Symposium 2025 - Vienna, Austria

---

MindEarth

www.mindearth.ai

MindEarth is a Swiss-Italian technology company specializing in geospatial technology, artificial intelligence, and Earth observation.

We collect and process complex geospatial data from satellites, sensors, and mobile sources to support decision-making across the public and private sectors.

## Mobility Data Analysis

Combination of large-scale, high-frequency location based data with POIs and road networks to obtain a dynamic population insights.

## Earth Observation data

Earth Observation data and advanced AI methods to develop customized services and solutions for emergency management, spatial planning, environmental monitoring and humanitarian aid.

## Mobile Mapping for Territorial Intelligence

Field surveys to collect high-frequency sequences of optical and sensor data. Extraction of valuable information on customer behavior, consumer goods, public spaces, and built environments.

**Alessandro Austoni** , Data scientist

This project is one of the four use cases of Destination Earth (DestinE), an initiative aimed at creating advanced digital models of the Earth to monitor and simulate natural phenomena and human activities.

An urban digital twin available on the DestinE DESP platform, which assesses environmental impacts resulting from changes to road networks, traffic systems, and urban services.

It combines Earth Observation data with empirical mobility data to create "what-if" scenarios and test sustainable mobility interventions and policies.

CITYNEXUS
Climate Resilient Urban Planning

# Why mapping $NO_2$?

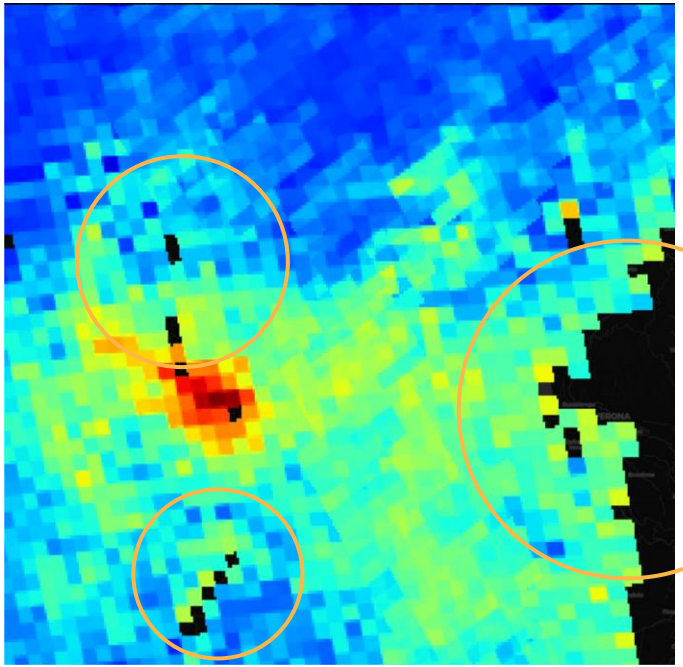
- Through everyday combustion activities nitrogen dioxide ($NO_2$) is released into the atmosphere.

  Once in the air, $NO_2$ reacts with sunlight, triggering a chemical transformation that produces ozone ($O_3$): a harmful pollutant when found near the ground.

  The pollutants affect us directly, contributing to both short and long-term health problems.

# Current state of $NO_2$ mapping

- **3.5 X 5 km** spatial resolution
- Has to go through the whole atmosphere
- Spatio-temporal **gaps**
- Many services only provide air quality **indexes**
- **Approximate** knowledge of air pollution
- Sum up all pollutants data from Sentinel-5p



Sentinel 5P, Milan (IT), missing values

## Provide high spatial resolution

Leverage AI techniques to create **maps at 100 m resolution**, enabling more accurate assessment of fine-scale pollution dynamics.

## Complement Sentinel 5p Availability

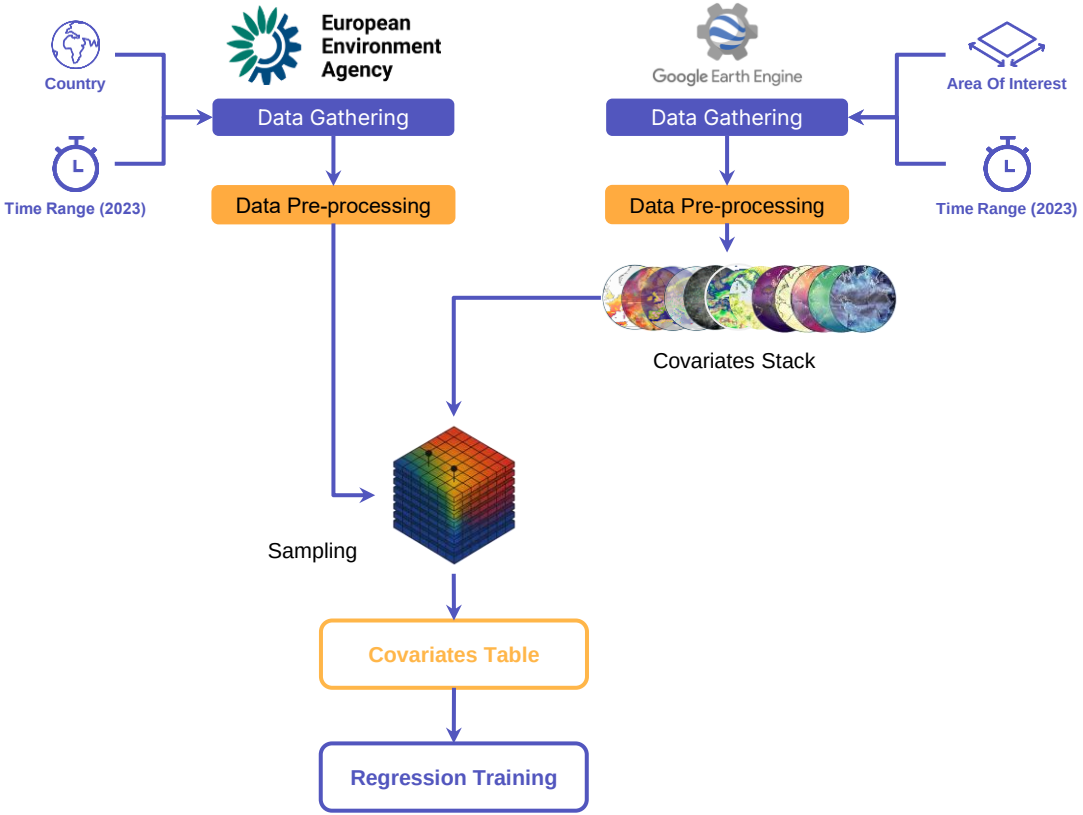
Provide an **alternative solution** when Sentinel-5P data is unavailable or limited due to spatio-temporal gaps.

## Provide ground $NO_2$ Observations

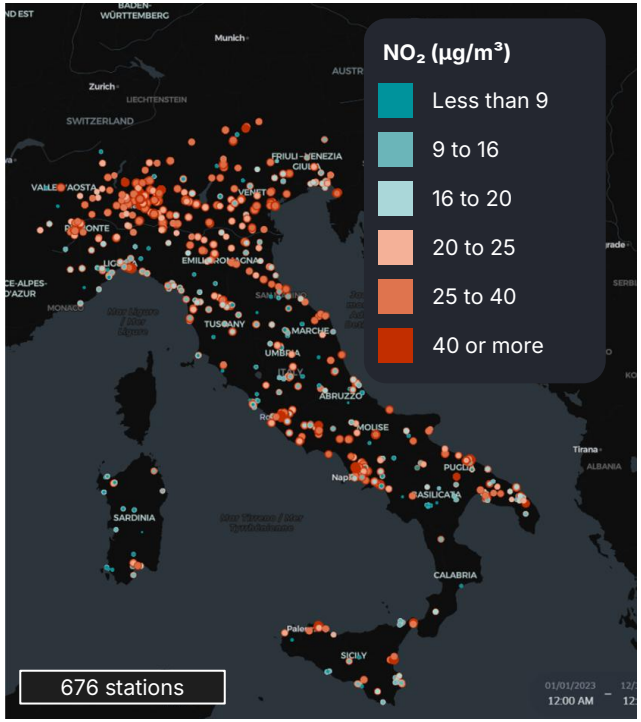
Shift the focus to **ground-level $NO_2$ concentrations**, rather than averaging values across the entire atmospheric column. This enables more accurate assessments of population exposure and health impacts.

CatBoost

# Method: model overview



# Data gathering: Ground Truth



$NO_2$ $SO_2$ HCHO

CO $CH_4$ $SO_2$

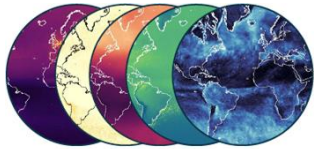
European Environment Agency

## Airbase

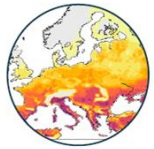
AirBase is a public air quality database containing air quality monitoring information for more than **35 countries** throughout Europe.

AirBase version **several data products** are available through the EEA data service.

# Data gathering: The Covariates



Sentinel-5P TROPOMI L2
(NO2, SO2, CO, O3 and clouds)



ECMWF
ERA5 hourly estimates

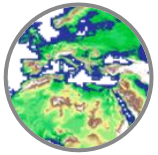
- NO2 (Sentinel-5p)
- SO2 (Sentinel-5p)
- O3 (Sentinel-5p)
- Clouds (Sentinel-5p)
- Surface Temperature (ERA5)
- Solar Radiation (ERA5)
- Total Precipitation (ERA5)
- Planetary Boundary Layer Height (ERA5)

**Time dependent**

**Spatial res:**
**3-28 km**

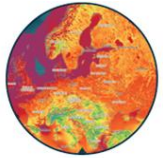
Road Density
OSM + Meta

CORINE Land Cover
at 100m resolution



GLO 30

Slope



Global Wind Atlas

- Land Cover (CORINE)
- elevation (GLO-30)
- Population (wsf)
- Slope (GLO-30)
- Wind speed (Atlas)
- Road Density (OSM + Meta)

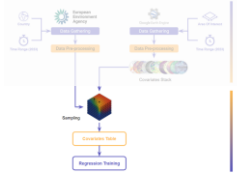
**Time independent**

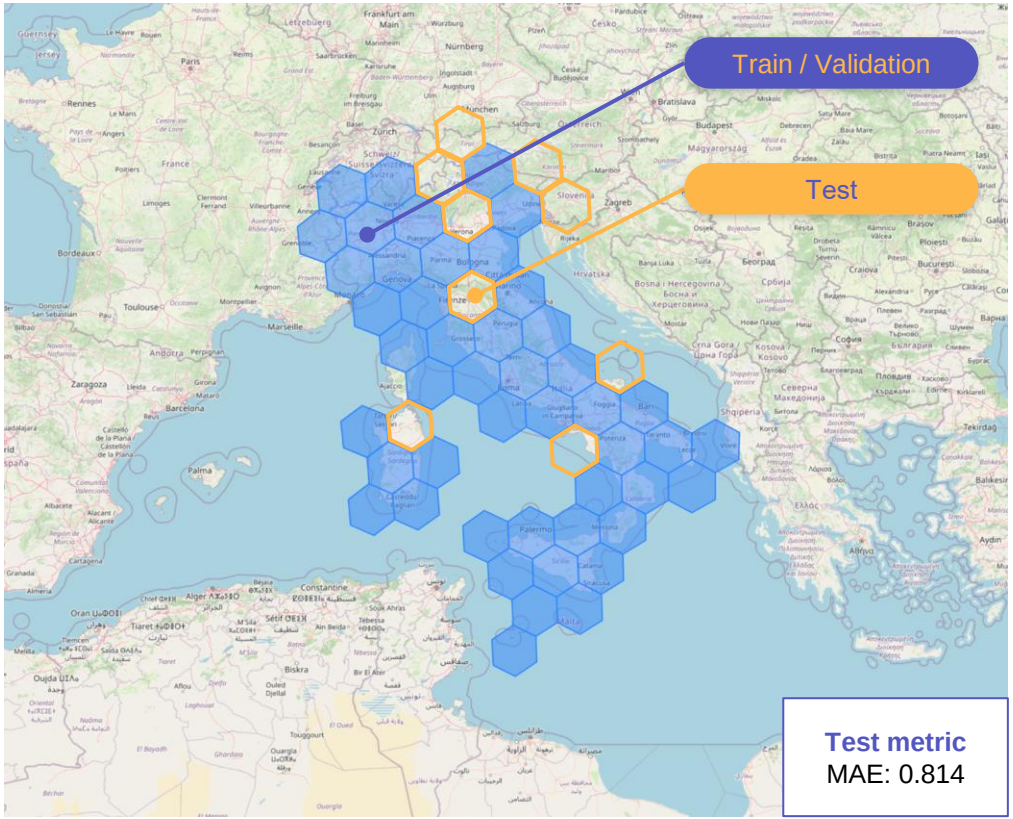
**Spatial res:**
**100-250 m**

\+

- Month
- Day of the Year
- Weekday
- Latitude
- Longitude

# Training: Train/Test split and performances

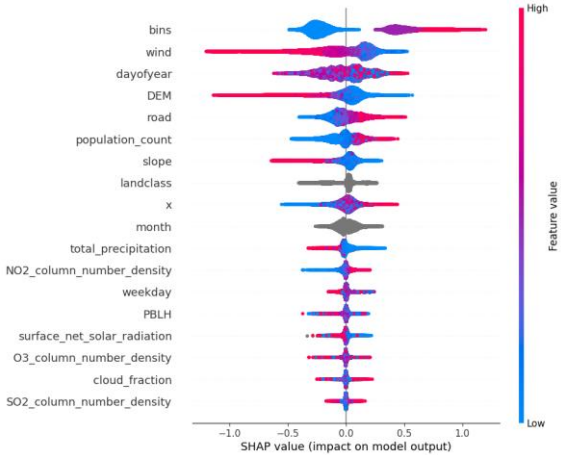





## SHAP Values

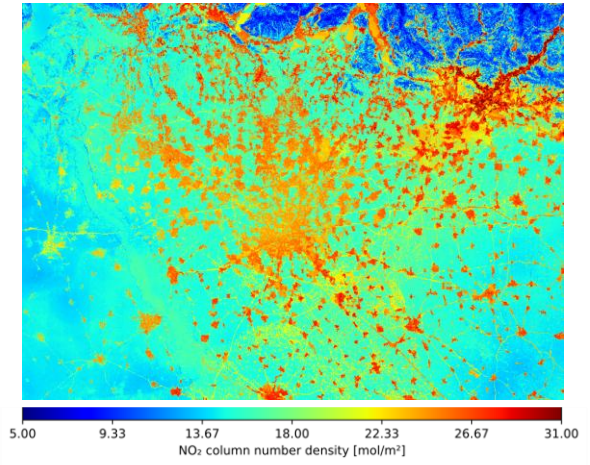
SHAP (SHapley Additive exPlanations) values are a way to explain the output of any machine learning model.

It uses a game theoretic approach that measures each player's contribution to the final outcome.

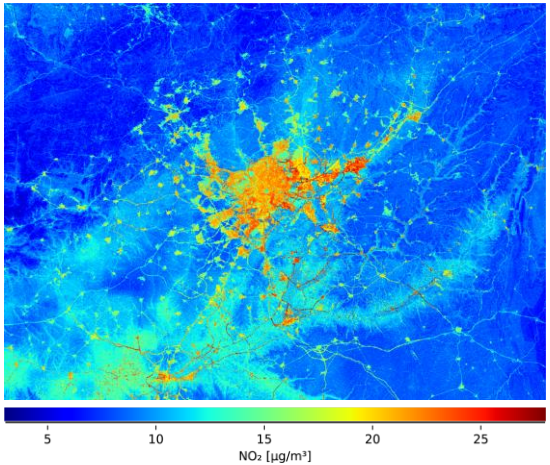


## Model Output: Daily Ground NO2 Inference Maps

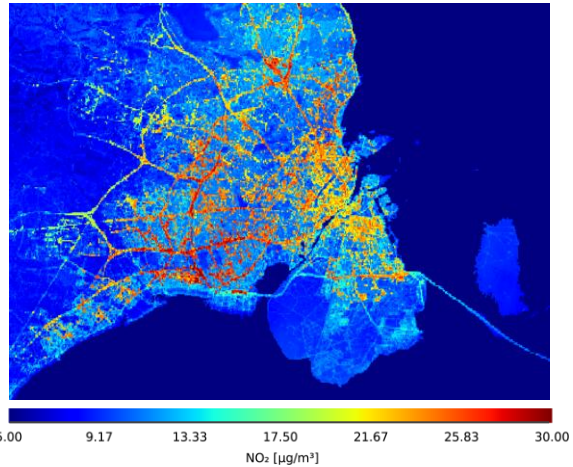
Spatial map showing estimated $NO_2$ concentrations at 100m resolution in µg/m³ for a specific date.
The model can generate such maps for any desired date, provided the necessary input data is available in Google Earth Engine.




Milano, Italy, 2025/03/01 - $NO_2$ (µg/m³), the inference shows the pollutant concentration in the city of Milano.



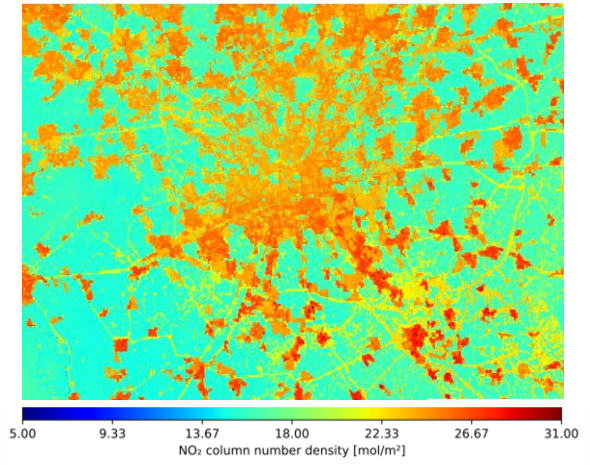

Madrid, Spain, 2025/03/01 - $NO_2$ (µg/m³), the inference shows a higher concentration of the pollutant in the city area.



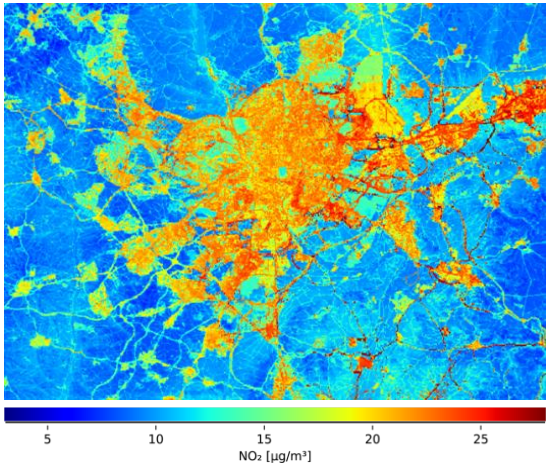

Copenhagen, Denmark, 2023/03/12 - $NO_2$ (µg/m³), the inference shows a higher concentration of the pollutant in the motorways around the city.

## Model Output: Daily Ground NO2 Inference Maps

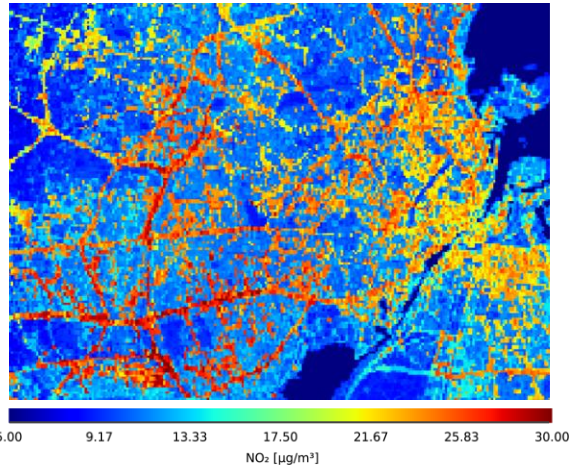
Spatial map showing estimated $NO_2$ concentrations at 100m resolution in $\mu g/m^3$ for a specific date.
The model can generate such maps for any desired date, provided the necessary input data is available in Google Earth Engine.




Milano, Italy, 2025/03/01 - $NO_2$ ($\mu g/m^3$), the inference shows the pollutant concentration in the city of Milano.

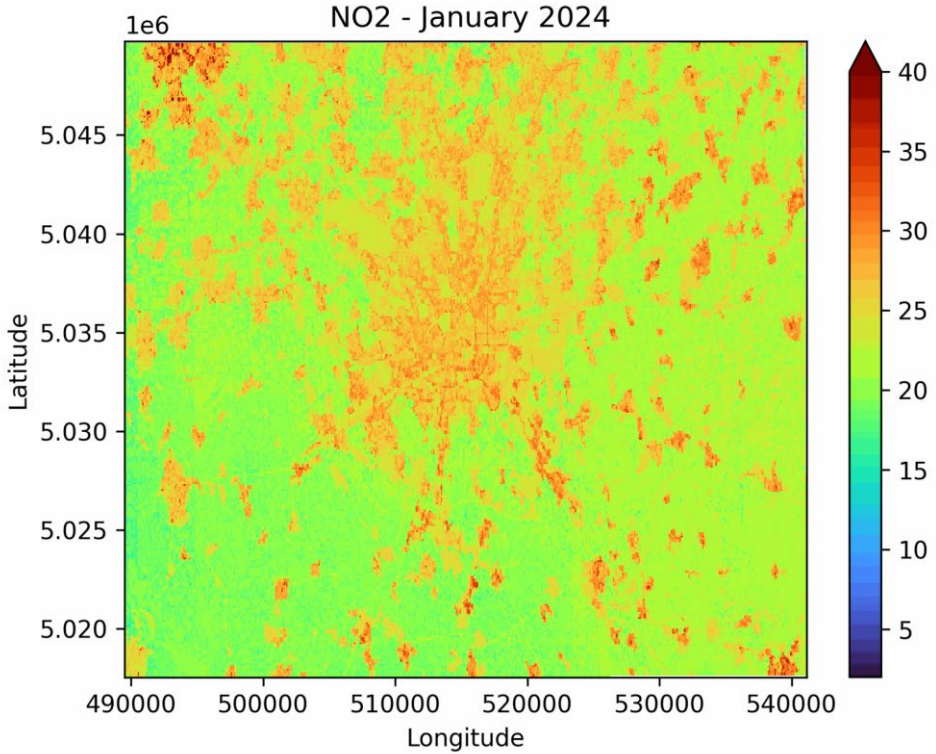



Madrid, Spain, 2025/03/01 - $NO_2$ ($\mu g/m^3$), the inference shows a higher concentration of the pollutant in the city area.




Copenhagen, Denmark, 2023/03/12 - $NO_2$ ($\mu g/m^3$), the inference shows a higher concentration of the pollutant in the motorways around the city.

- Monthly means: $NO_2$ levels tend to peak during the winter months, an annual trend that is clearly visible here and consistently captured by both our model reflecting the trend of the Airbase ground stations.



### A scalable $NO_2$ estimation pipeline

Development of a **scalable** regression model that predicts **ground-level** $NO_2$ concentrations at **100m** resolution with **daily** outputs, enabling insights into spatial and temporal pollution patterns and supporting real-time monitoring and intervention.

### Detection of Local Changes

The Airbase database includes pollutants beyond those from Sentinel-5P, enabling our model to generate high-resolution predictions for a wider range of pollutants.

### Model Performances

In the Italy use case, the model achieved a test **MAE** of **0.814**, demonstrating strong accuracy in predicting ground-level $NO_2$.

### NO2 4-day forecast

Using forecast data from the Extremes DT by ECMWF as a covariate, our model can generate high-resolution $NO_2$ forecasts up to 4 days ahead.

### Test other Airbase pollutants

The Airbase database includes pollutants beyond those from Sentinel-5P, enabling our model to generate high-resolution predictions for a wider range of pollutants.

### Global training

We currently train separate models per country, but with more computational resources, we could develop a unified global model for broader application.

### Hourly estimates

Instead of daily aggregation, the model could be trained and run at hourly intervals for finer temporal resolution.

You can find us at
**booth S27, Level 1**

# Thank you