Improving Real-time High-Resolution Estimations of $PM_{2.5}$ Concentration Fields for the present time in Urban Areas by the SmartAQ+ System with Data Fusion and Machine Learning

Dr. Ioannis D. Apostolopoulos
Postdoctoral Researcher, FORTH/ICE-HT

LAQS LABORATORY OF AIR QUALITY STUDIES FORTH
C·STACC Center for the Study of Air Quality & Climate Change

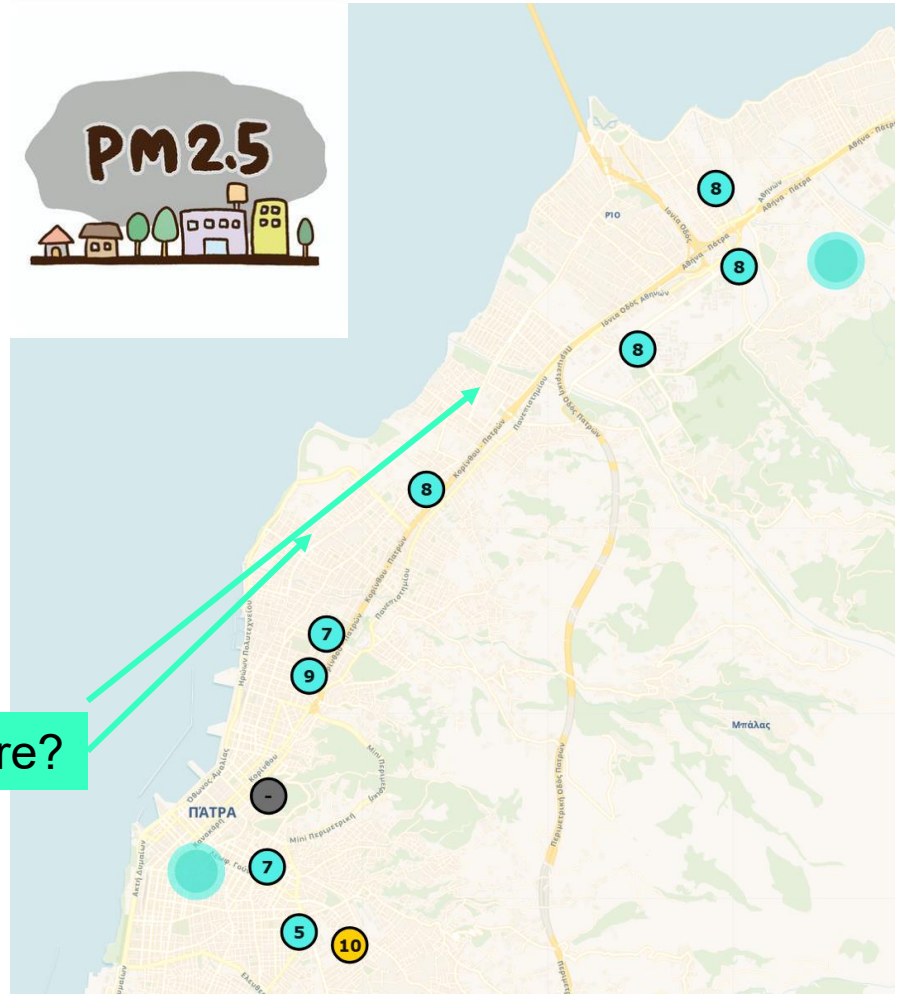
The problem (1/2)

- Monitor $PM_{2.5}$ in real time
- High-resolution → greater coverage
- Provide maps for the general public

- Limited number of regulatory measurement sites
- Limited low-cost sensor coverage

What about here?

Present time $PM_{2.5}$ concentrations by the low-cost sensor network in Patras



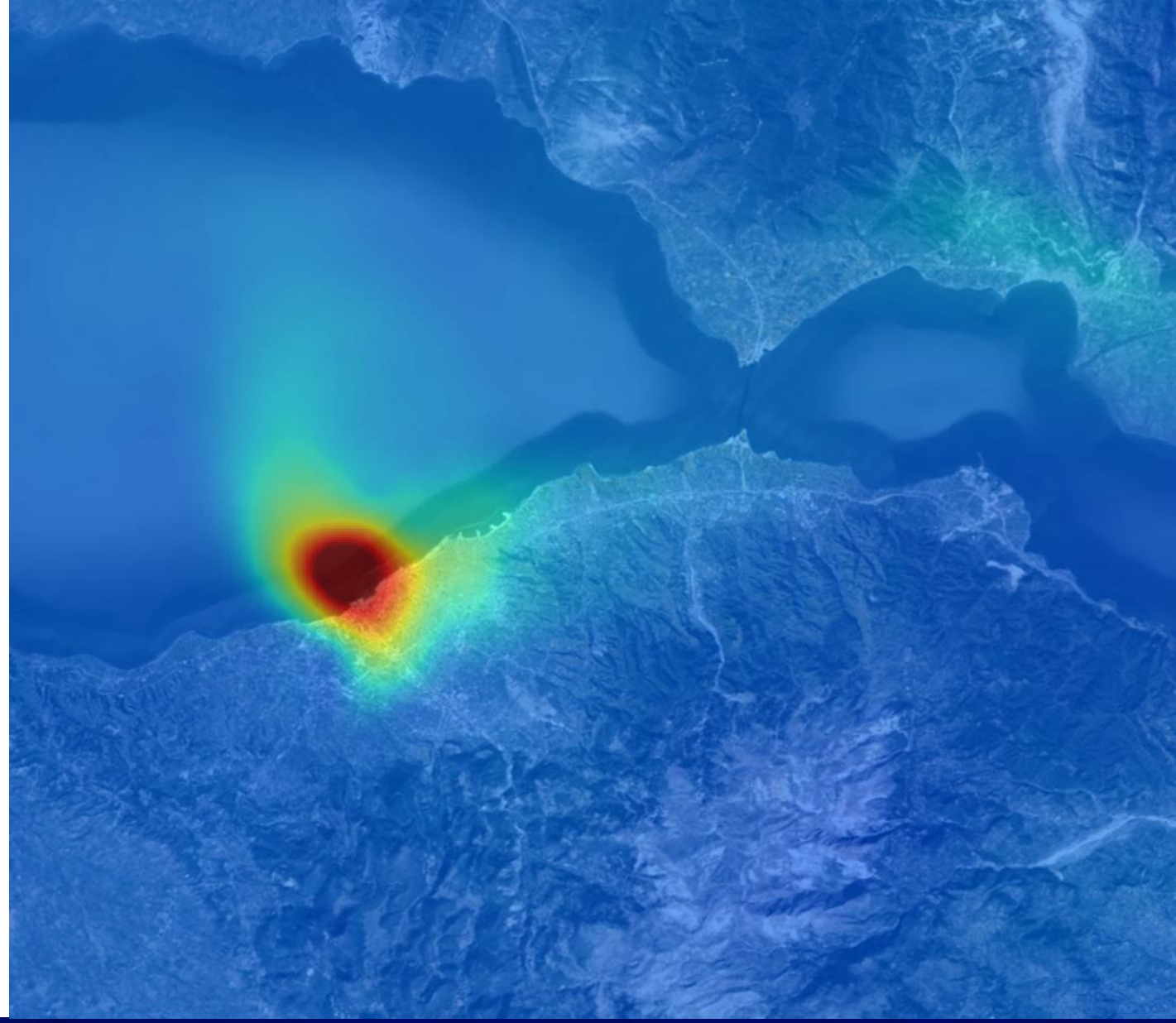
One solution

Chemical Transport Models

- High resolution
- Model the underlying physical and chemical phenomena

- Errors in emissions
- Errors in weather forecasting
- No real-time updates

Errors and Biases

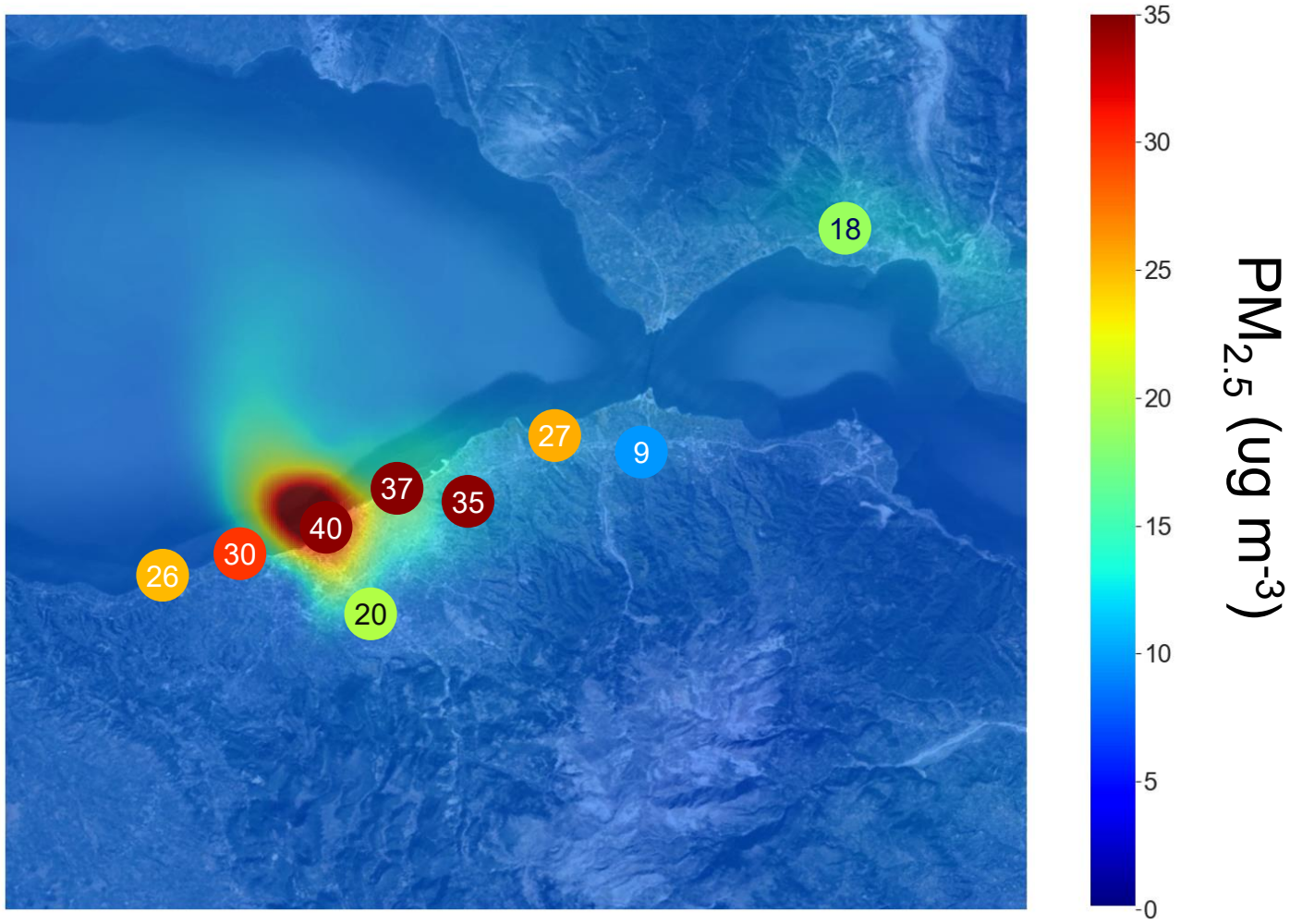


Example of SmartAQ

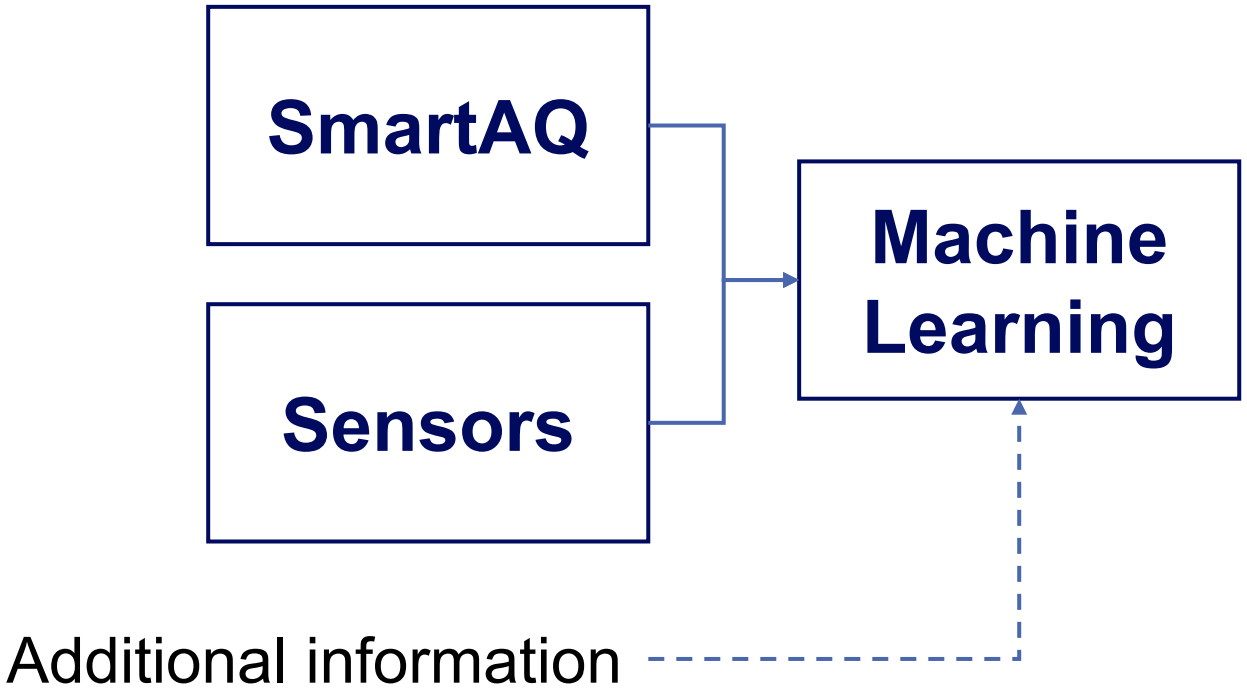
January 28, 2023 at 21:00

Can we improve it ?

SmartAQ



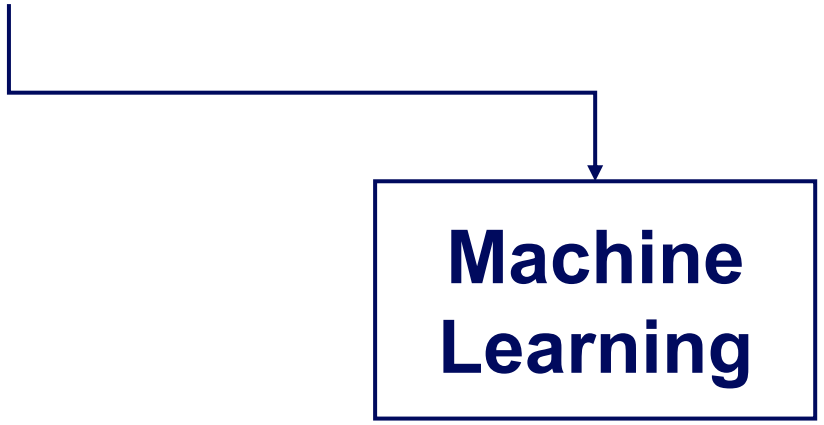
Proposed approach



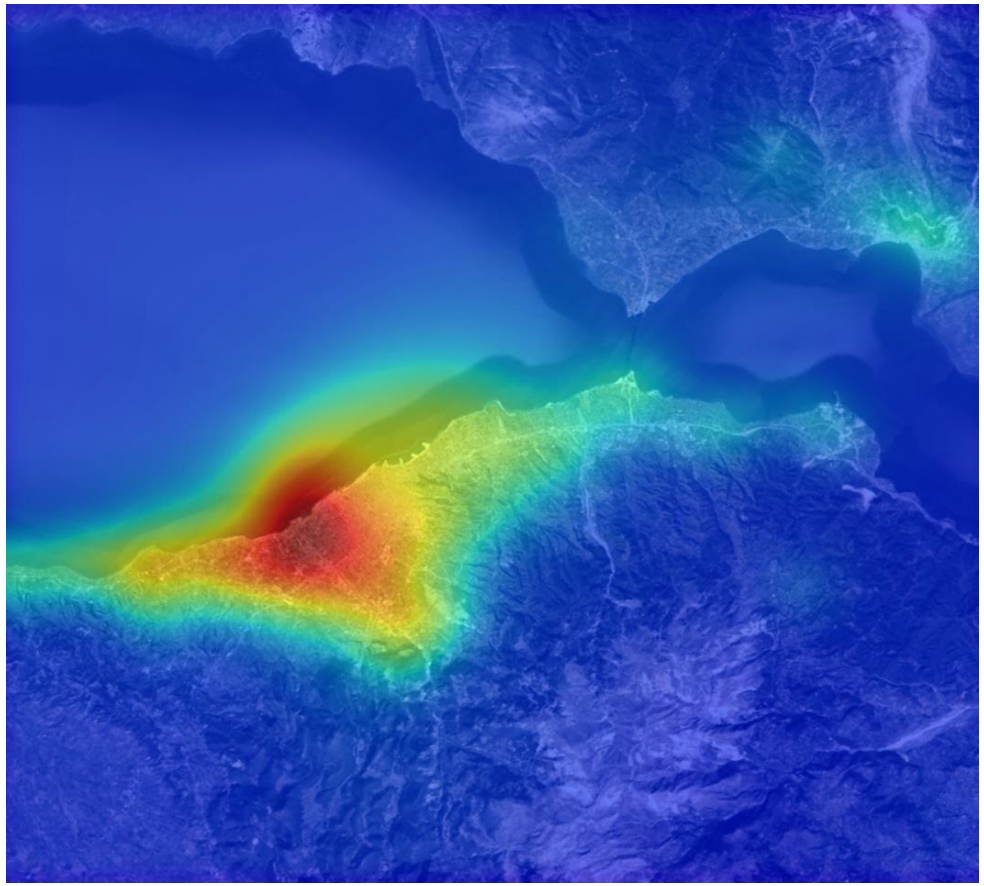
Improved estimations of $PM_{2.5}$ concentration fields for the present time at 1 x 1 Km^2 resolution

Motivation

- Historical sensor data are valuable
- Sensor placement
- $PM_{2.5}$ concentration trends follow patterns
- Patterns might be clustered based on time-variables (like weekday, month, hour)
- Real-time weather effects
- Emissions, land use, and population can be a valuable predictor in remote areas

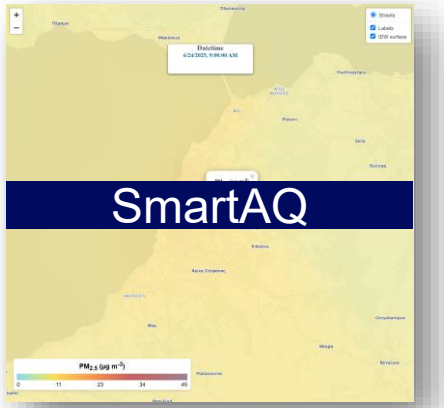


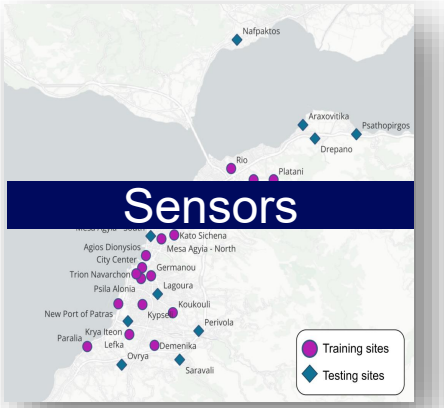
Model outputs

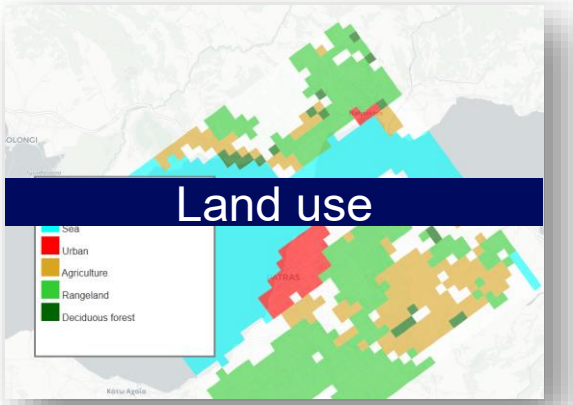


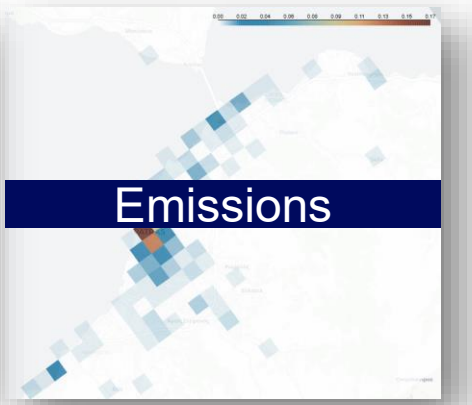
- Present-time $PM_{2.5}$
- 36 x 36 Km^2 area
- 1 x 1 Km^2 resolution

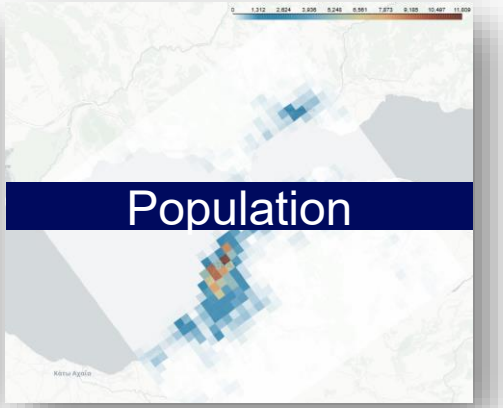
Model Inputs





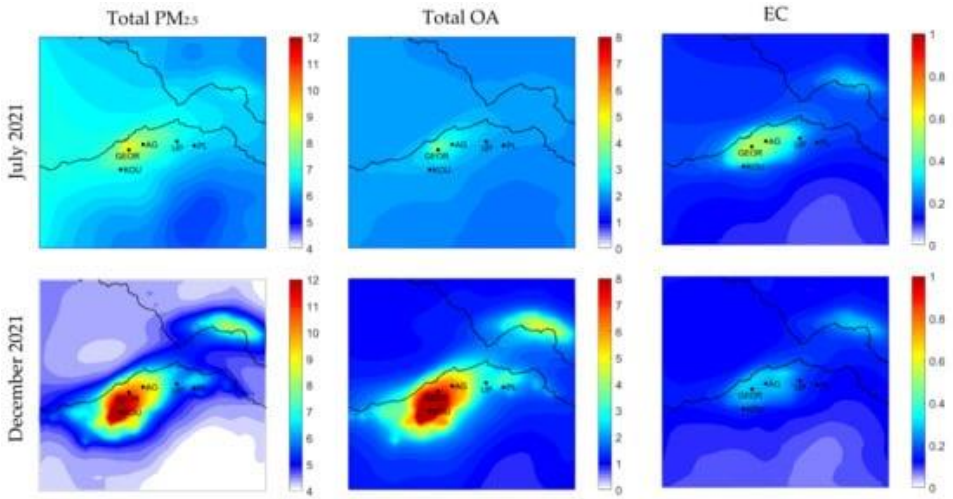






SmartAQ model

- Operates in real time
- Provides three-day forecast of the concentration of gas-phase air pollutants
- Provides the complete aerosol size/composition distribution
- source contributions for all primary and secondary pollutants
- Area of 36 x 36 km^2
- Resolution of 1 x 1 km^2



Siouti E, Skyllakou K, Kioutsioukis I, Patoulias D, Fouskas G, Pandis SN. Development and Application of the SmartAQ High-Resolution Air Quality and Source Apportionment Forecasting System for European Urban Areas. Atmosphere. 2022; 13(10):1693

Low-cost sensor network in Patras

Data period and train-test split

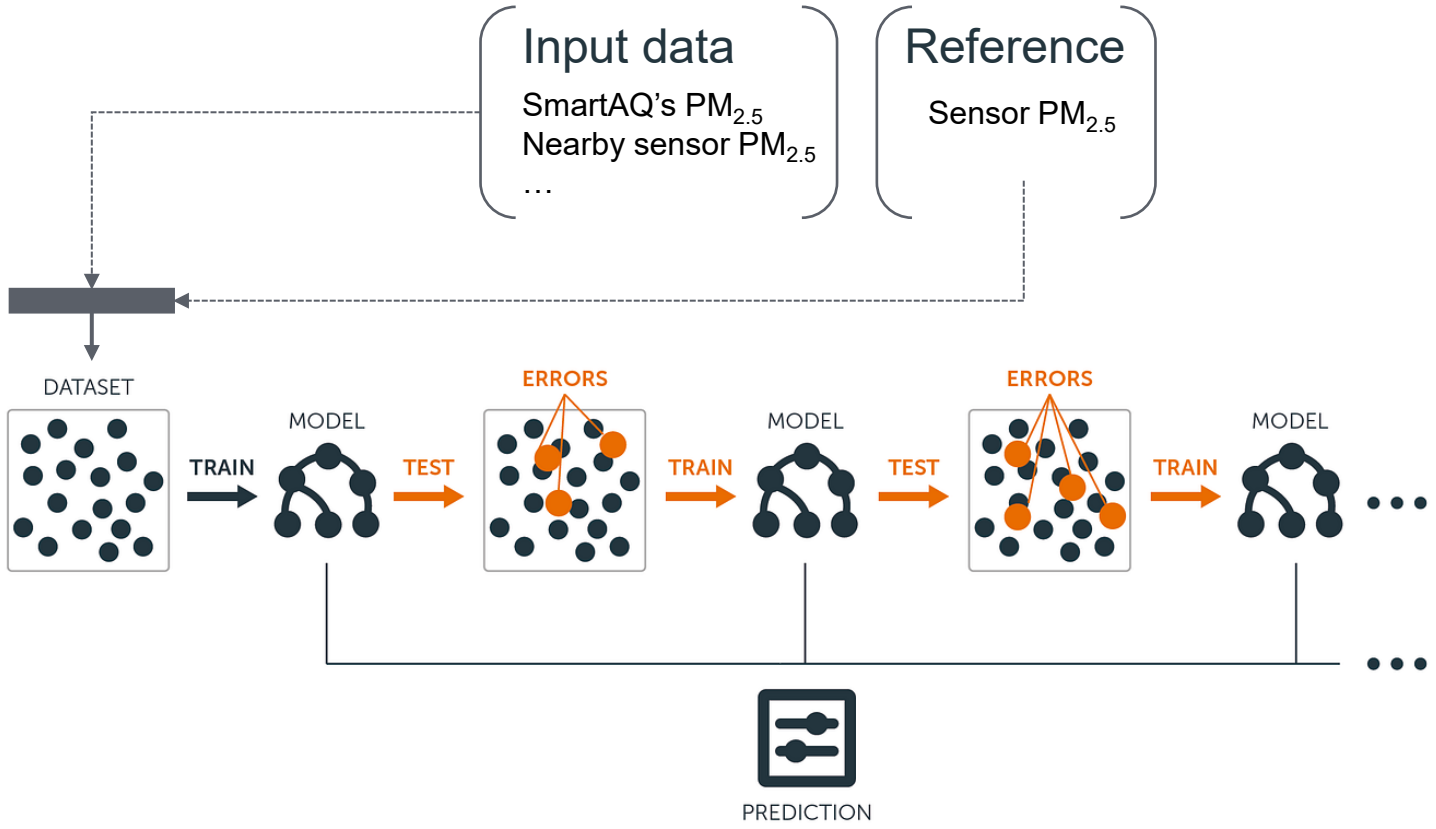
We simulated the retrospective model run every hour during 2023 and compared with historical measurements per location with a sensor

Machine Learning with eXtreeme Gradient Boosting regression

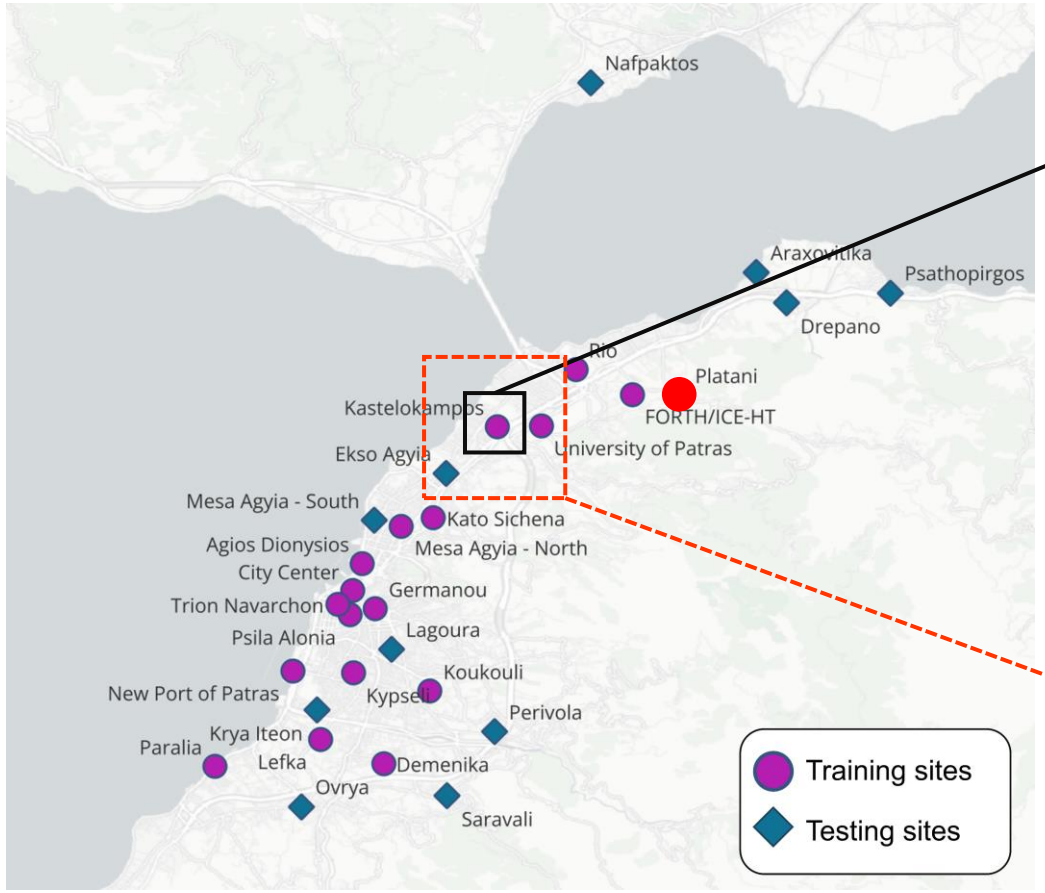
Ensemble Machine Learning

Decision Tree-based

Tries to minimize
Mean Squared Error



Machine Learning model sensor inputs per cell



Forecast model

- SmartAQ prediction

Meteorology

- Weather Data from the nearest station

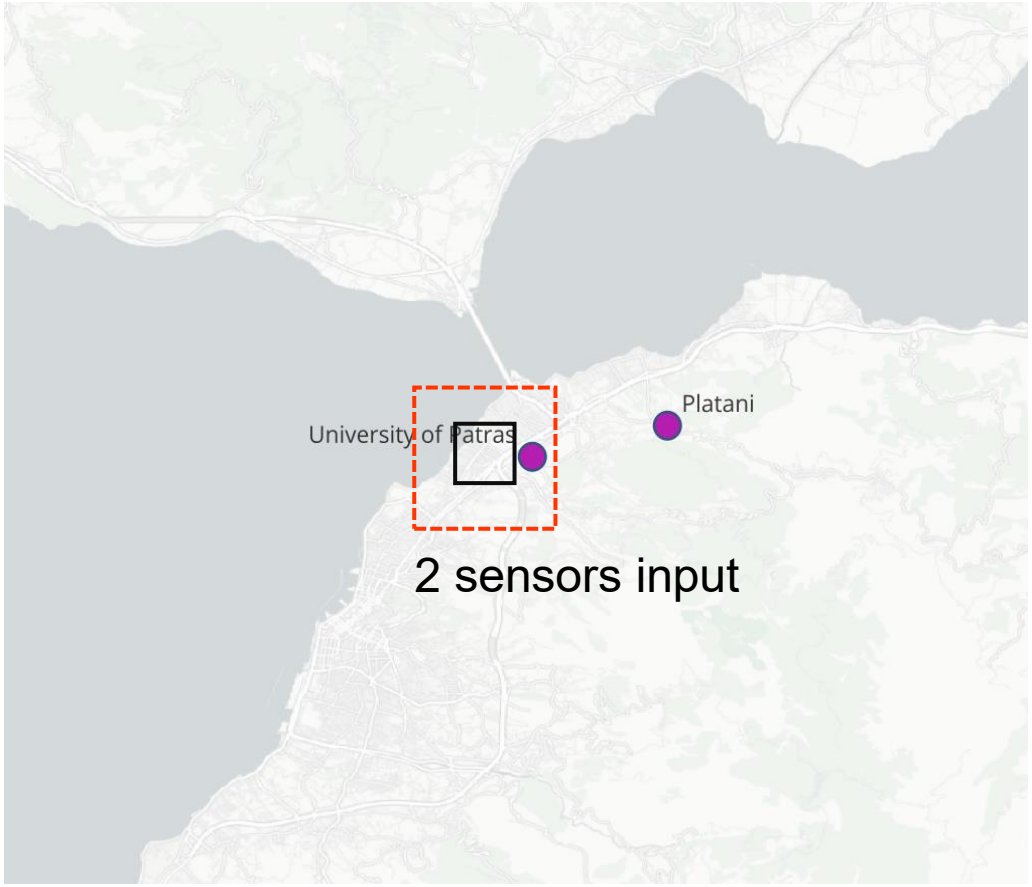
Cell's characteristics

- Land use
- Emissions
- Population

Sensors

- Any sensor inside the 4x4 Km^2 box, excluding the cell of interest
- Background sensor

Machine Learning model sensor inputs per cell

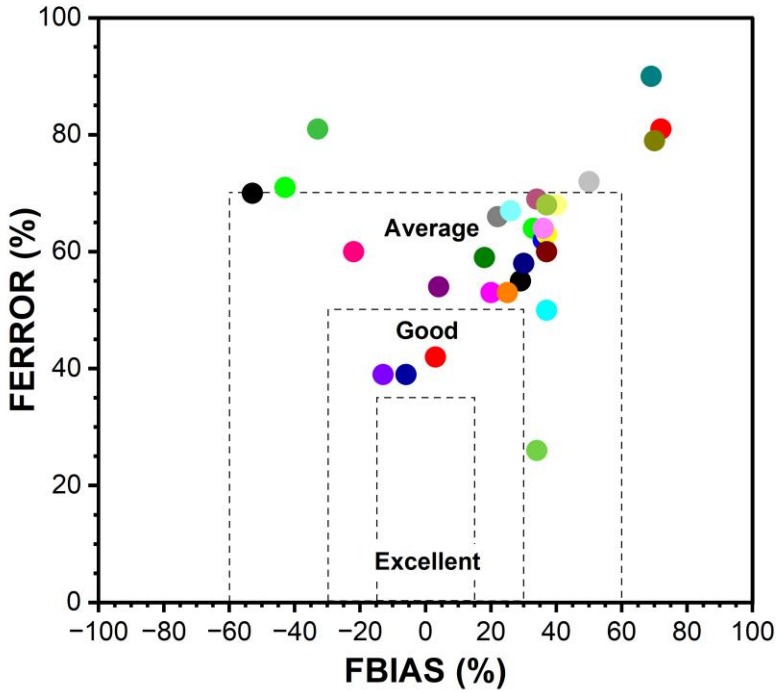


Results

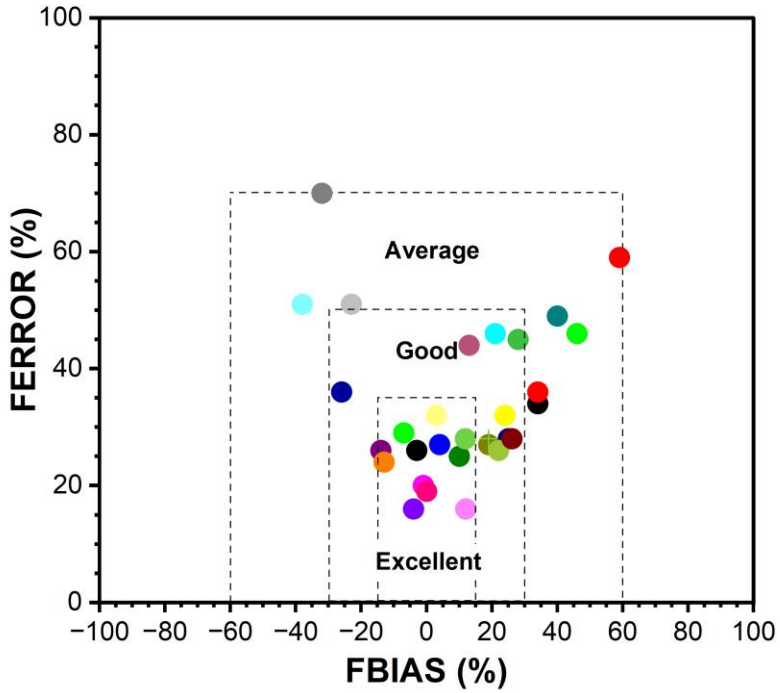
Performance in predicting the average monthly patterns

Performance in predicting the average monthly patterns per location

SmartAQ



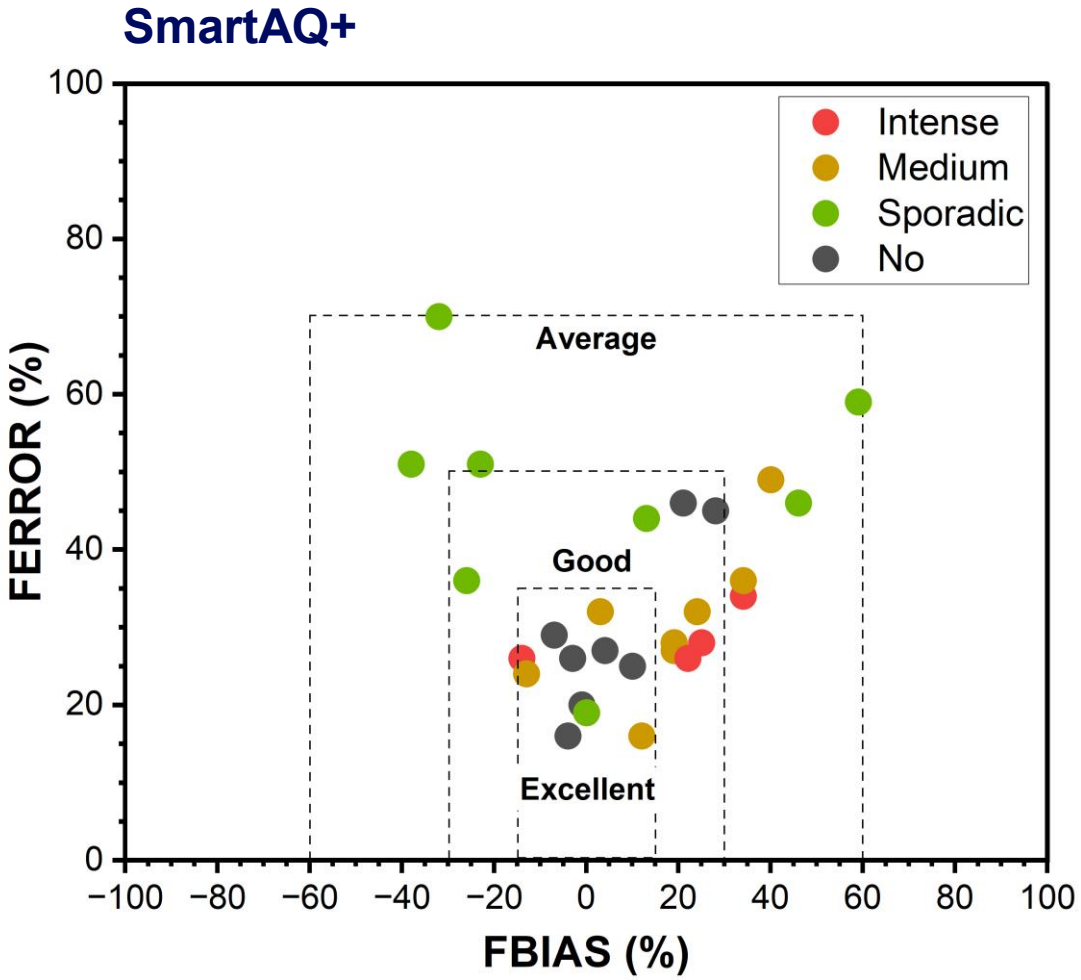
SmartAQ+



Morris, R.E., McNally, D.E., Tesche, T.W., Tonnesen, G., Boylan, J.W., Brewer, P., 2005. Preliminary Evaluation of the Community Multiscale Air Quality Model for 2002 over the Southeastern United States. J. Air Waste Manag. Assoc. 55, 1694–1708. https://doi.org/10.1080/10473289.2005.10464765

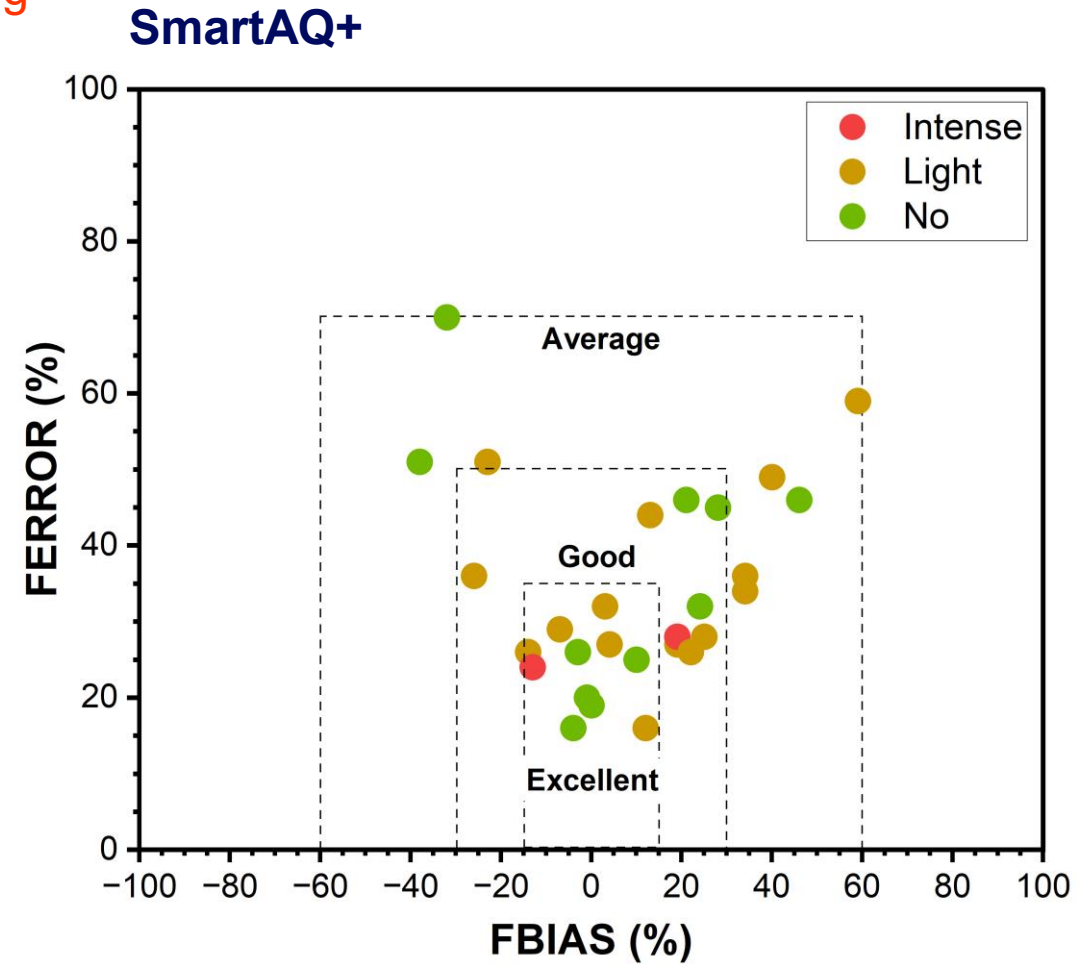
Performance in predicting the average monthly patterns

Locations grouped by BB



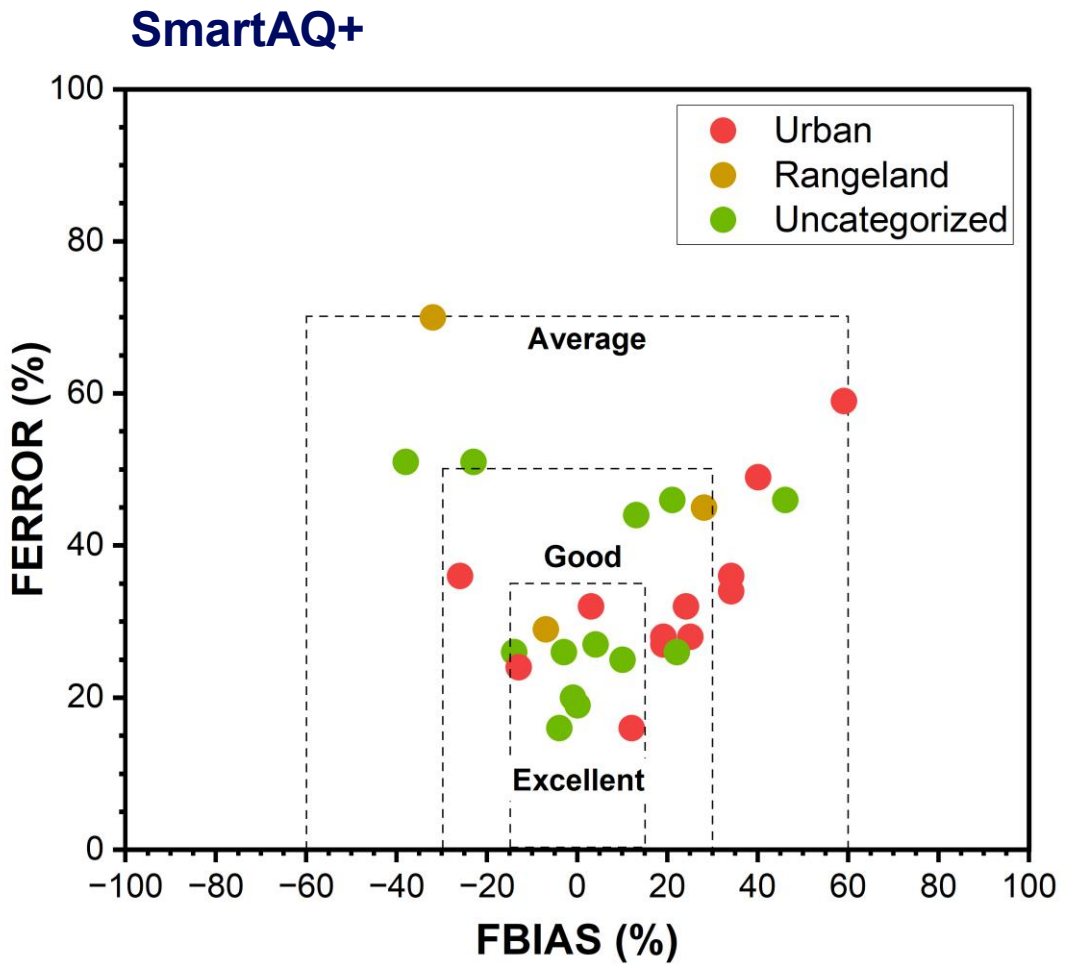
Performance in predicting the average monthly patterns

Locations grouped by Cooking



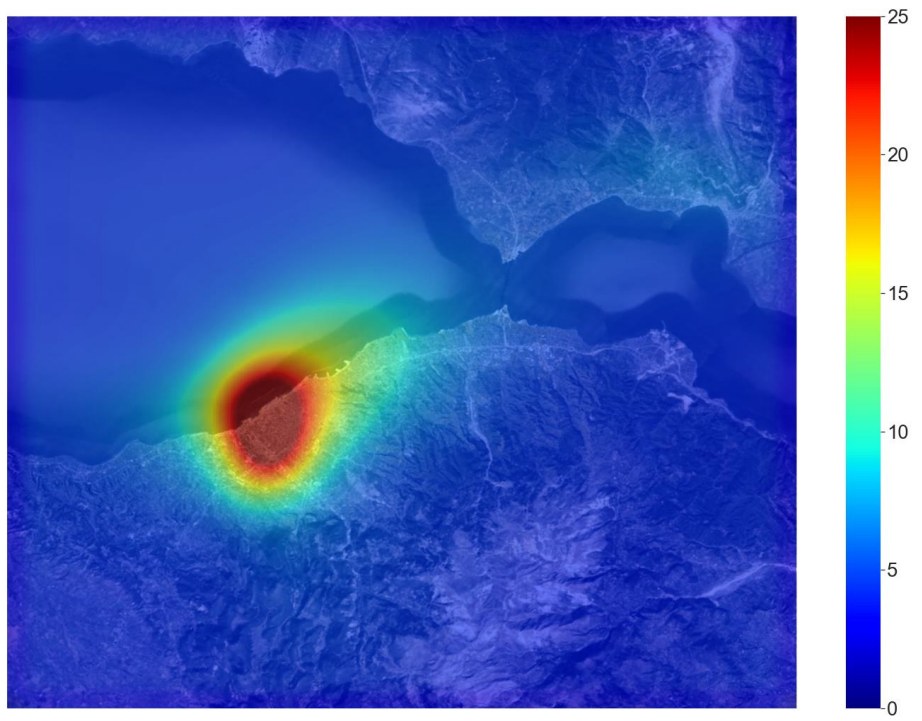
Performance in predicting the average monthly patterns

Locations grouped by Land use

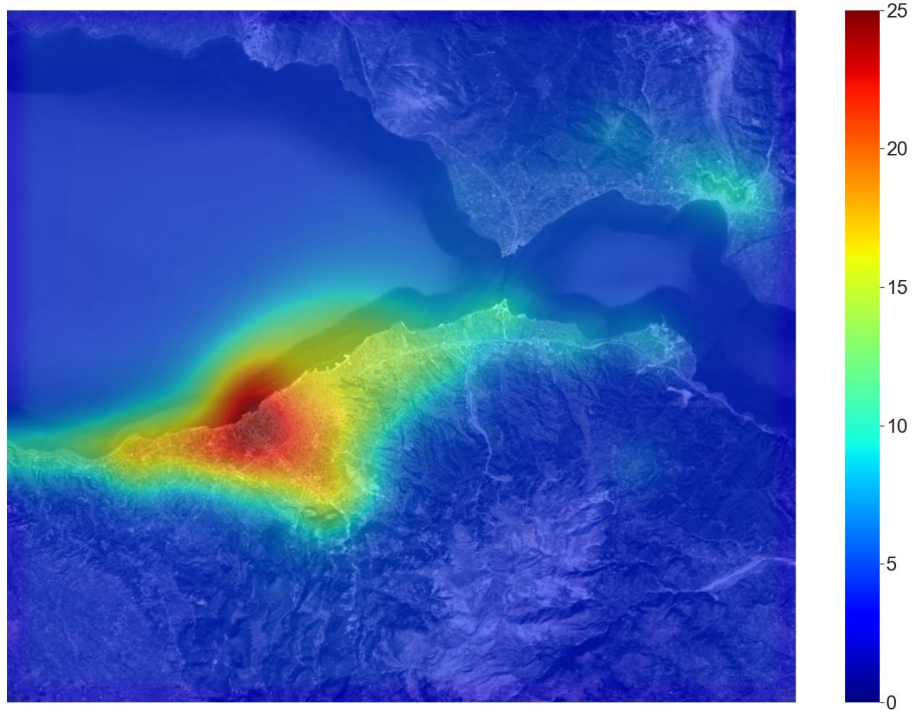


Average predicted $PM_{2.5}$ concentrations in µg m^{-3} during January 2023

SmartAQ

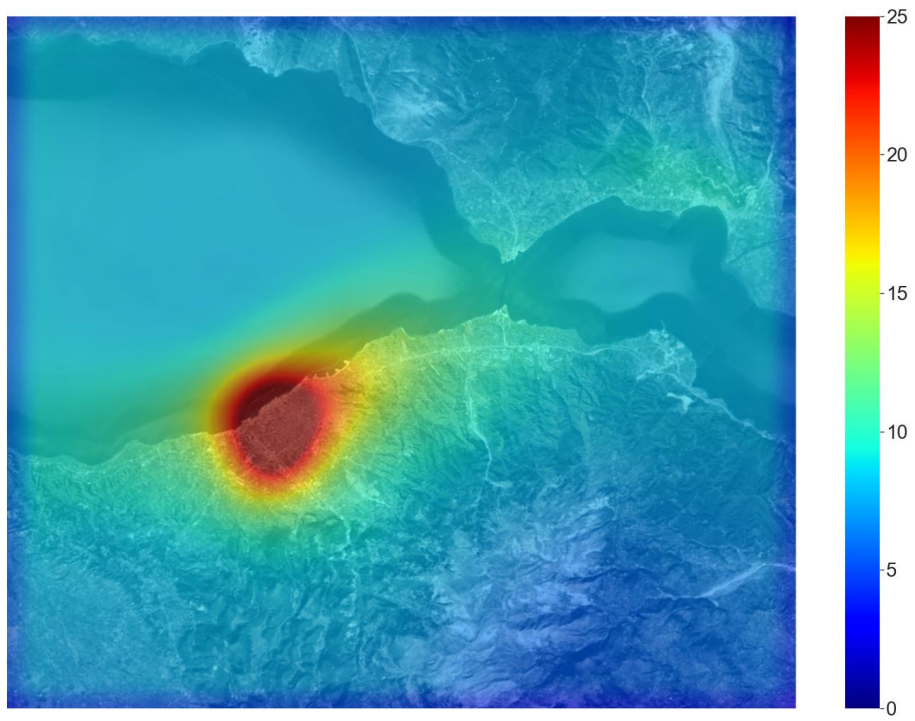


SmartAQ+

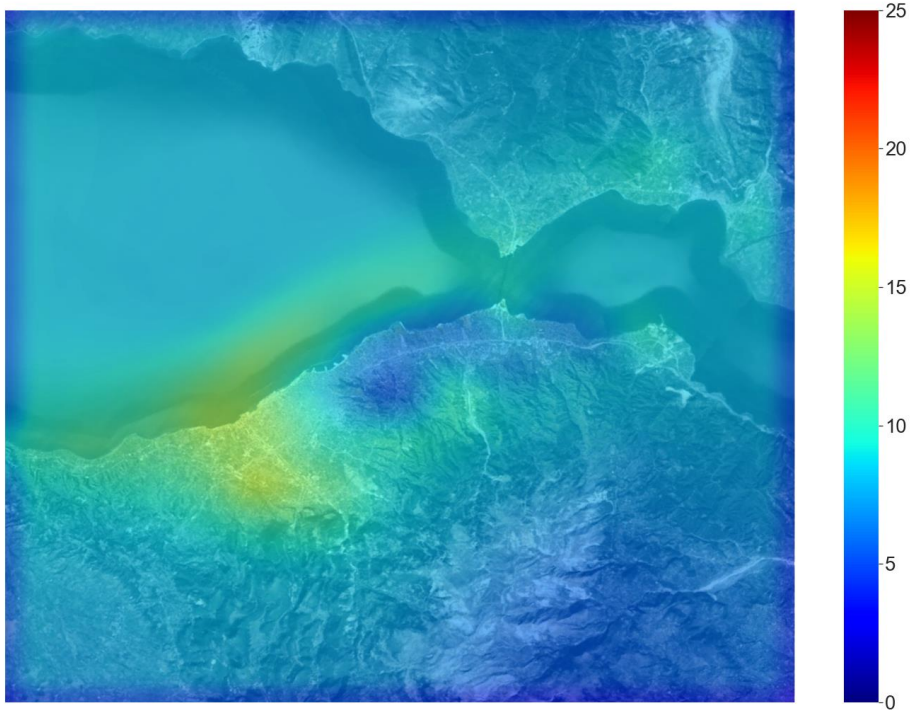


Average predicted $PM_{2.5}$ concentrations in μg m^{-3} during April 2023

SmartAQ



SmartAQ+



Average predicted $PM_{2.5}$ concentrations in µg m^{-3} during July 2023

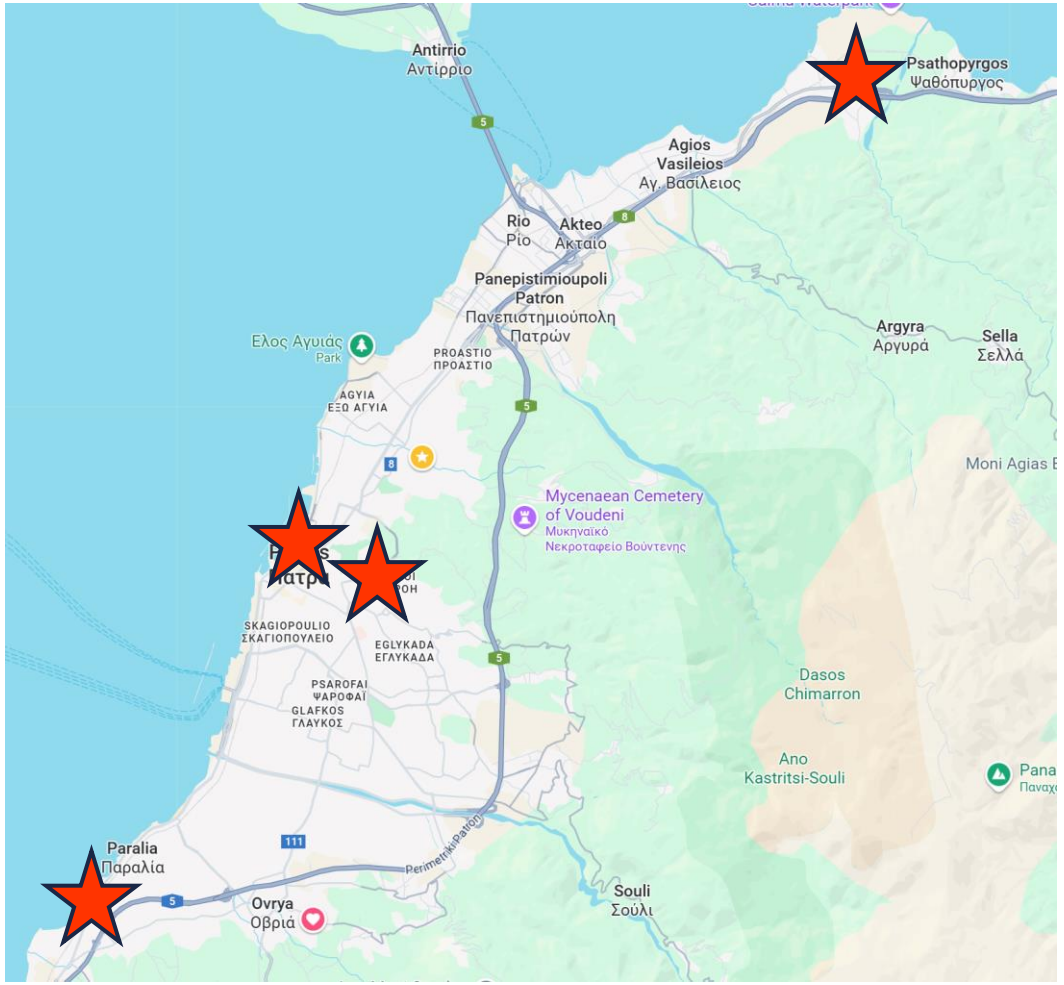
SmartAQ

SmartAQ+

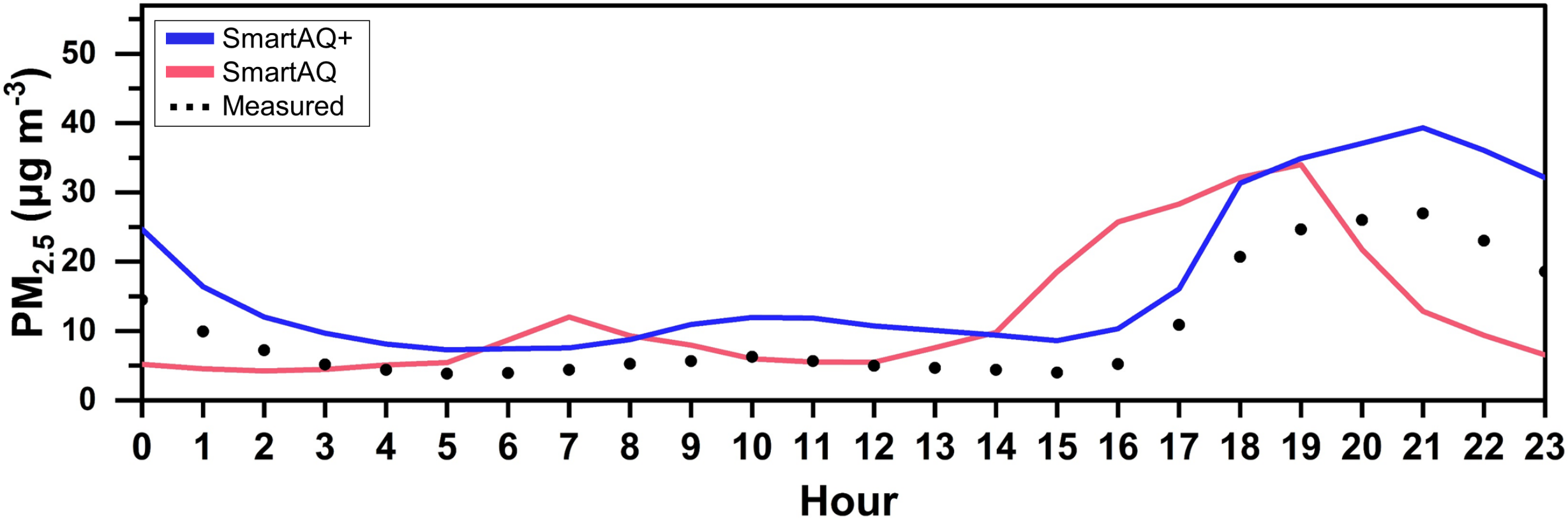
Performance in predicting the average daily patterns

Evaluation sites

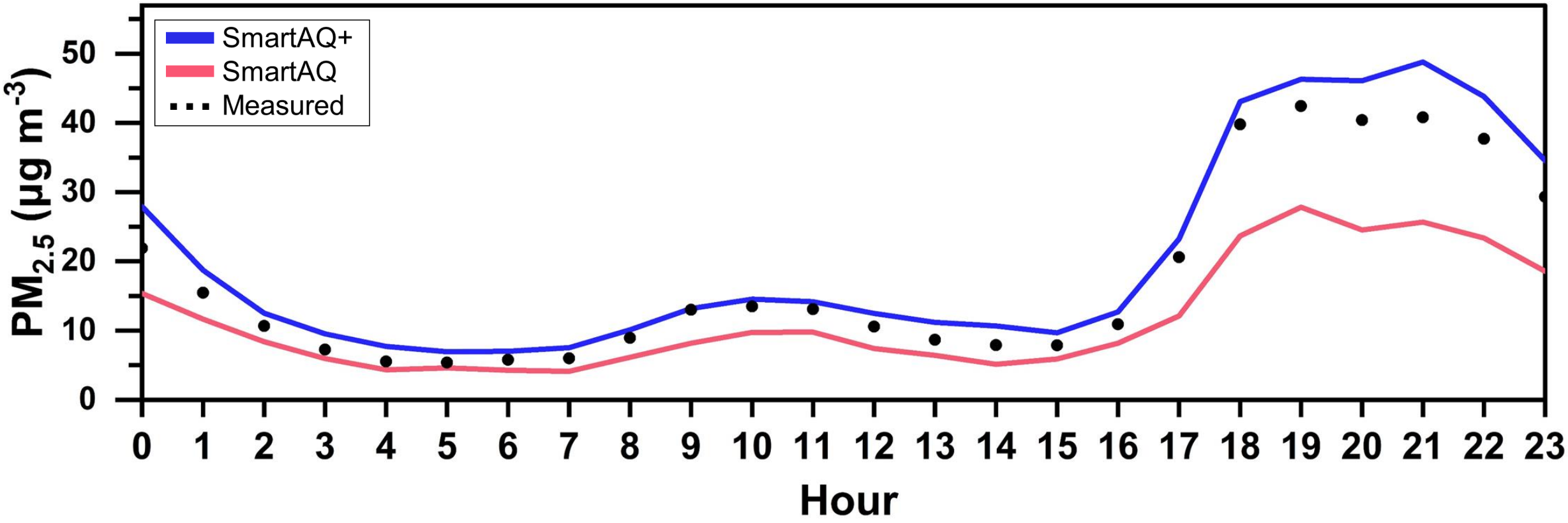
- City center
- Lagouras (near the city center) – **hidden site**
- Paralia (remote)
- Psathopirgos (remote) – **hidden site**



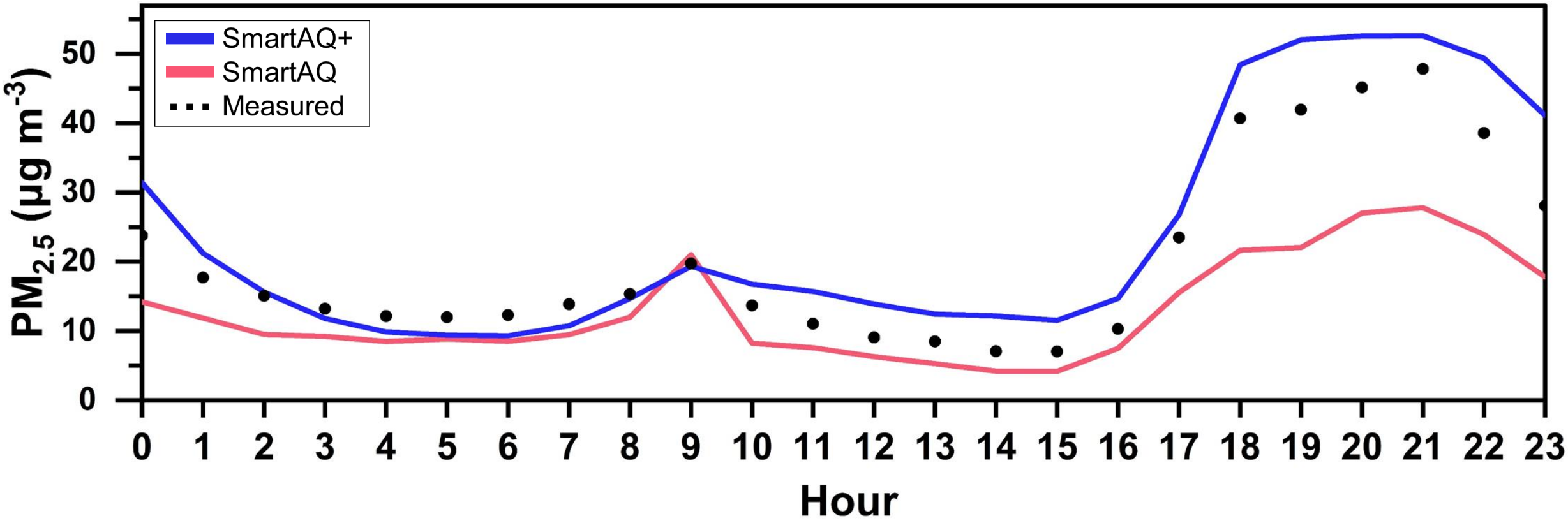
Average diurnal profile during January 2023 in the center of Patras



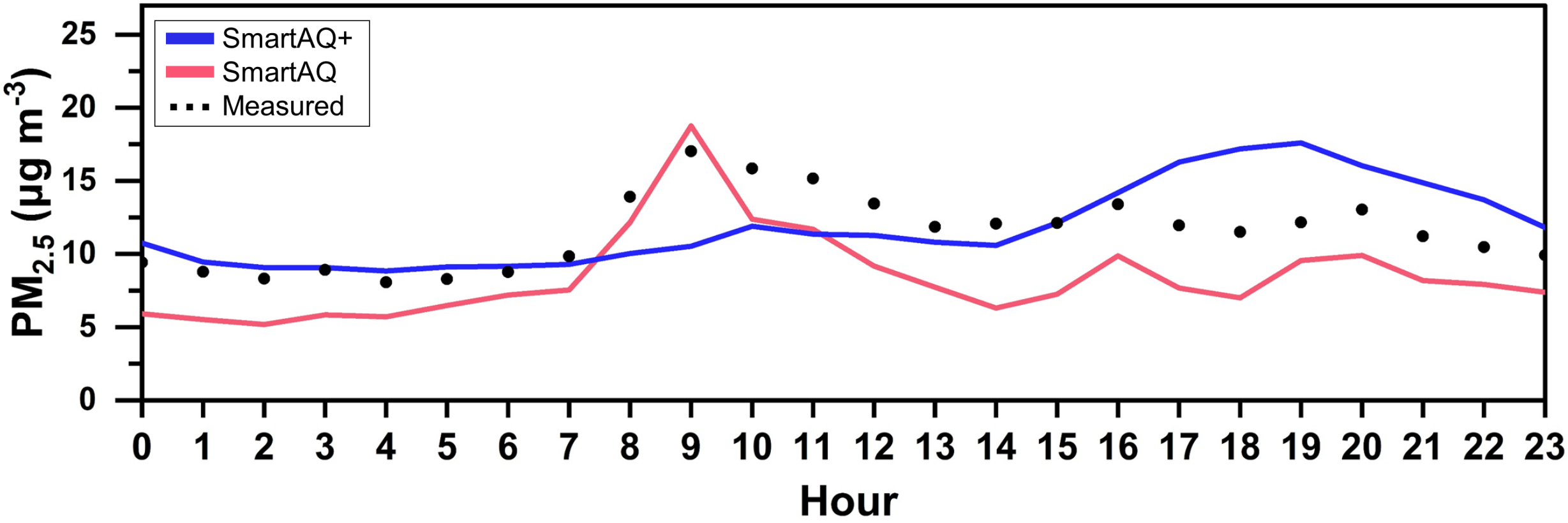
Average diurnal profile during January 2023 in Lagouras



Average diurnal profile during January 2023 in Paralia



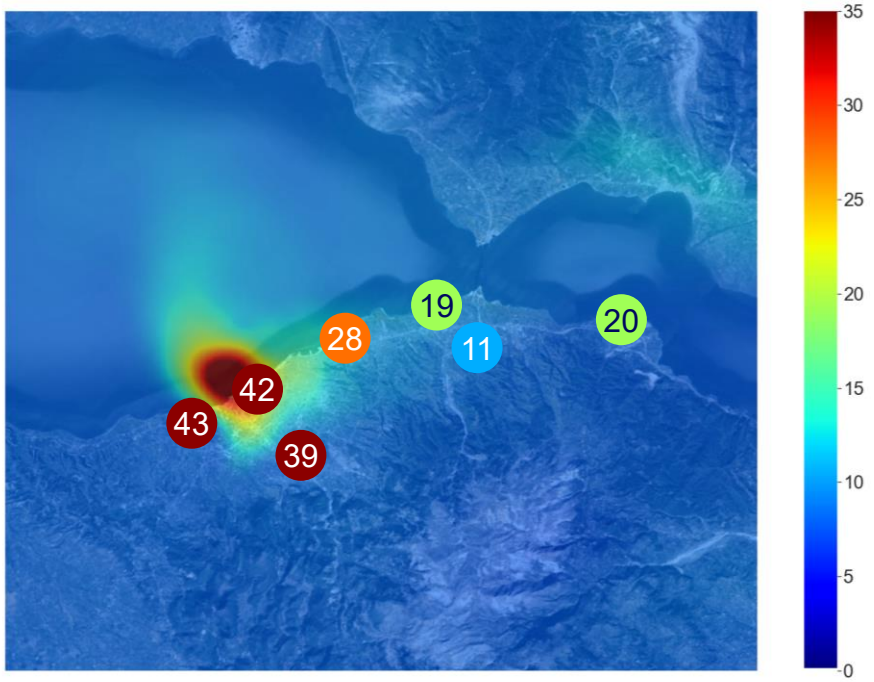
Average diurnal profile during January 2023 in Psathopirgos



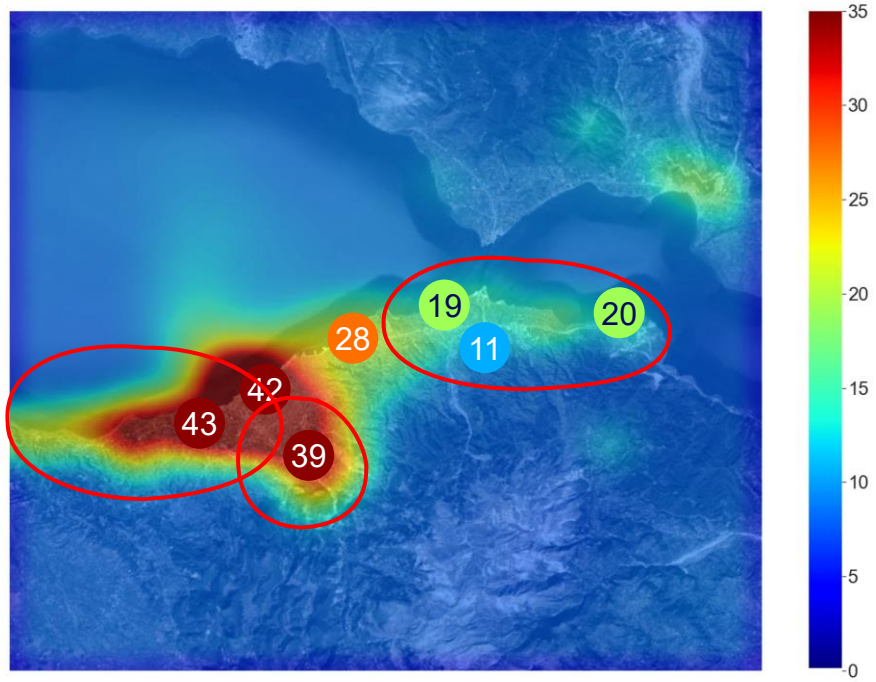
Predicted and measured $PM_{2.5}$ map on January 28, 2023 at 21:00

Average Temperature 7 °C

SmartAQ



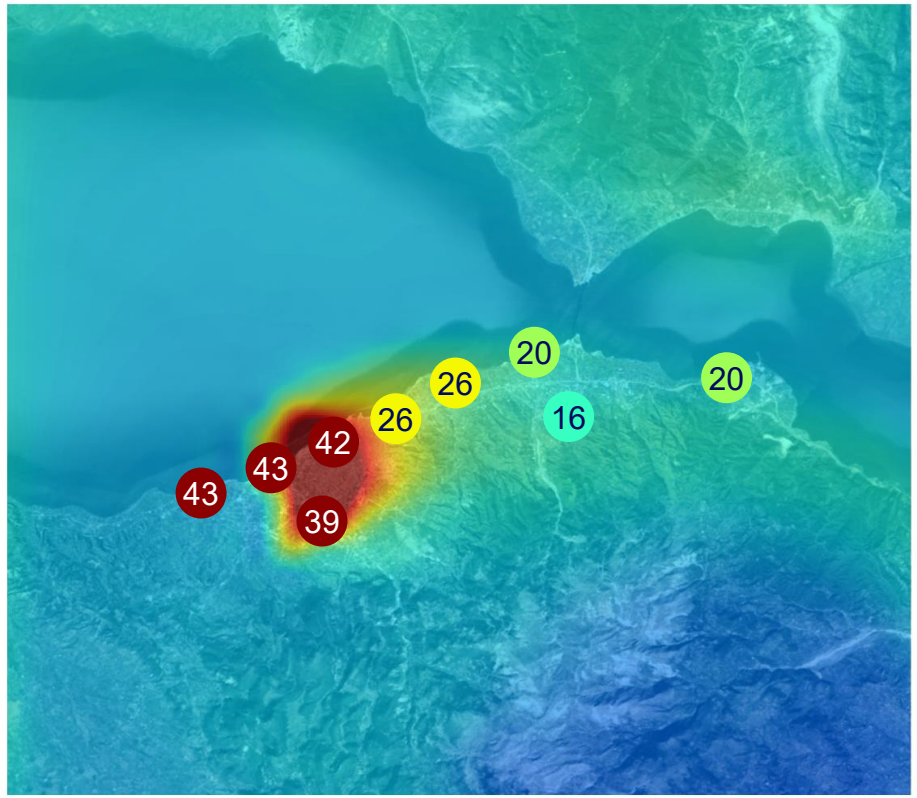
SmartAQ+



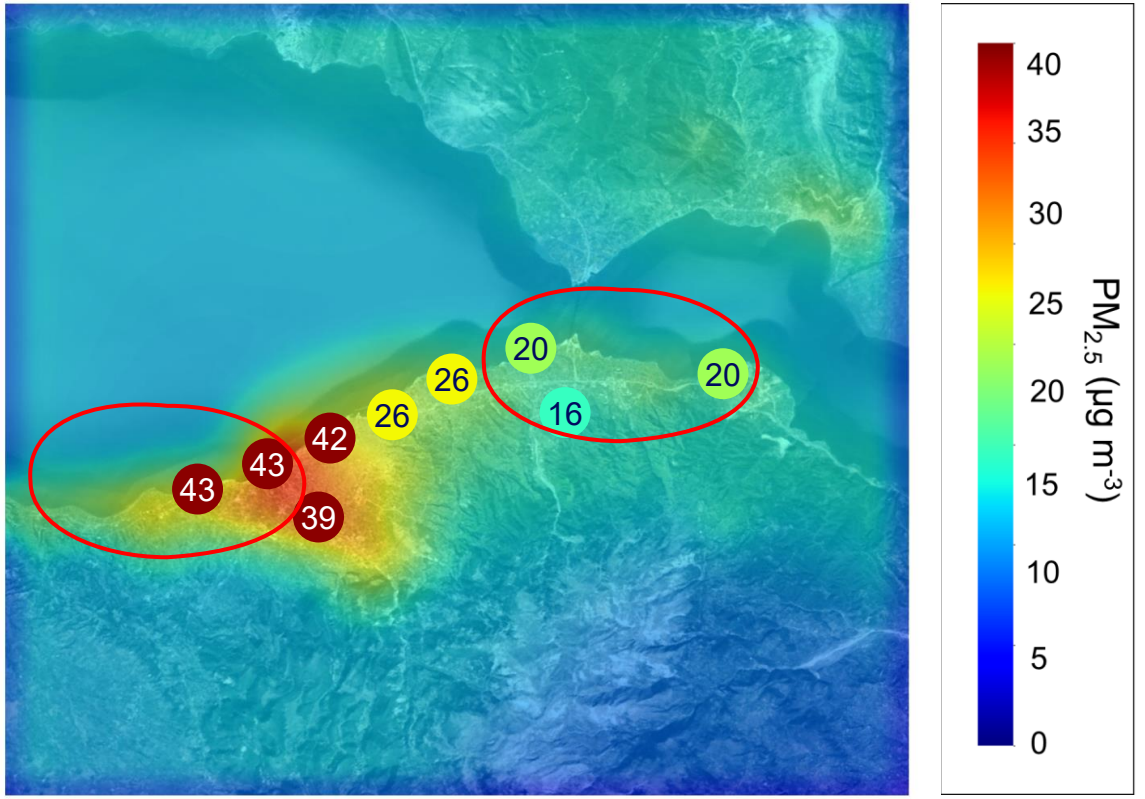
Average predicted and measured $PM_{2.5}$ map during Fat Thursday 2023

February 16, 2023

SmartAQ

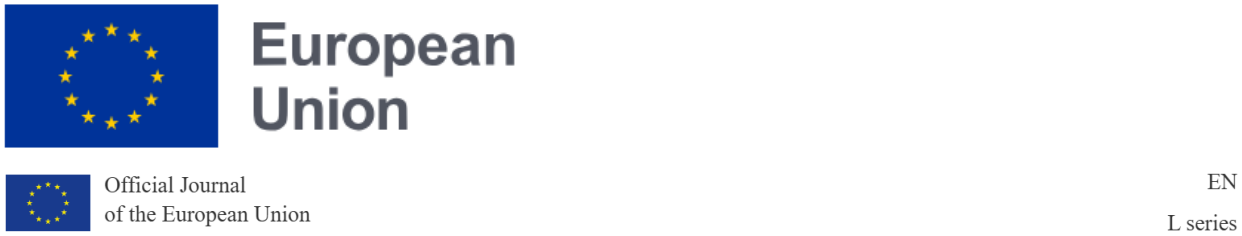


SmartAQ+



Performance in predicting daily $PM_{2.5}$ limit exceedance

New EU Directive



European Union

Official Journal of the European Union

EN

L series

2024/2881

20.11.2024

DIRECTIVE (EU) 2024/2881 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 23 October 2024

on ambient air quality and cleaner air for Europe

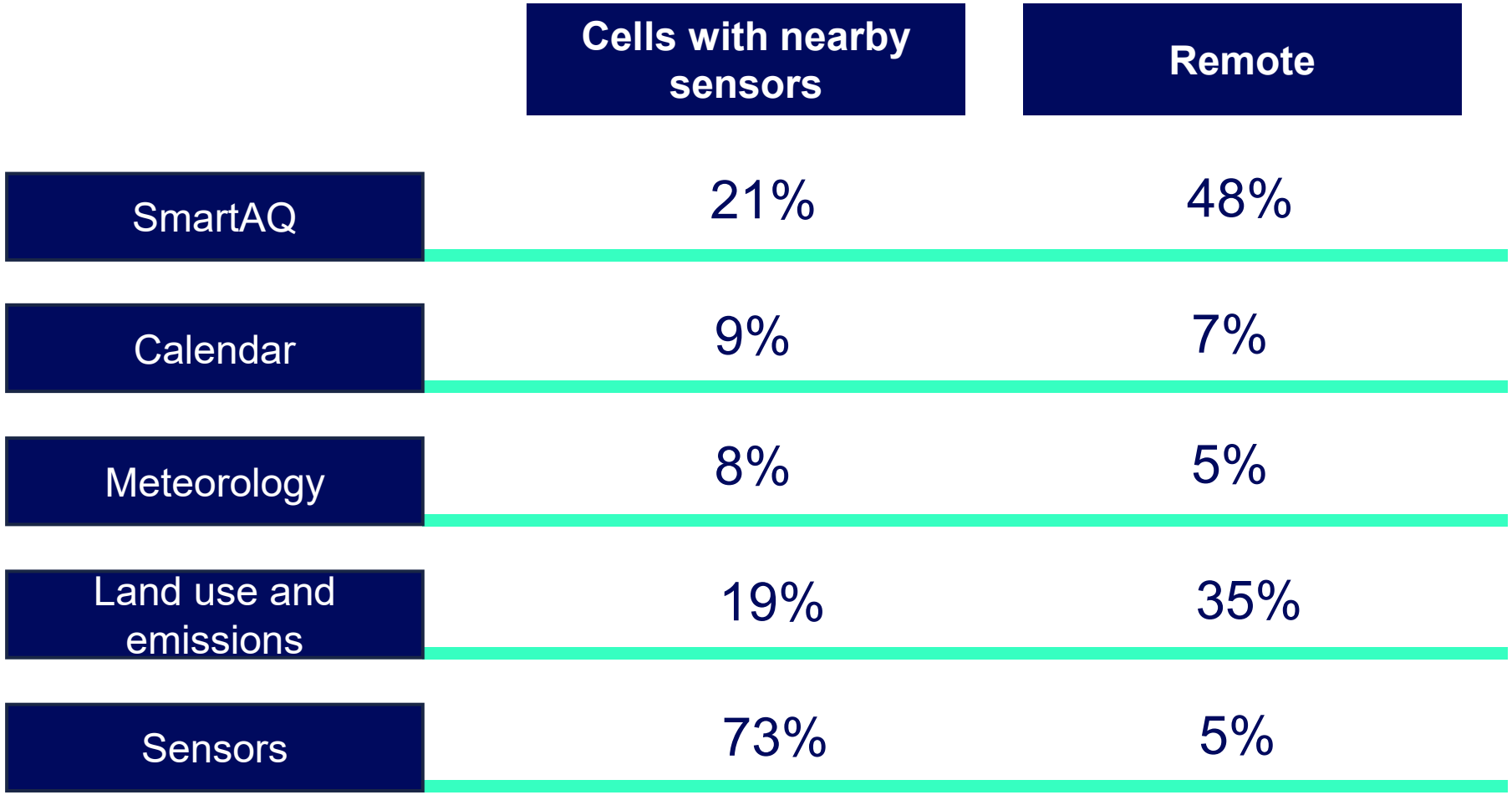
(recast)

Daily average $PM_{2.5}$ limit at 25 µg m^{-3}

SmartAQ+ performance in predicting daily $PM_{2.5}$ limit exceedance at all evaluation sites

SmartAQ+ and SmartAQ accuracy in predicting daily $PM_{2.5}$ limit exceedance at the 6 most polluted sites

ML model input importance based on SHapley Additive exPlanations (SHAP)



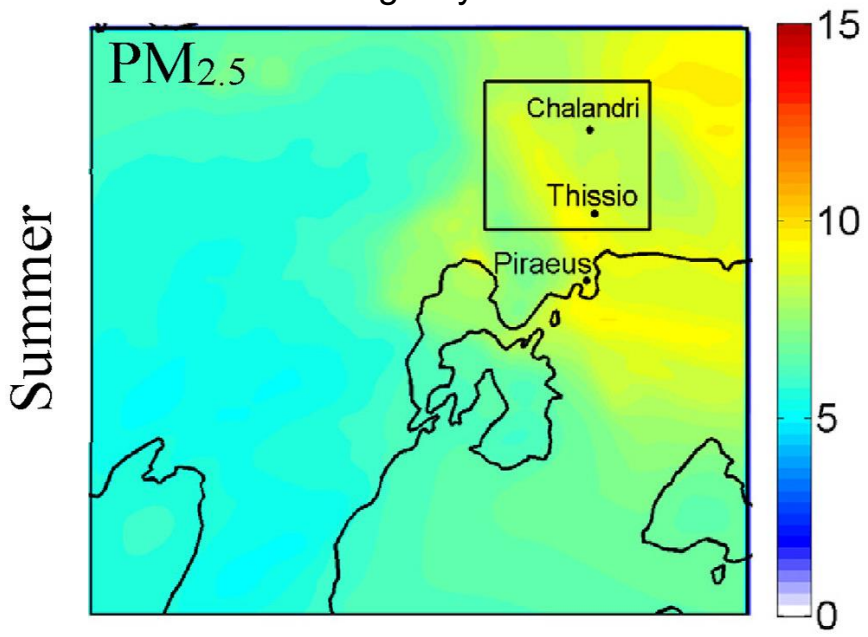
Summary of the results

- Localized real-time data helped SmartAQ+ improve the accuracy of $PM_{2.5}$ estimations
- SmartAQ+ characterized spatial variability in pollution fields better than SmartAQ
- SmartAQ+ relies more on SmartAQ predictions in regions with sparse or no sensor coverage
- SmartAQ+ correctly identified more daily $PM_{2.5}$ limit exceedance events and produced fewer false positives and missed events compared to SmartAQ

Future

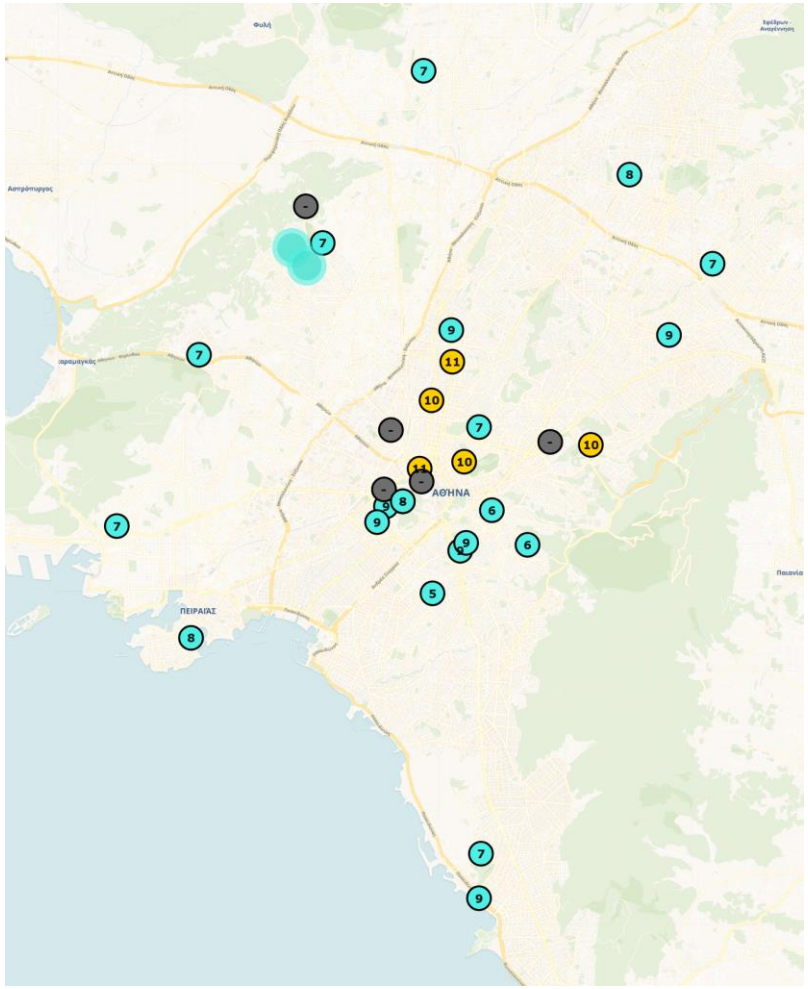
Deployment in Athens

Average SmartAQ-predicted ground concentrations (μg m^{-3}) of $PM_{2.5}$, during July 2019



Siouti, Evangelia, et al. "High resolution source-resolved PM2. 5 spatial distribution and human exposure in a large urban area." Atmospheric Environment (2025): 121277.

Low-cost sensor network in Athens



Deployment in other cities

Requirements for instant deployment

No ML model retraining needed

- SmartAQ model
- Low-cost sensor network
- Emissions
- Land use

Requirements for improvements based on historical data

ML model retraining

- SmartAQ model historical predictions (at least one year)
- Low-cost sensor network past measurements (at least one year)
- Any extra information on emissions (e.g. marine)
- Land use

Contributors

- Dr. Evangelia Siouti
- Dr. George Fouskas
- Prof. Spyros N. Pandis


Thank You

Extra Slides

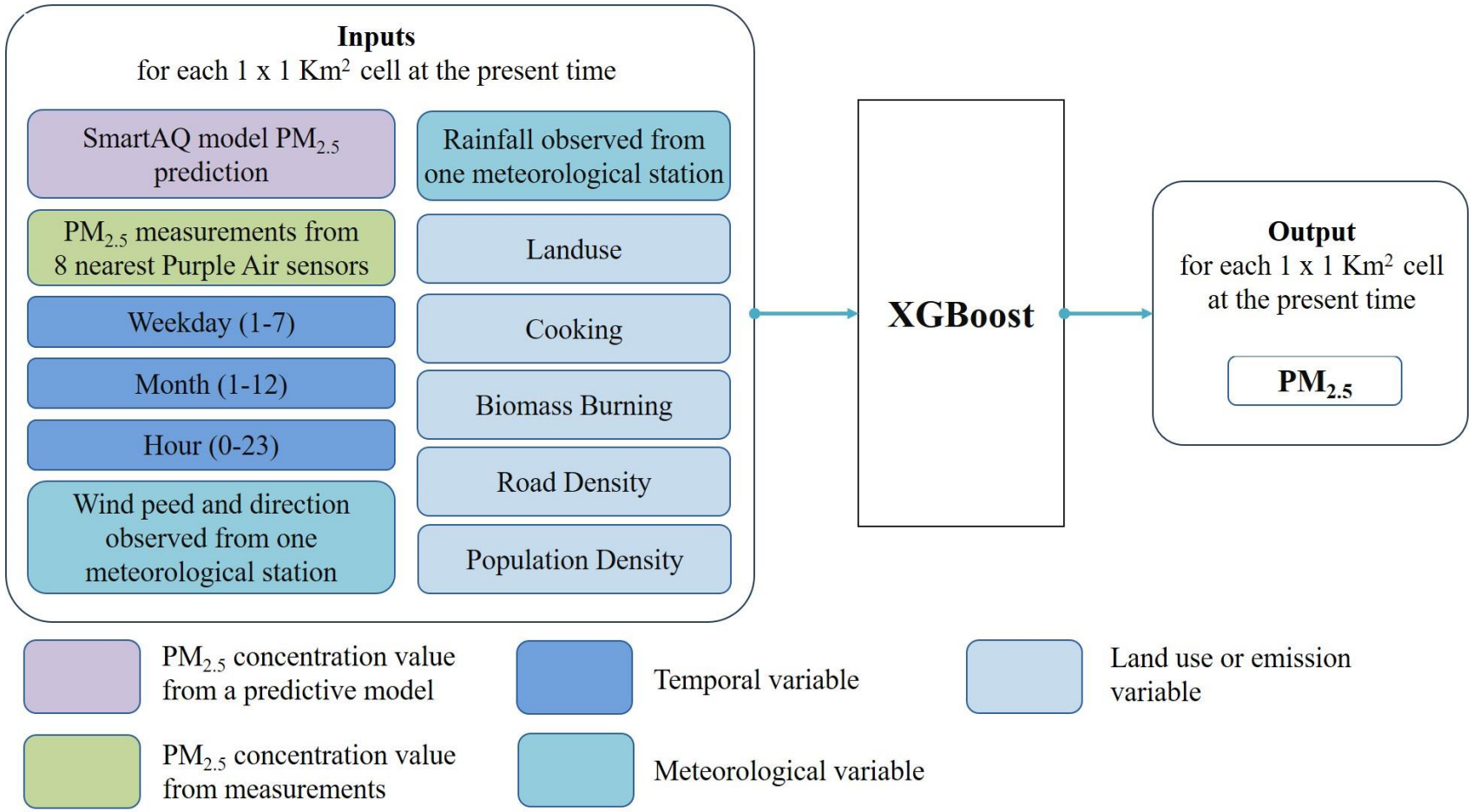
Limitations

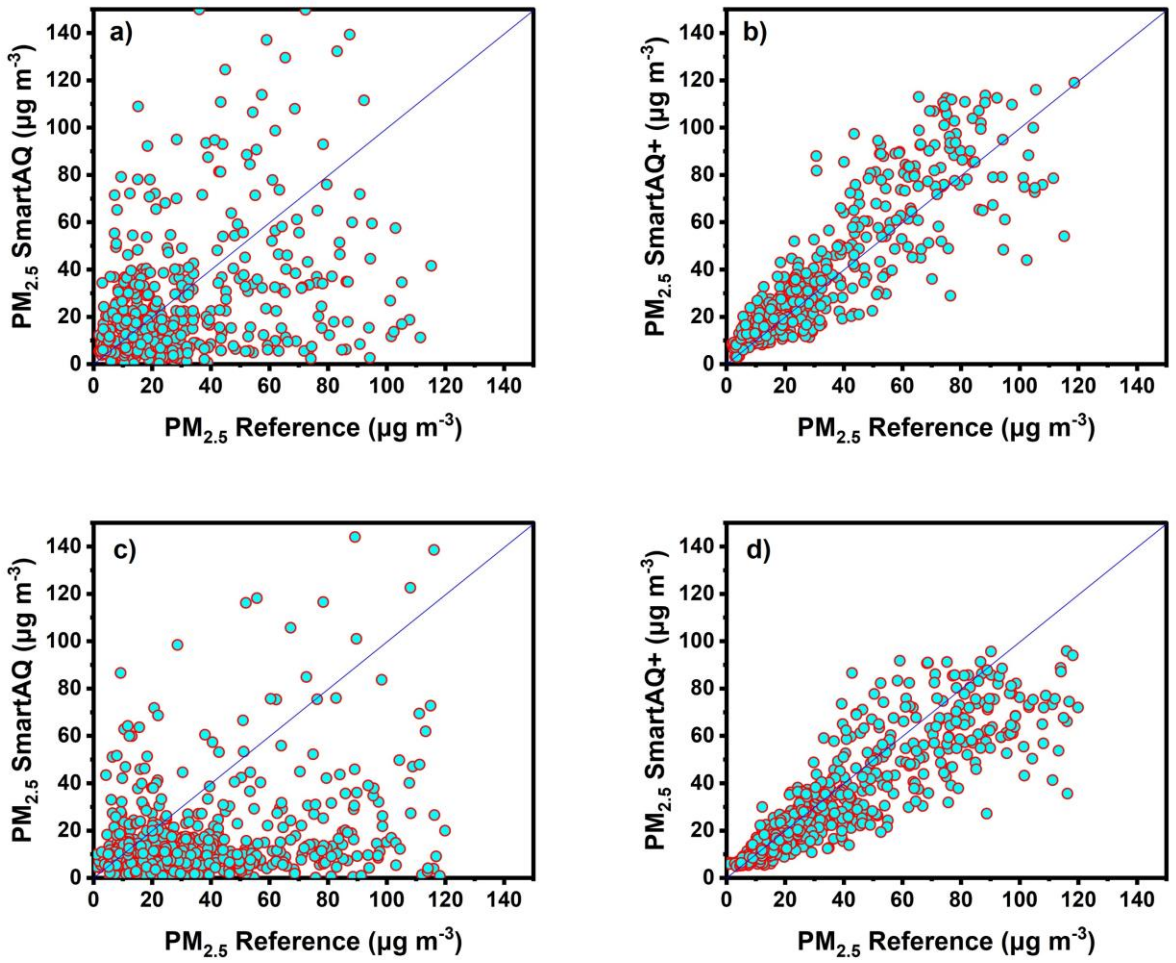
- Low-cost sensor performance
- Transferability – model might be overfitting to Patras conditions
- Rare weather extremes
- Works for the present time – not inherently able to forecast
- Inherits all SmartAQ limitations, which are reflected in its predictions (applies mainly to remote areas)

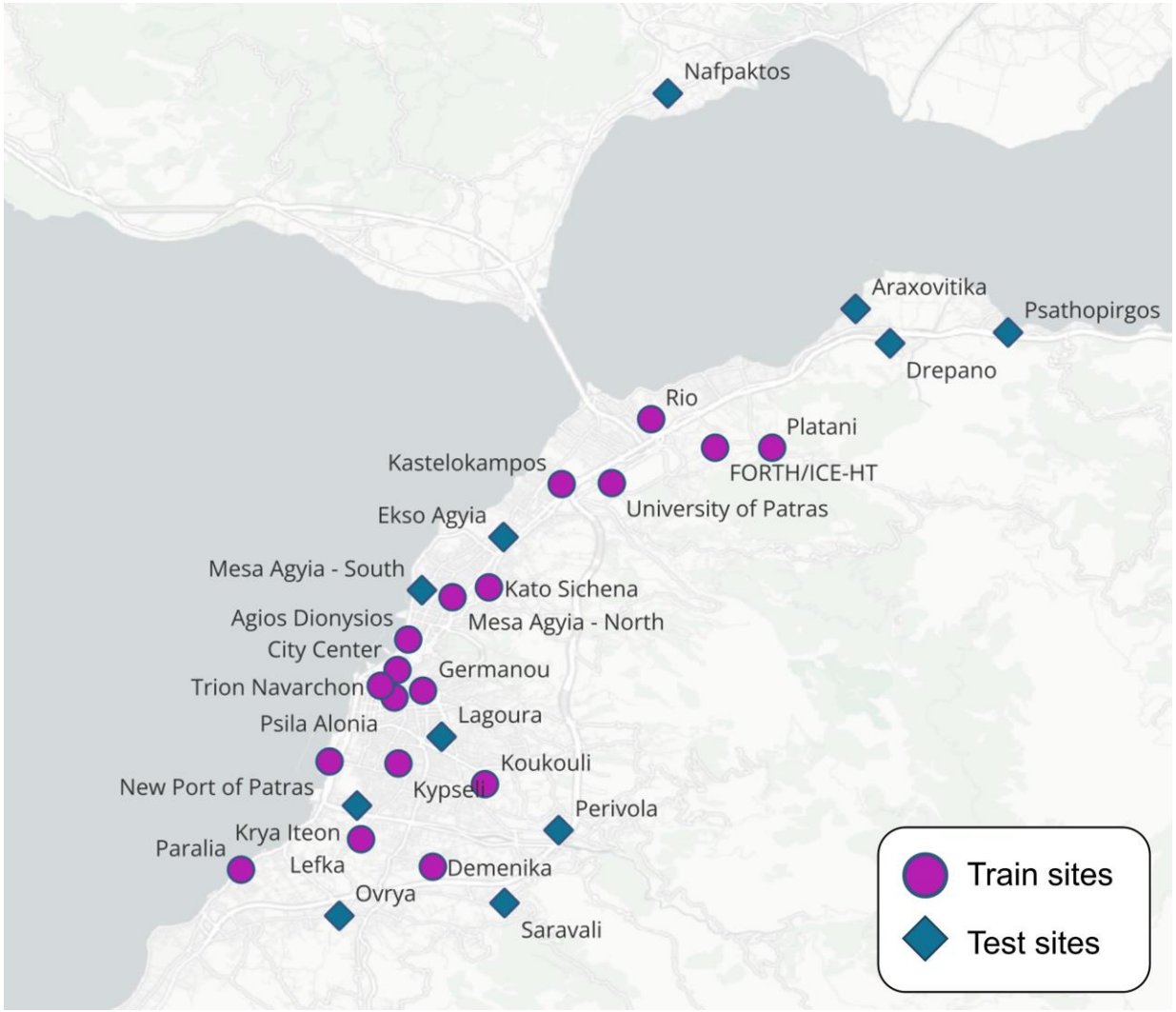
From the present to the future

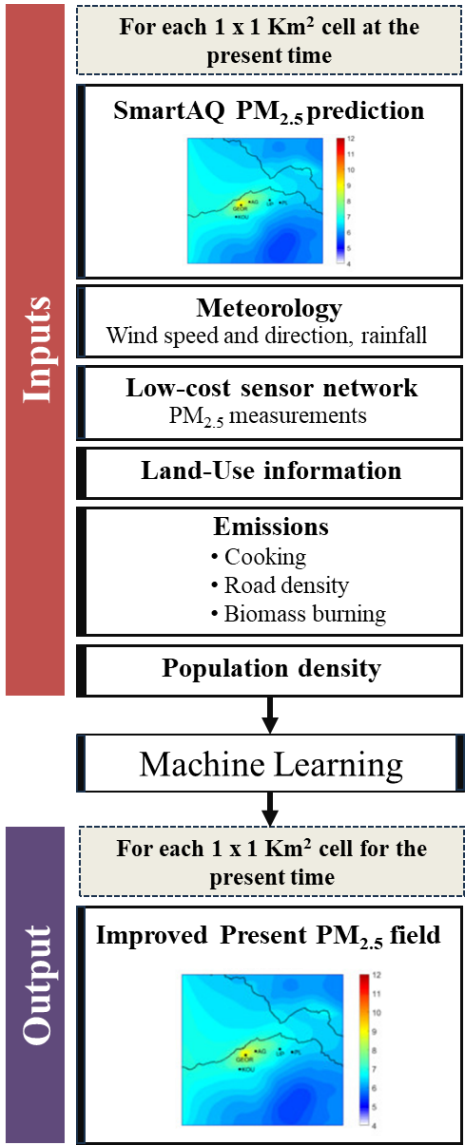
- Switch to an LSTM-based network, or Transformer (very hot)
- WRF forecasts will be necessary – extra biases
- Balance between historical observation and forecasts (ability to adapt ?)
- Computational resources
- Data demands


Legend
Empty
Sea
Urban
Agriculture
Rangeland
Deciduous forest

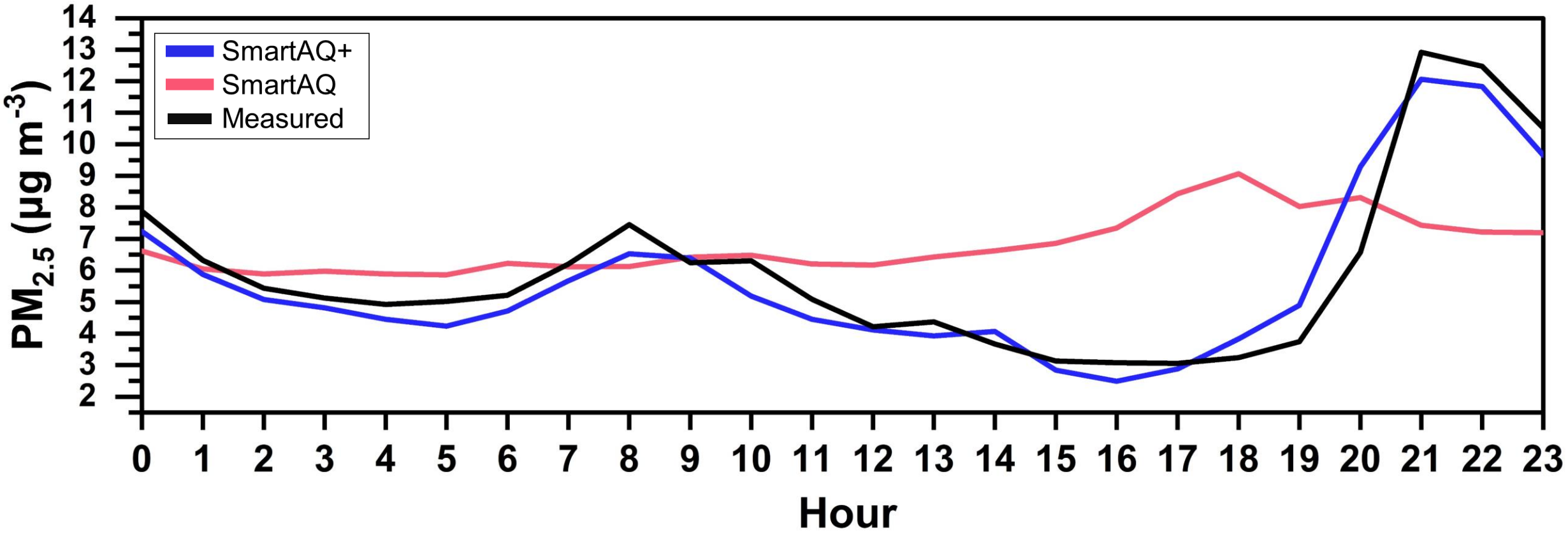

Inputs
for each 1 x 1 Km² cell at the present time
SmartAQ model PM2.5 prediction
PM2.5 measurements from 8 nearest Purple Air sensors
Weekday (1-7)
Month (1-12)
Hour (0-23)
Wind peed and direction observed from one meteorological station
Rainfall observed from one meteorological station
Landuse
Cooking
Biomass Burning
Road Density
Population Density
XGBoost
Output
for each 1 x 1 Km² cell at the present time
PM2.5
PM2.5 concentration value from a predictive model
PM2.5 concentration value from measurements
Temporal variable
Meteorological variable
Land use or emission variable


a)
b)
c)
d)
PM2.5 SmartAQ (µg m-3)
PM2.5 SmartAQ+ (µg m-3)
PM2.5 Reference (µg m-3)
0
20
40
60
80
100
120
140

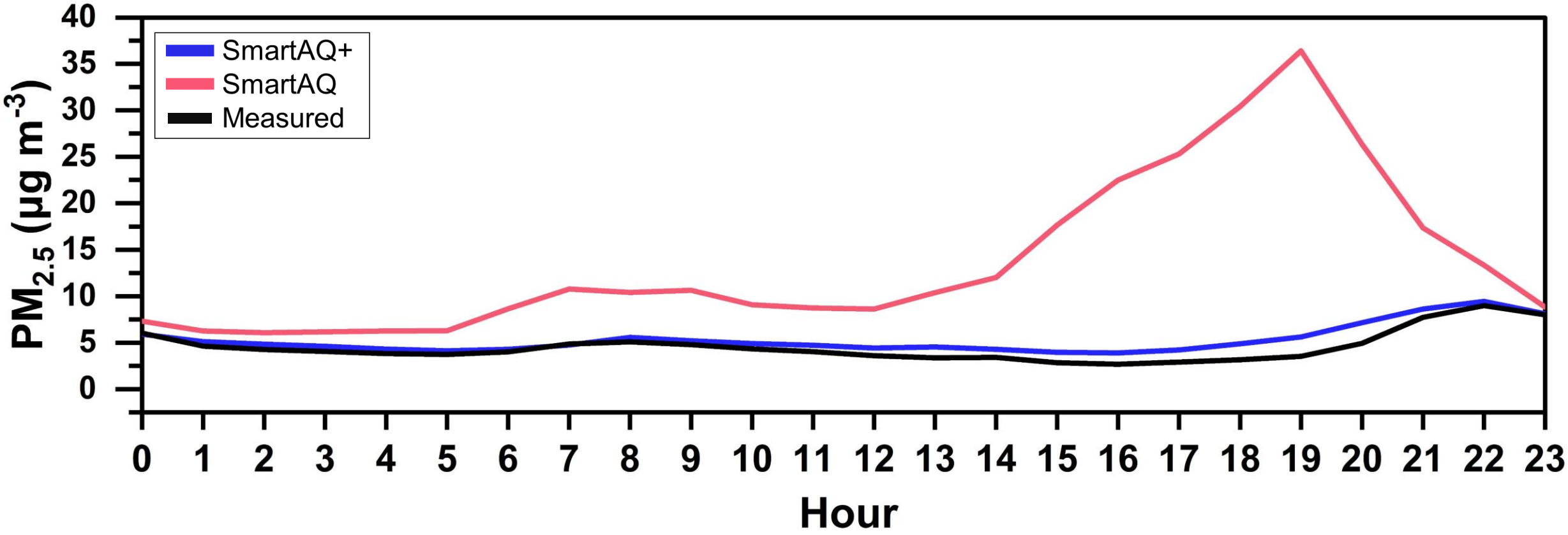

Nafpaktos
Araxovitika
Psathopirgos
Drepano
Rio
Platani
FORTH/ICE-HT
Kastelokampos
University of Patras
Ekso Agyia
Mesa Agyia - South
Kato Sichena
Mesa Agyia - North
Agios Dionysios
City Center
Germanou
Trion Navarchon
Psila Alonia
Lagoura
Koukouli
Kypseli
New Port of Patras
Perivola
Krya Iteon
Paralia
Lefka
Demenika
Ovrya
Saravali
Train sites
Test sites


Inputs
For each 1 x 1 Km² cell at the present time
SmartAQ PM2.5 prediction
Meteorology
Wind speed and direction, rainfall
Low-cost sensor network
PM2.5 measurements
Land-Use information
Emissions
• Cooking
• Road density
• Biomass burning
Population density
Machine Learning
Output
For each 1 x 1 Km² cell for the present time
Improved Present PM2.5 field

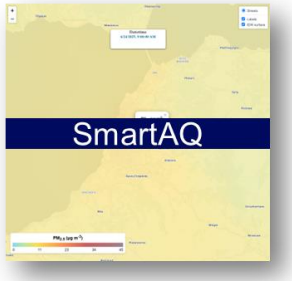
Average diurnal profile during April 2023 in Paralia



Average diurnal profile during April 2023 in the center of the city



Data processing

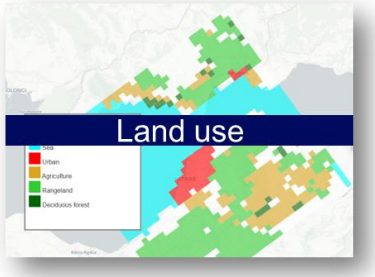


Scaling (standard)

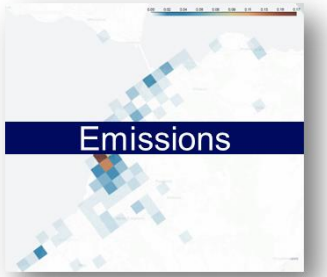
- Calibration
- Outliers (by hand)
- Channel selection (continuous)
- Scaling

- Wind speed (scaling)
- Wind Direction (from degrees to categorical)
- Precipitation (scaling)

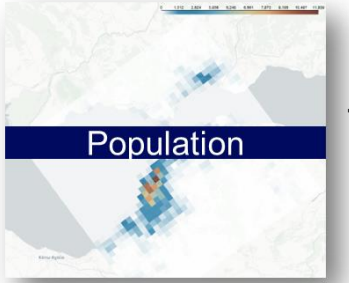
- Month (1-12)
- Weekday (1-7)
- Hour (1-24)



- Sea, Urban, Agriculture, Rangeland, Forest, Uncategorized (6 categories)
- Label encoding (e.g. Sea = 1)



- Cooking, Biomass burning as a percentage of total cell's emissions (scaling)



- Habitants per cell (scaling)