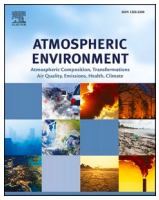
Atmospheric Environment 364 (2026) 121665

Contents lists available at ScienceDirect

Atmospheric Environment

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



Improving real-time high-resolution estimates of $PM_{2.5}$ concentration fields in urban areas by the SmartAQ+ system with data fusion and machine learning

Ioannis D. Apostolopoulos [a], Evangelia Siouti [a,b], George Fouskas [a], Spyros N. Pandis [a,b,*]

[a] Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Greece
[b] Department of Chemical Engineering, University of Patras, Patras, 26504, Greece

HIGHLIGHTS

- SmartAQ + improves SmartAQ's estimations of $PM_{2.5}$ with machine learning.
- It reduced mean error, bias, and fractional error by ~50 % versus SmartAQ baseline.
- SmartAQ + detected 70 % of $PM_{2.5}$ daily limit exceedances, tripling SmartAQ's accuracy.
- SmartAQ + performs well in locations without sensor data.

ARTICLE INFO

Keywords:
Fine particle matter
Machine learning
Data fusion
Chemical transport model

ABSTRACT

Monitoring $PM_{2.5}$ (mass of particles with diameter less than 2.5 μm) concentrations is challenging due to the limited number of ground-level monitoring stations and the limitations of existing modeling approaches. This study introduces SmartAQ+, a data fusion model that combines a chemical transport model-based system (SmartAQ) with low-cost sensor measurements and machine learning (ML) to enhance high-resolution $PM_{2.5}$ estimations at the present-time at a 1×1 km^2 scale. SmartAQ+ integrates real-time data from low-cost $PM_{2.5}$ sensors, weather stations, and land-use information to improve the accuracy of present-time $PM_{2.5}$ estimations at all locations in an urban area. SmartAQ+ demonstrated superior performance compared to SmartAQ that does not use real-time measurements in estimating the present-time $PM_{2.5}$, reducing the corresponding mean error, fractional bias (FBIAS) and fractional error (FERROR) by a factor of two. SmartAQ+ correctly identified 132 out of 190 $PM_{2.5}$ exceedance events of the daily limit of 25 μg m^{-3}, compared to SmartAQ's 34, while reducing false positives by a factor of 2 and missed events by a factor of 3. The performance gains depended on the availability of nearby sensors. In data sparse zones and during unusual events the model can inherit biases from the chemical transport model and can underestimate extremes. The study highlights the potential of data fusion models to address the limitations of standalone approaches, offering more precise air quality estimations in areas of a city in which there are no measurements.

1. Introduction

According to the World Health Organization (2018), 92 % of the global population is exposed to pollutant levels that exceed the air quality standards considered safe for human health. $PM_{2.5}$ (mass of particles with diameter less than 2.5 μm) pose a major health risk. Exposure to $PM_{2.5}$ can lead to various health problems, including a higher risk of heart disease, greater chances of heart attacks and strokes, impaired lung development, and an increased likelihood of developing lung diseases (Landrigan et al., 2018).

The limited number of ground-level air quality monitoring stations restricts our capacity to monitor spatiotemporal variations in pollutant concentrations (Tang et al., 2024). Several air quality forecasting models have been developed and assessed for their performance on hourly, daily, and seasonal timescales. Both statistical methods and chemical transport models (CTMs) have been used for air quality assessment and

* Corresponding author. Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Greece.
E-mail address: spyros@chemeng.upatras.gr (S.N. Pandis).

https://doi.org/10.1016/j.atmosenv.2025.121665
Received 11 July 2025; Received in revised form 30 October 2025; Accepted 8 November 2025
Available online 10 November 2025

forecasting (Delle Monache et al., 2020; Kaya and Gündüz Öğüdücü, 2020; Kukkonen et al., 2012). Statistical models are based on historical air quality and meteorological data and require minimal computational time because they do not account explicitly for pollutant emissions and atmospheric processes (Hamill and Whitaker, 2006; Pappa and Kioutsioukis, 2021). CTMs simulate air quality by modeling chemical and physical processes occurring in the atmosphere. These models do not require historical data and can offer insights into pollutant sources, whether local or resulting from long-range transport, as well as the formation of secondary pollutants (Zhang et al., 2012). CTMs depend on detailed information about pollutant sources, including industrial emissions, traffic, biomass burning, and other human activities, to produce their forecasts. However, emissions inventories can be incomplete, outdated, or inaccurate, leading to errors and biases (Hua et al., 2024). CTMs are also limited by uncertainties in meteorological forecasting. Since air quality is heavily influenced by weather conditions, such as temperature, wind speed, and precipitation, errors in weather forecasts propagate into the air quality predictions.

Grid resolution plays a crucial role in CTM studies targeting major urban areas, as sources like on-road traffic, commercial cooking, and biomass burning often exhibit steep gradients at the urban scale (Allan et al., 2010; Lanz et al., 2007). High-resolution pollutant concentration predictions enable better exposure assessments, facilitating comparisons among subpopulations within a metropolitan area (Wolf et al., 2020).

Machine learning (ML) models have demonstrated good performance in air pollution modeling due to their ability to capture complex nonlinear relationships among air pollutant concentrations and various predictors, including satellite data, meteorological variables, and land use information (Arowosegbe et al., 2022; De Hoogh et al., 2019; Lee et al., 2011; Tang et al., 2024). Support vector machines and artificial neural networks have shown promising results (Bai et al., 2016; Karimian et al., 2019; Prasad et al., 2016; Voukantsis et al., 2011; Zhou et al., 2019). Random forest (RF) ML algorithms have been used to estimate nitrogen dioxide (NO_2) and $PM_{2.5}$ levels across various regions, time periods, and spatial resolutions (De Hoogh et al., 2018; Stafoggia et al., 2019). Ensemble models have been developed to map air pollutant concentrations (Requia et al., 2020; Yu et al., 2022). Deep neural networks have been implemented to further enhance modeling performance (Li and Wu, 2021).

A key limitation of ML-based models is their reliance on historical data. Their predictive accuracy is heavily influenced by the quality, quantity, and diversity of the training data. In estimating $PM_{2.5}$ levels, historical data typically include past air quality measurements, meteorological conditions, and occasionally traffic and industrial activity patterns. This dependency makes ML models underperform in situations for which they have not been trained. For example, sudden shifts in meteorological patterns, changes in pollution sources (e.g., new regulations or industrial activities), or unprecedented events are challenging for ML models. Moreover, ML models, particularly those lacking domain-specific knowledge, treat data as abstract inputs without explicitly considering the physical and chemical processes that drive $PM_{2.5}$ formation, transport, and dispersion. ML-based predictions may also suffer from overfitting, where the model becomes overly specialized in the specific characteristics of the training data, leading to poor generalization in new conditions. Finally, increased model complexity, such as that introduced by large ensembles or deeper neural networks, can reduce interpretability and raise computational demands, often without yielding significant improvements in predictive performance (Kerckhoffs et al., 2019).

Researchers are increasingly exploring hybrid frameworks that integrate the strengths of CTMs and ML through data fusion techniques. This approach seeks to integrate multiple auxiliary inputs—such as real-time ground-based measurements, high-resolution emission inventories, satellite-derived observations (e.g., aerosol optical depth), real-time meteorological data, and land use characteristics—to improve the accuracy of the predictions.

Past studies have successfully used ML to transform coarse CTM outputs into fine-resolution (1×1 km^2) pollution maps. In a recent study (Dinkelacker et al., 2023) the authors developed an RF model to downscale PMCAMx predictions in southwestern Pennsylvania, achieving low bias and good performance across species and sources. However, their model had limitations in representing long-range pollution transport and relied solely on CTM output without observational correction. Another study (Bi et al., 2022) used RF to downscale global GEOS Composition Forecasting (GEOS-CF) predictions over China's Fenwei Plain to 1 km resolution, showing strong short-term forecast performance. Lv et al. (2021) applied ML models like RF and Support Vector Regression in China to reduce bias in $PM_{2.5}$ component forecasts, with high accuracy but limited urban-scale spatial testing.

More advanced frameworks like those by Malings et al. (2024) and Fang et al. (2023) combine CTMs' output with satellite and ground observations. Malings et al. (2024) introduced a modular, uncertainty-aware NO_2 prediction framework using satellite data and CTM outputs (GEOS-CF), achieving sub-city scale forecasts (~5 km). Fang et al. (2023) used an Ensemble Kalman Filter to merge ML predictions with a GEOS-Chem ensemble, reducing the root mean square Error (RMSE) and improving forecasting but at high computational cost.

Studies like those of Koo et al. (2023) in South Korea and Xu et al. (2021) in Shanghai demonstrated that ML can significantly improve short-term (6–48 h) $PM_{2.5}$ forecasts by correcting CTM biases using observations from the national regulatory network. Testing of these models indicated improved agreement with ground truth and reduction of false alarms.

Though regulatory-grade measurements are frequently used for model training and evaluation, limited site coverage reduces the generalizability and applicability of these models to all environments, especially in under-monitored or resource-limited areas. While some studies achieve 1 km spatial resolution, many remain at coarser scales (12–36 km). Few models are explicitly evaluated for their effectiveness in detecting pollution hotspots or extreme $PM_{2.5}$ events. Evaluation often focuses on overall statistical performance rather than spatial or temporal extremes.

This study proposes a hybrid approach (SmartAQ+) that combines the CTM-based SmartAQ system and ML to enhance the precision of $PM_{2.5}$ estimation fields. SmartAQ+ builds on the SmartAQ system, which uses PMCAMx and high-resolution meteorological data to provide forecasts at a 1×1 km^2 resolution.

SmartAQ's ability to capture pollution levels differs across city zones and seasons and varies from average to excellent (Siouti et al., 2022, 2023b). Errors are due to uncertainties in emissions, in the simulation of atmospheric chemistry and transport, but also to the fact that errors in the WRF model are inherited by SmartAQ (Pappa et al., 2023). Wind speed is systematically overestimated, particularly in colder months, influencing the accuracy of $PM_{2.5}$ predictions, as shown by the increased error at most monitoring stations during winter. Rainfall predictions are accurate about 50 % of the time (Pappa et al., 2023), which directly affects modeled wet deposition processes. Soil moisture is also consistently overestimated, which may alter atmospheric stability and humidity-driven chemistry near the surface.

SmartAQ+ enhances the SmartAQ's estimation for the present by incorporating real-time, localized data from nearby low-cost $PM_{2.5}$ sensors, city-wide weather stations, and land-use variables such as cooking, biomass burning, road density, and population density. SmartAQ+ differs from recent hybrid frameworks that fuse CTM output with observations. The low-cost sensor network is used to correct present-time $PM_{2.5}$ estimates from the CTM based SmartAQ system. SmartAQ+ also operates on top of a CTM that already resolves the urban domain at 1×1 km^2, so the ML component focuses on bias correction and spatial refinement of an existing fine resolution field.

2. Methods

2.1. Input data

2.1.1. SmartAQ system

The SmartAQ forecasting system utilizes six key models to predict weather conditions, anthropogenic, biogenic and marine emissions, pollutant concentrations, and pollutant sources (cooking, biomass burning, long-range transport, transportation, ships, etc.) (Siouti et al., 2023a, 2023b). It delivers three-day forecasts for major particle-phase pollutants (PM_1, $PM_{2.5}$, PM_{10}), gas-phase pollutants (e.g., NO_x, SO_2, CO, O_3, and volatile organic compounds) and the chemical composition and size distribution of aerosols. It also estimates the sources of all pollutants. The forecasts have an hourly time resolution. To enhance focus on the desired European urban area, three nested grids with progressively increasing spatial resolution are employed.

The SmartAQ system uses the WRF model (Skamarock et al., 2019) to compute key meteorological fields (e.g., cloud cover, precipitation, temperature and humidity) required for air quality predictions. For natural and terrestrial ecosystems, MEGAN (Model of Emissions of Gases and Aerosols from Nature) estimates gas and aerosol emissions, integrating WRF meteorological data with land use information (Guenther et al., 2006, 2012). Marine emissions, including sea salt and organics, are calculated using the O'Dowd and Monahan algorithms (Monahan et al., 1986; O'Dowd et al., 2008), which rely on WRF-predicted wind speeds over the sea surface. Anthropogenic emissions are derived from the TNO emission inventory (Kuenen et al., 2022), adjusted to the specific simulation day and month. To model air pollution within the target domain, the PMCAMx chemical transport model is applied, while the Particulate Source Apportionment Technology (PSAT) algorithm (Wagstrom et al., 2008) determines the contributions of individual sources to pollutant concentrations. Additional details are available in the study of Siouti et al. (2022).

In the first application of SmartAQ, the city of Patras, Greece, was used as a test case. The outer domain, which covers all Europe, has 36 × 36 km^2 horizontal spatial resolution and covers a region of 5400 × 5832 km^2, while the three nested domains, which are parts of Greece, regions of 276 × 276, 114 × 114 and 36 × 36 km^2, respectively (Fig. 1). In the vertical, PMCAMx uses 14 layers up to 10 km for all the modeling domains. The surface layer extends approximately up to 50 m.

2.1.2. Meteorology

Apart from the WRF predictions used in the SmartAQ system, we used real-time meteorological observations from a weather station

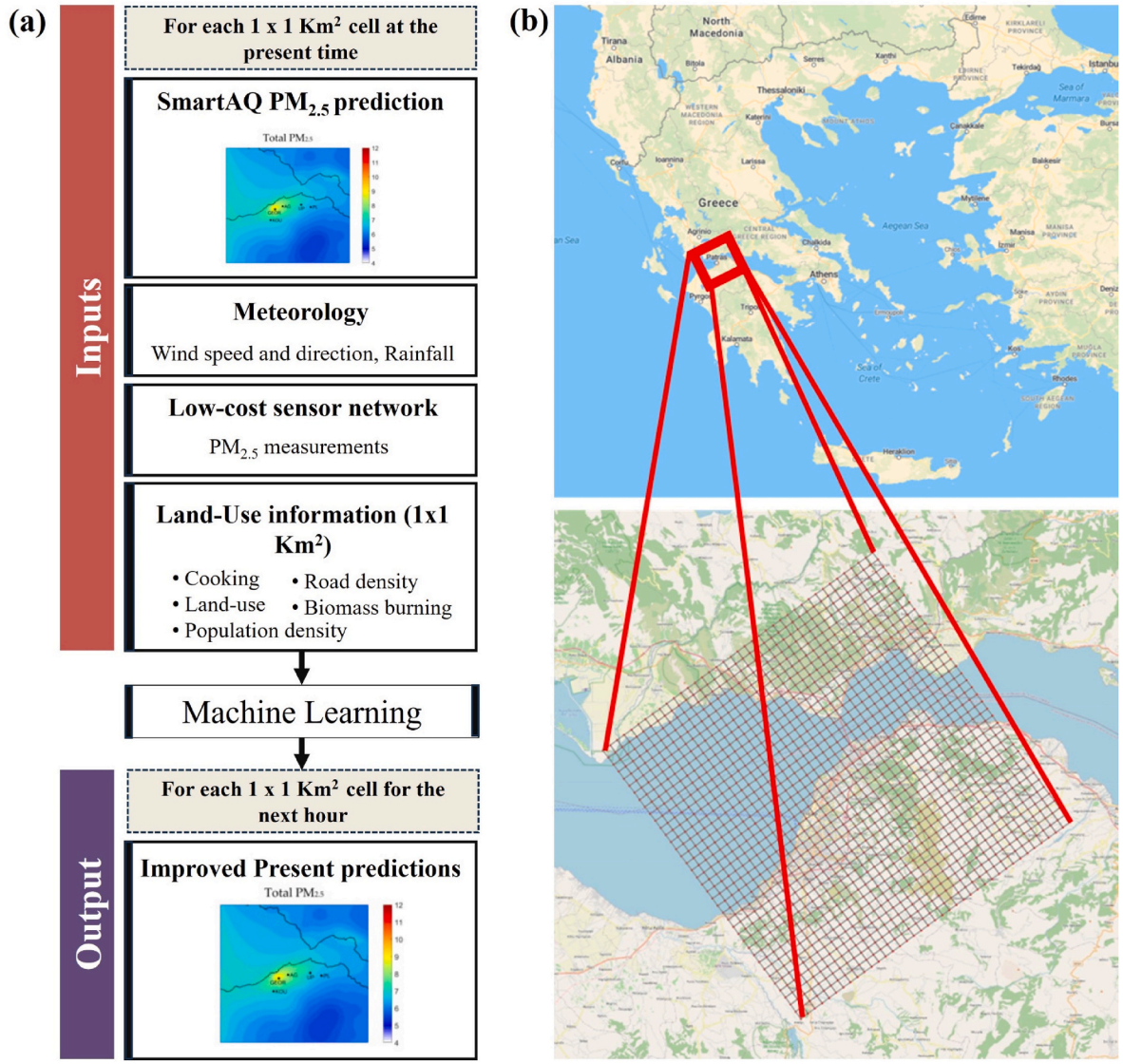


Fig. 1. (a) Overview of the ML methodology and (b) the study area.

located at the University of Patras in Rio as input to the ML model. The station provides hourly measurements of the wind speed, wind direction, and rainfall. At each hour, the measurements from the meteorological station at the University of Patras were paired with every grid cell and used as time varying predictors for the whole domain. These variables complemented the spatially resolved meteorological information, deriving from WRF, which was already embedded in the SmartAQ inputs.

2.1.3. Land use features

Although the TNO emissions inventory is already used as an input to SmartAQ, we also supplied spatial indicators of land use and emissions directly into the ML model. The ML does not have access to the internal information of the SmartAQ system and these spatially varying yet temporally invariant variables act as stable priors that guide the SmartAQ+ correction especially where sensor coverage is sparse.

The ML model's input includes information for each grid cell regarding its land use, population distribution, biomass burning and cooking emissions, and the percentage of the total road surface (Fig. 1). The United States Geological Survey (USGS) geographical datasets for topography and land use, which are also inputs to WRF, were utilized to classify the land use within the modeling domain. The land is categorized into sea, urban, agricultural, rangeland, forest, and uncategorized areas.

The population distribution was calculated based on population data provided by the European Union. The results were derived from the most recent database, the Eurostat census grid 2021 (2021). The data pertain to 2021 and are available in a 1×1 km^2 grid.

Residential biomass burning emissions at spatial resolution of 1×1 km^2 for the Patras area have been estimated by Siouti et al. (2023a). Their spatial distribution is based on the density of houses in the modeling domain of Patras. Cooking emissions are spatially distributed based on the density of the restaurants in each 1×1 km^2 grid cell of the modeling domain (Siouti et al., 2021).

The percentage of total road surface within each 1×1 km^2 grid cell was quantified using ArcGIS Explorer Desktop (ESRI). Road transport emissions are then spatially allocated according to this road surface fraction.

2.1.4. Low-cost $PM_{2.5}$ sensor network

A low-cost sensor network consisting of 29 Purple Air devices (PurpleAir PA-II) was used to provide the $PM_{2.5}$ concentration at fixed locations in the study area (Fig. 2). Thirteen devices were located in urban cells, thirteen in uncategorized, and three in rangeland cells (Table S1). We used measurements from January 1, 2021, to December 31, 2023 for this study. All measurements were averaged to 1 h. Each device contained two identical sensors (PMS5003). We used the average of the two sensors of each device. In cases where one of the sensors was constantly reporting extreme values (e.g. above 1000 μg m^{-3}) or too low (e.g. 0.5 μg m^{-3}) that sensor was excluded and the measurements of the other were used. The $PM_{2.5}$ sensor's measurements were corrected using the methodology suggested by Kosmopoulos et al. (2020). Their study suggested a linear calibration formula that reduces the measured $PM_{2.5}$ by approximately half. Using this correction reduced the hourly $PM_{2.5}$ relative mean error to 18 % (1.1 μg m^{-3}), with negligible bias.

This calibration was applied uniformly to all sensors for the full study window (January 1, 2021–December 31, 2023), and any future SmartAQ+ deployment should likewise ingest corrected sensor measurements to operate as intended.

2.2. Data processing

The output of SmartAQ was free of missing values, requiring no further imputation or processing. Temporal features, including the day of the week (represented as integers 1–7), the month (1–12), and the hour of the day (0–23), were added to the dataset to account for

Fig. 2. Low-cost sensor network in Patras.The 19 sites used for training/validation and the 10 sites used for testing are shown.

potential temporal variations in $PM_{2.5}$ concentrations. The above features were included as integers because the tree-based model can capture cyclic structure through splits on these variables.

Weather data processing involved converting wind direction, originally measured in degrees, into eight categorical bins corresponding to cardinal and intercardinal directions. These categories were defined as North, Northeast, East, Southeast, South, Southwest, West, and Northwest, enabling the incorporation of wind direction as a categorical variable in the model.

The contribution of cooking, biomass burning, and roads to the total emissions in individual computational cells ranged from zero to 17 %, 2 %, and 6 %, respectively. Population density values spanned from 0 to 11,809 inhabitants per square kilometer. These features were inserted as float values to let the ML algorithm choose the optimal split threshold. Additionally, land-use classifications were assigned integer labels corresponding to different categories: 0 for uncategorized, 1 for sea, 2 for urban, 3 for agriculture, 4 for rangeland, and 5 for forest.

No imputation or specific handling of missing values was performed for any type of data, as the ML model manages missing data during training.

2.3. Machine learning model

Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) is an advanced implementation of gradient-boosted decision trees that handles large and complex datasets (Ma et al., 2020). Its ability to build an ensemble of decision trees sequentially, where each new tree focuses on correcting the errors of the previous ones, makes it particularly effective in capturing complex patterns and relationships within the data. XGBoost is well-suited for the diverse and multifaceted features of the present application because its tree-based structure naturally handles different types of features without the need for extensive preprocessing or scaling. This allows the model to effectively capture the non-linear interactions between variables, such as the impact of land use on $PM_{2.5}$ levels during specific hours or the seasonal effects of biomass burning.

XGBoost demonstrates strong resilience to noisy and missing data, which is a common challenge when working with low-cost sensors that may occasionally provide unreliable measurements. Its built-in mechanisms for handling missing values ensure that the model remains stable even when some sensor data is unavailable (Liu et al., 2021). Understanding the relative impact of different factors on $PM_{2.5}$ predictions produced by XGBoost is possible though the quantification of the importance of each feature.

We used XGBoost with a regularizing hyperparameter setting guided by preliminary tuning on the training data, using the mean error as the loss function. The final configuration used 300 trees with a learning rate of 0.1, a maximum depth of 10, strong stochastic regularization through row subsampling of 0.2 and column subsampling per tree of 0.1, and L1 and L2 penalties of 1.0 and 0.8 to shrink complex trees and promote sparsity. All remaining parameters were left at their defaults.

During the development of SmartAQ+, we evaluated several other ML algorithms, including RF (Svetnik et al., 2003), Multi-Layer Perceptron (MLP), CatBoost, and Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017). XGBoost consistently outperformed them in terms of prediction accuracy and computational efficiency. Random Forest, for instance, had slower convergence and lower accuracy. MLPs required extensive tuning, while CatBoost and LightGBM were less effective.

The trained XGBoost model in SmartAQ+ operates every hour using the most recent $PM_{2.5}$ predictions from the SmartAQ system, the latest $PM_{2.5}$ measurements from the sensor network, and the rest of the variables to refine SmartAQ's prediction for each grid cell at the present-time (Fig. 1).

2.4. Use of sensor inputs in the ML model

Only the 7 nearest sensors and a complementary sensor located at a background site (Platani) are selected for input to the ML model when processing data for estimating the $PM_{2.5}$ concentration at each specific cell of the grid. If the nearest sensor is more than 4 km away, it is excluded from the input data. This exclusion ensures that the model only processes sensor data that is geographically relevant.

By using only the nearest sensors, the model better reflects the real-time conditions of the target cell, accounting for local environmental or operational factors that may influence the data. This approach also reduces the risk of irrelevant or outdated information from distant sensors skewing the model's predictions. If fewer than 7 sensors fall within the 4 km radius, the model uses only the available sensors, forcing itself to rely more on the other input variables. In essence, the latter methodology helps SmartAQ+ form its estimations at remote areas, where no sensor data are available and forces the model to rely more on the SmartAQ predictions and the background sensor for these areas.

2.5. Training and validation

We trained the SmartAQ+ model using data from 19 measurement sites (65 % of the total available sites), located at different cells, from January 1, 2021, to December 31, 2022. This group of sites is called training-validation group hereafter (Table S1). $PM_{2.5}$ measurements from this sensor group, from January 1, 2023, to December 31, 2023, were used for evaluation.

Additionally, a sub-network of 10 sites (Table S1) was selected as a complementary test set, called testing group hereafter. These devices were installed during 2022 in the area and were better suited for test purposes, because of the limited training data they could provide. They were hidden from the model at the training phase and their location differed from the location of the training-validation group. We used the testing group to provide the $PM_{2.5}$ concentration as an extra step in assessing the model's performance in "new" locations during 2023. For testing group of sensors, measurements from January 1, 2023, to December 31, 2023 were used.

For the evaluation of the present-time capability we executed retrospective simulations with model outputs at an hourly time step for the entire year 2023. We stored the resulting hourly concentration fields for the full annual cycle and compared the model values at the sensor locations with the temporally aligned corrected measurements.

2.6. Performance metrics

The mean error (ME), fractional bias (FBIAS), and fractional error (FERROR) were used to evaluate the model performance:

$$\mathrm{ME} = \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}} |P_i - O_{\mathrm{i}}| \tag{1}$$

$$\mathrm{FBIAS} = \frac{2}{N}\sum_{i=1}^{N} \frac{(P_i - O_i)}{(P_i + O_i)} \tag{2}$$

$$\mathrm{FERROR} = \frac{2}{N}\sum_{i=1}^{N} \frac{|P_i - O_i|}{(P_i + O_i)} \tag{3}$$

where, N is the total number of measurements, P_i is the predicted concentration and O_i is the corresponding reference concentration.

Based on the study by Morris et al. (2005), $PM_{2.5}$ model performance for daily average values is considered excellent for FBIAS $\leq \pm 15$ % and FERROR $\leq \pm 35$ %, good for FBIAS $\leq \pm 30$ % and FERROR $\leq \pm 50$ %, average for FBIAS$\leq \pm 60$ % and FERROR $\leq \pm 75$ %, while there are fundamental problems in the modeling system for higher FBIAS and FERROR.

The European Union has established a daily average limit value for $PM_{2.5}$ concentrations at 25 μg m^{-3}, which should not be exceeded more than 18 times per calendar year by 2030 (European Parliament, 2024). We assessed the performance of the SmartAQ and SmartAQ+ models in identifying days when $PM_{2.5}$ concentrations exceeded the EU-defined limit somewhere in the urban area. For each location where a sensor existed, we compared the number of times the model-predicted and sensor-observed daily average $PM_{2.5}$ levels surpassed the 25 μg m^{-3} threshold. This analysis aimed to determine the precision and reliability of both SmartAQ and SmartAQ+ systems in capturing exceedance events, which is critical for ensuring compliance with regulatory standards and informing public health advisories. A True Positive (TP) event is when both the observed and the predicted daily average $PM_{2.5}$ exceeded 25 μg m^{-3}. A False Positive (FP) when the model predicted $PM_{2.5}$ above the threshold, but the observed $PM_{2.5}$ concentration was below it. Under the same reasoning, we considered the True Negative (TN) and the False Negative (FN) cases. We computed the True Positive Rate (TPR) and False Positive Rate (FPR) of each model based on the following equations:

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

Practically, TPR is the proportion of days where the model correctly identified $PM_{2.5}$ exceedance when the observed concentration was above the 25 μg m^{-3} threshold and FPR is the proportion of days where the model incorrectly predicted $PM_{2.5}$ exceedance when the observed concentration was below the threshold.

3. Results

3.1. Performance in predicting average monthly patterns

We first used the monthly average predicted and measured $PM_{2.5}$ concentrations for evaluating SmartAQ+ for the test period (2023). Fig. 3 displays the average FERROR and FBIAS of the SmartAQ+ and SmartAQ models for 2023 for each location. Fig. 4 illustrates the mean monthly FBIAS of SmartAQ and SmartAQ+ at selected locations. Table S2 presents the SmartAQ+ evaluation metrics at each location. Based on the average FERROR and FBIAS for 2023, SmartAQ+ was excellent for 7 training-validation sites, good for 7, and average for 5. As far as the hidden 10 test sites are concerned, SmartAQ+ was excellent for 5, good for 2 and average for 3. The ME of SmartAQ+ was 2.1 ± 1 μg m^{-3} at the training-validation sites and 2.3 ± 1 μg m^{-3} at the 10 test sites. FBIAS was 9 ± 23 % and FERROR 34 ± 13 % at the training-validation sites and approximately the same at the test sites. There were two sites (Nafpaktos, Ovrya) where SmartAQ+ performance in terms of FBIAS and FERROR was significantly worse than the rest. Nafpaktos is a small city approximately 10 km from Patras (38° 23′ 39.0588″ N, 21° 50′ 4.9488″ E). The absence of nearby low-cost sensors for training had a negative impact on the performance of the SmartAQ+ model in that small city (FERROR = 59 %, FBIAS = 59 %). Ovrya is a suburban location approximately 7 km from Patras (38° 11′ 26.844″ N, 21° 43′ 45.12″ E). Its distance from the core of the low-cost sensor grid played a role, because the SmartAQ+ model relied on the SmartAQ model more and inherited its errors. SmartAQ's FERROR was 73 % and FBIAS 51 %. SmartAQ+ decreased these errors (FERROR = 70 %, FBIAS = −32 %) but its performance remained worse than the other sites. Out of the four sites affected by intense biomass burning, SmartAQ+ was excellent for one, good for two, and average for the remaining one. SmartAQ+ was excellent for 3 urban sites, good for 5, and average for the remaining 4 (Fig. S1). SmartAQ+ performance was classified as excellent and good for the two sites with intense cooking emissions.

SmartAQ+ performed better than SmartAQ at all training-validation and test sites (Table S5) based on monthly-averaged values. SmartAQ was good for 3 sites, average for 20, and below average for 6 (Fig. 3).

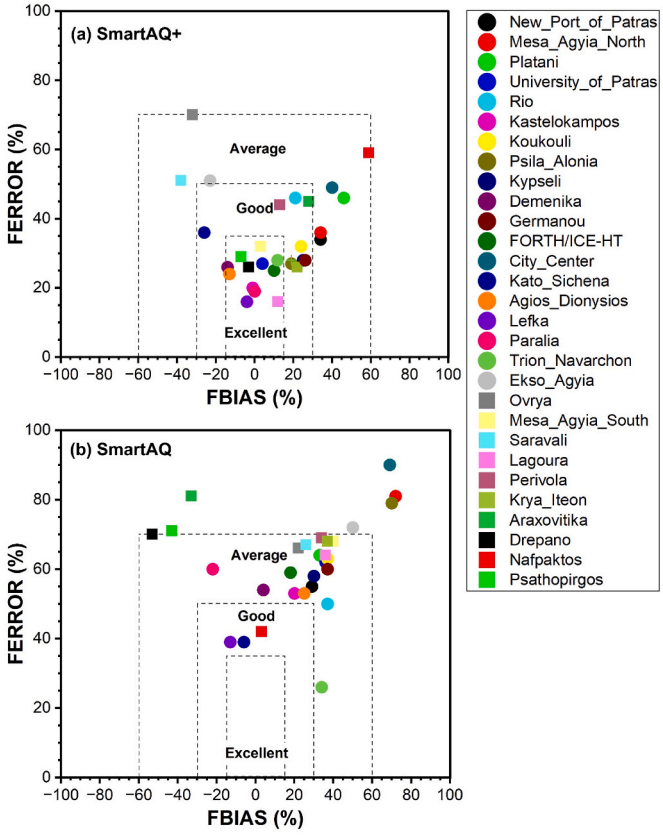


Fig. 3. (a) SmartAQ+ and (b) SmartAQ evaluation using FERROR (%) versus FBIAS (%) of monthly $PM_{2.5}$ concentrations for all sites in Patras during 2023. Circles denote training sites and squares denonte testing sites.

The average monthly FBIAS (Fig. 4) shows that SmartAQ+ reduced SmartAQ FBIAS by approximately 100 % for most months and sites, with the largest gains (Lagoura, Paralia, Mesa Agyia South, City Center) at the urban locations and limited improvement at the remote site of Nafpaktos.

We used the mean daily values for evaluating SmartAQ+ at all locations for January, April, July, and October 2023 (Table S3, Table S4). We selected January, April, July, and October 2023 to represent one month per season—winter, spring, summer, and autumn, respectively, as we expect different patterns and influencing factors across seasons. Fig. 5 shows average $PM_{2.5}$ predictions by the SmartAQ and SmartAQ+ models for January, April, July, and October 2023.

In January, SmartAQ+ estimated a mean $PM_{2.5}$ concentration of 16 μg m^{-3} at an area including the center of Patras and its east and west suburbs. The mean measured $PM_{2.5}$ concentrations of sensors located inside that area was 13 μg m^{-3}. SmartAQ predicted an average of 27 μg m^{-3} at the urban core and 6 μg m^{-3} at the suburbs. The sensors located in the city center (New Port of Patras, Psila Alonia, Kypseli, Germanou, City Center, Trion Navarchon, Agios Dionysios) measured on average 15 μg m^{-3}. SmartAQ+ produced more accurate concentration estimates within the urban core, whereas SmartAQ consistently overpredicted the $PM_{2.5}$ concentrations. Also, SmartAQ+ yielded correct estimates for suburban areas, in which the SmartAQ system underestimated $PM_{2.5}$.

For April, SmartAQ+ and SmartAQ predictions differ by a factor of two (~10 μg m^{-3}) near the city center and the south suburbs. The average estimated concentrations of SmartAQ+ for the urban core was 11 μg m^{-3}. Sensors in the urban core measured on average 8 μg m^{-3}. SmartAQ+ had a lower average FERROR (18 %) than SmartAQ (37 %) at these locations, based on sensor measurements. SmartAQ+ estimates

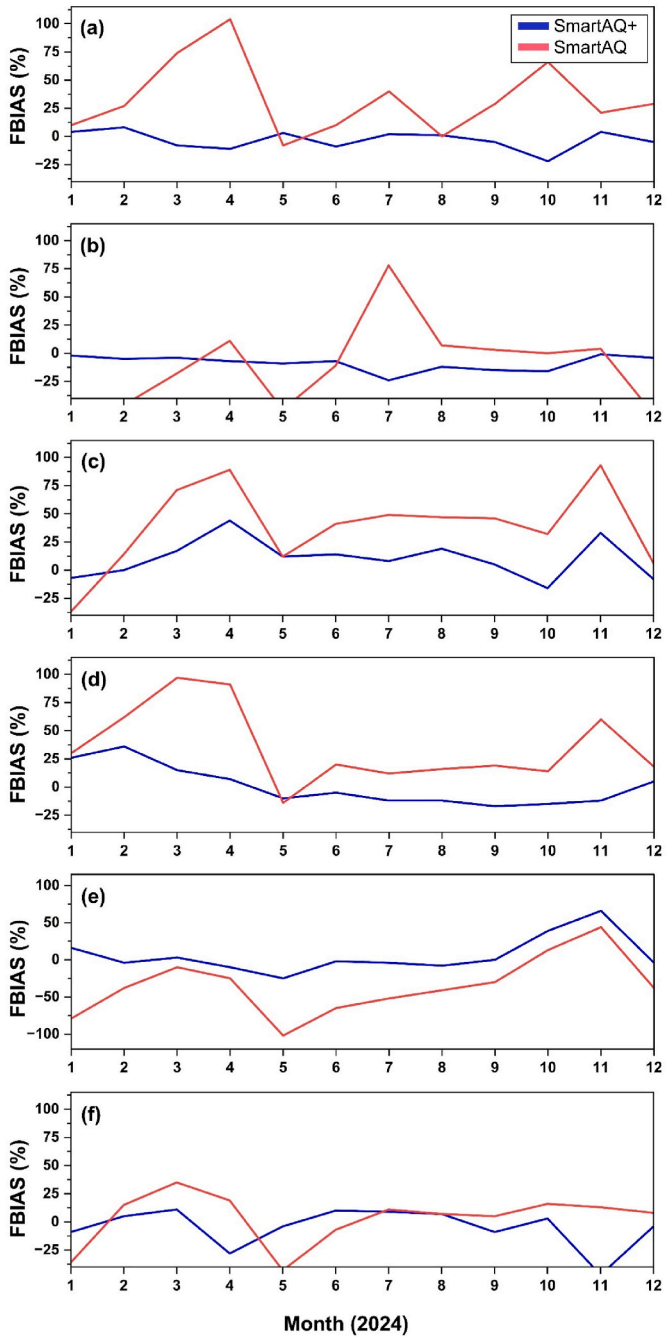


Fig. 4. Mean monthly FBIAS of SmartAQ and SmartAQ+ at a) Lagoura (test site), b) Paralia, c) Mesa Agyia South (test site), d) City center, e) Psathopirgos (test site), and f) Nafpaktos (test site).

differed from SmartAQ in background sites by an average of 4 μg m^{-3} (30 %). The sensor at Platani (background site) measured an average of 4 μg m^{-3} for April, while SmartAQ+ estimated 5 μg m^{-3} and SmartAQ 9 m^{-3}. In summary, SmartAQ+ halved SmartAQ's overestimation both within the urban core and at background sites, reducing the corresponding biases by 50 %.

For July, SmartAQ+ estimated an average of 8 μg m^{-3} in the urban core and SmartAQ 10 μg m^{-3}. The sensors inside that area measured on average 7 μg m^{-3}. At background sites, the predicted concentrations by the two models differ by 5 μg m^{-3}, with SmartAQ+ estimating on average 7 μg m^{-3}. The sensor at Platani measured on average 5 μg m^{-3}. At lower concentration levels in July, SmartAQ+ markedly reduced SmartAQ's positive bias, decreasing the urban-core FERROR from 23 % to 16 % and the background-site FERROR from 43 % to 28 %.

For October, SmartAQ+ estimated an average $PM_{2.5}$ concentration of 8 μg m^{-3} at locations near the city center, where sensors measured 7 μg m^{-3}. Similar to July, the SmartAQ system overestimated $PM_{2.5}$ with an average predicted value of 11 μg m^{-3}. At background sites, SmartAQ+ estimated on average 5 μg m^{-3} and the sensor at Platani measured 3 μg m^{-3}. SmartAQ predicted an average of 9 μg m^{-3} in Platani.

3.2. Performance in predicting average daily patterns

Fig. 6 illustrates the average diurnal $PM_{2.5}$ profiles for January 2023 at 6 selected sites (5 urban and 1 rural). Four of the sites (Mesa Agyia South, City Center, Lagoura, and Paralia) are located in Patras, where the majority of the study's sensors exist. Psathopirgos and Nafpaktos are locations 15 and 18 km away from Patras. Mesa Agyia South and Nafpaktos are testing sites and the rest are training sites.

During January, elevated $PM_{2.5}$ concentrations were measured by the sensors at all five selected urban sites after 17:00. This is due to intense biomass burning during winter in Patras (Kaltsonoudis et al., 2025). SmartAQ underestimated the concentrations during the afternoon and evening by a factor of 2. SmartAQ+ had a better performance at the four urban sites during the same hours, decreasing the FERROR of SmartAQ from 61 % to 23 %.

Fig. 7 illustrates the average diurnal $PM_{2.5}$ profiles for July 2023 at the same sites. Except for Nafpaktos and Psathopirgos, SmartAQ overestimates $PM_{2.5}$ concentrations during all hours by 2–4 μg m^{-3}. At Psathopirgos there is an underestimation of 5–7 μg m^{-3} by SmartAQ. SmartAQ+ manages to mitigate these biases with a mean error of 1–2 μg m^{-3} at all sites, except for Nafpaktos. At Nafpaktos, SmartAQ+ overestimates $PM_{2.5}$ by 6 μg m^{-3}.

3.3. SmartAQ + performance in predicting daily $PM_{2.5}$ limit exceedance

We evaluated the SmartAQ and SmartAQ+ models in detecting exceedance events by comparing their predictions with sensor measurements at training-validation and testing sites together. A TP occurred when both predicted and observed values exceeded the limit, while an FP occurred when only the model did.

Based on the sensors' measurements, the daily-average $PM_{2.5}$ limit (25 μg m^{-3}) during 2023 was exceeded for 10 or more days at 12 of the study's sites (Table S6). At these sites, the total number of exceedances was 190. SmartAQ+ identified correctly 132 events and SmartAQ 34. On the other hand, SmartAQ+ had 67 false-positive cases and SmartAQ 152. SmartAQ+ missed 56 events and SmartAQ 134. Fig. 8 illustrates the TPR and FPR values of SmartAQ and SmartAQ+ at the 12 most polluted locations. SmartAQ+ exhibits significantly better TPR and FPR rates at all locations.

We selected the two most polluted locations (Kypseli and Lefka) and we examined the hourly-averaged measurements and model predictions during the exceedance days (Fig. 9). SmartAQ+ showed better performance in capturing the $PM_{2.5}$ higher concentrations (>25 μg m^{-3}) at both sites. At Lefka, SmartAQ+ slightly underestimated the $PM_{2.5}$ levels (Fig. 9d).

We also inspected the average observed and predicted $PM_{2.5}$ concentration during Fat Thursday (February 16) when people grill and feast on large amounts of meat, emitting significant amounts of cooking organic aerosol across the entire city (Kaltsonoudis et al., 2017). Fig. 10 displays the average $PM_{2.5}$ on that day as predicted by SmartAQ+, SmartAQ and measured by multiple Purple Air sensors. The measured concentrations were above 35 μg m^{-3} at a large area covering the city center and the outskirts. SmartAQ estimated lower concentration values by 5 μg m^{-3} compared to the observed ones at a relatively smaller area (Fig. 10b) than the actual. SmartAQ+ improved the prediction by expanding the affected area (Fig. 10a) but the predicted concentrations were lower than the observed by 15–25 % at the suburbs of the city. The presence of real-time sensor measurements across the city played a

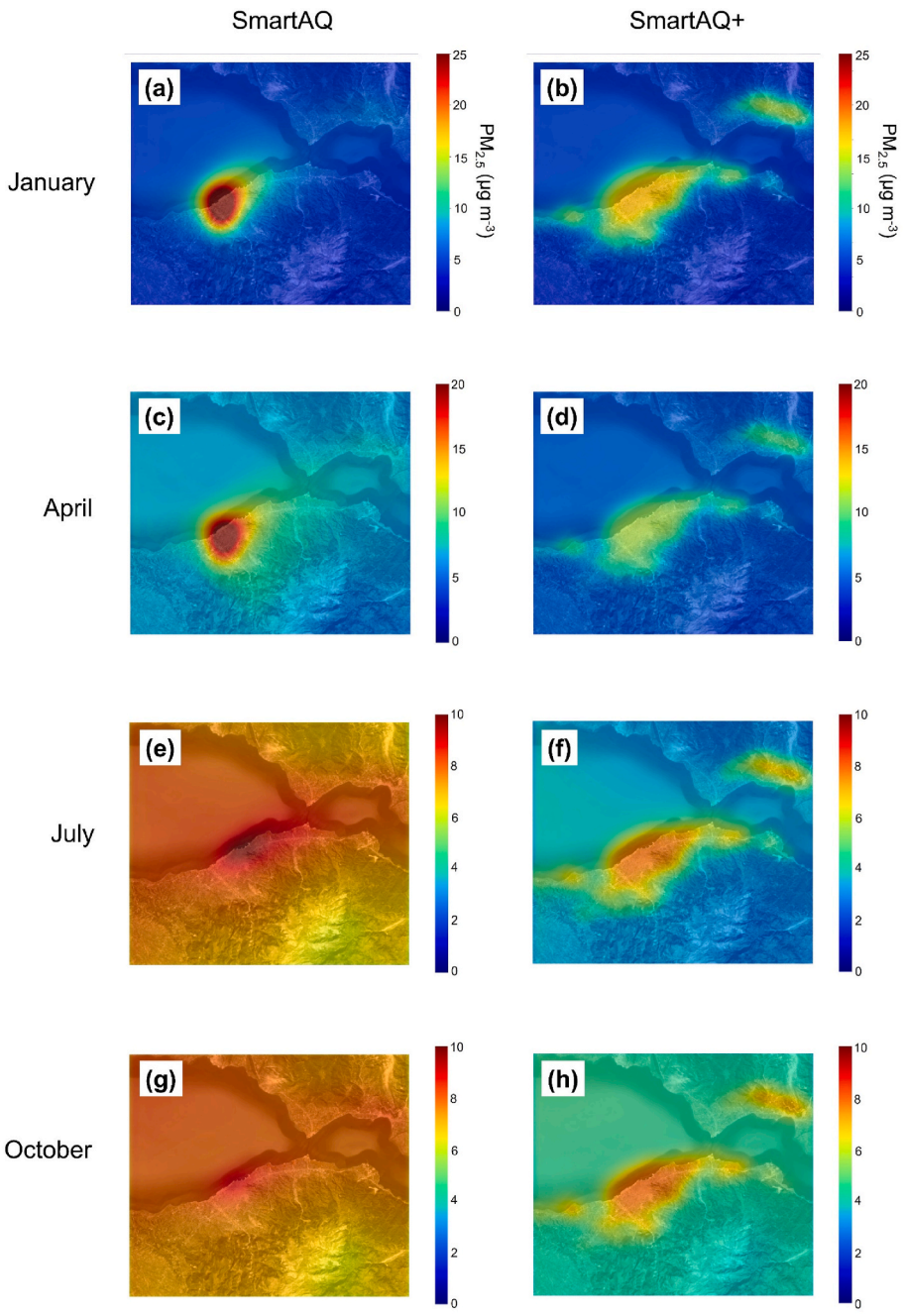


Fig. 5. Predicted monthly-average $PM_{2.5}$ (μg m^{-3}) by SmartAQ and SmartAQ+ for January, April, July, and October 2023.

significant role in adjusting the SmartAQ prediction by the SmartAQ+ system during this day. Finally, both models performed well (within 2 μg m^{-3}) in predicting the average $PM_{2.5}$ concentration (28 μg m^{-3}) at Nafpaktos.

3.4. Feature importance

We performed a SHapley Additive exPlanations (SHAP) analysis (Lundberg and Lee, 2017) on the SmartAQ+ model across the 36 × 36 km^2 geographical domain with a 1 × 1 km^2 resolution. SHAP ranks the ML input features by how much they move a prediction away from a baseline. It borrows Shapley values from game theory to fairly split the total prediction difference among the features, so each feature's score represents its relative importance.

The SHAP analysis was conducted for every grid cell using all available data test data from 2023. The feature importances varied depending on the presence of a reference sensor within a 4 km radius (Fig. 11). To investigate this effect, we categorized the SHAP importances into two groups: (i) grid cells with at least one nearby sensor within 4 km and (ii) grid cells without any nearby sensor in this radius (Table S7). The importance of the SmartAQ prediction was 21 % at cells where nearby sensors existed and 53 % at cells with no nearby sensor. The importance of land-use variables was 19 % for cells with nearby sensor(s) and 35 % for cells without nearby sensors. Similarly, the importance of sensor measurements was 43 % for cells with nearby sensors and zero for cells without.

4. Discussion and conclusions

This study developed and evaluated SmartAQ+, a hybrid approach

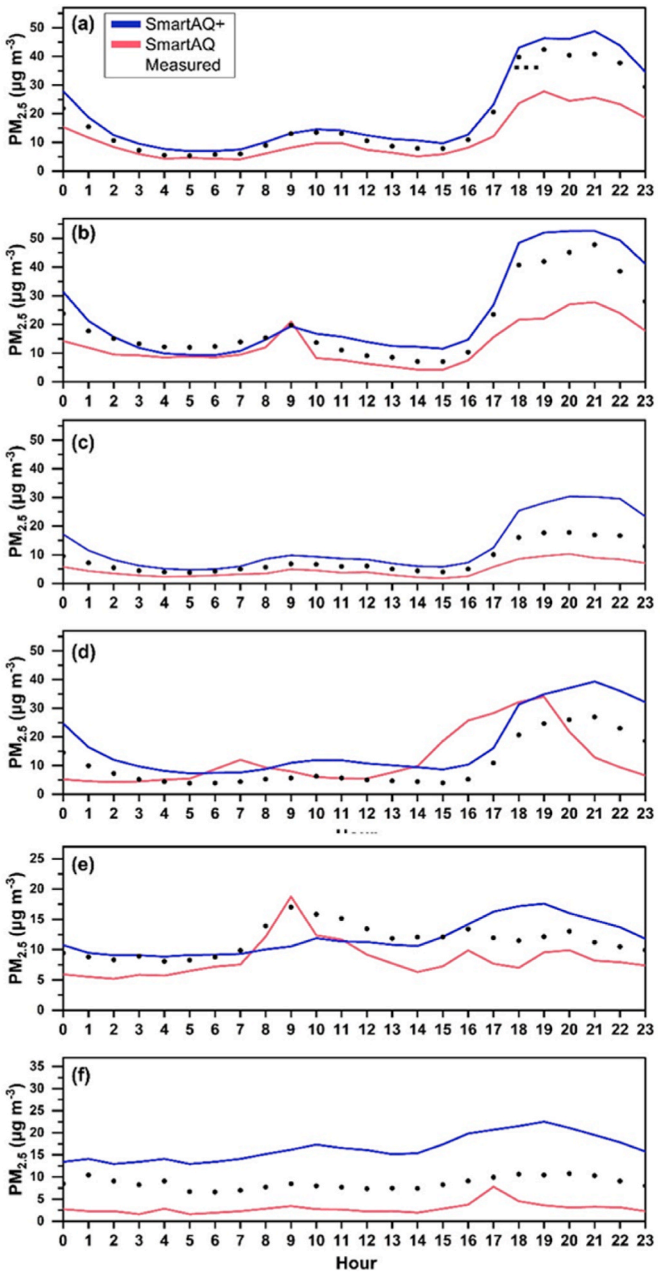


Fig. 6. Average diurnal $PM_{2.5}$ profiles for January 2023 at a) Lagoura (test site), b) Paralia, c) Mesa Agyia South (test site), d) City center, e) Psathopirgos (test site), and f) Nafpaktos (test site).

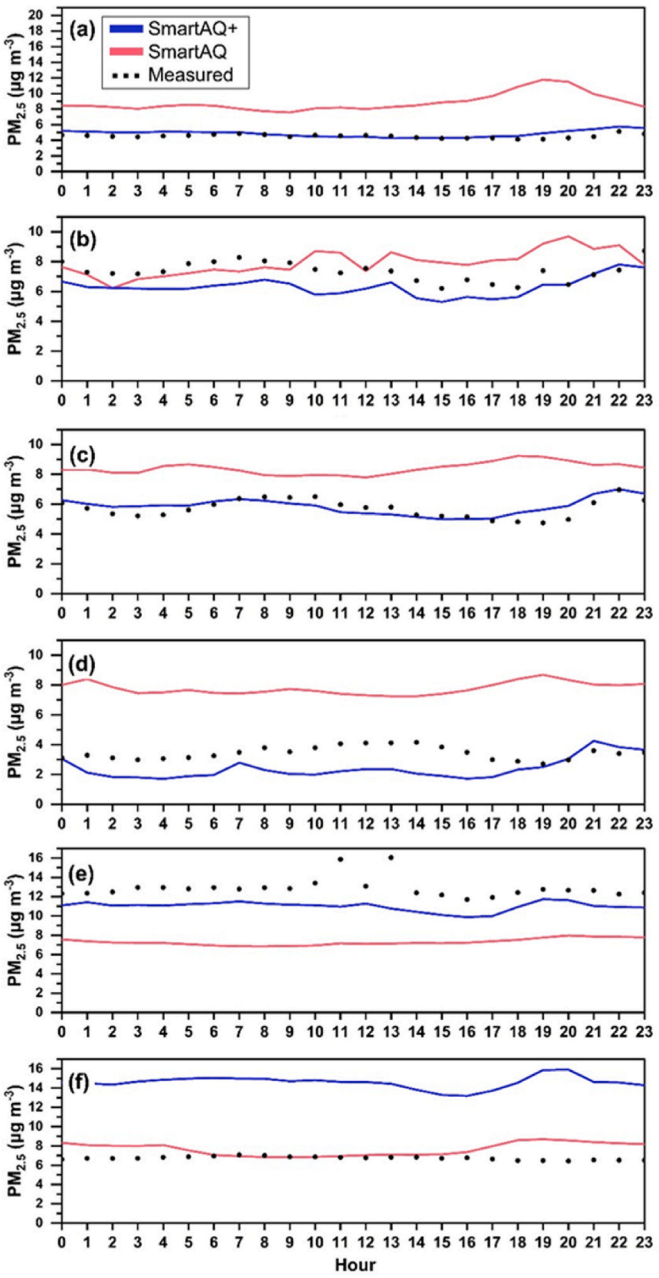


Fig. 7. Average diurnal $PM_{2.5}$ profiles for July 2023 at a) Lagoura (test site), b) Paralia, c) Mesa Agyia South (test site), d) City center, e) Psathopirgos (test site), and f) Nafpaktos (test site).

that integrates a CTM (SmartAQ) with ML to enhance the accuracy of high-resolution $PM_{2.5}$ estimations at the present time (or the past). Data from calibrated low-cost $PM_{2.5}$ sensors, one weather station, and land-use variables, helped SmartAQ+ improve the accuracy of $PM_{2.5}$ estimations at a 1×1 km^2 resolution.

Overall, SmartAQ+ performed significantly better than SmartAQ, especially during the winter, with substantial improvements in both error and bias metrics. At the training sites, SmartAQ+ reduced the average FERROR from 62 % to 38 % in winter and from 49 % to 21 % in summer. FBIAS also decreased in winter (from 38 % to 33 %) and in summer (from 39 % to 3 %). At the test sites, the average FERROR dropped from 74 % to 40 % in winter and from 55 % to 33 % in summer. SmartAQ+ showed a lower FERROR compared to SmartAQ at nearly all locations. For FBIAS, SmartAQ+ showed improvements at 12 out of 29 locations in winter and at 24 out of 29 locations in summer, highlighting its more robust performance during the summer.

Beyond general improvements, SmartAQ+ better characterized spatial variability in air pollution fields. This is evident in the seasonal maps, where SmartAQ+ consistently offered more localized and topographically relevant predictions. This difference is because SmartAQ+ integrates sensor data from multiple locations, including nearby areas, to refine its predictions. The SmartAQ model is unaware of the low-cost sensor measurements.

SmartAQ+ appears to rely more on SmartAQ predictions in regions with sparse or no sensor coverage, such as marine and rangeland areas compared to areas with available nearby sensor data. This pattern is particularly evident in Fig. 5b, where SmartAQ+ and SmartAQ display agreement in isolated locations while diverging in urban-adjacent cells.

SmartAQ+ correctly identified more daily $PM_{2.5}$ limit exceedance events and produced fewer false positives and missed events compared

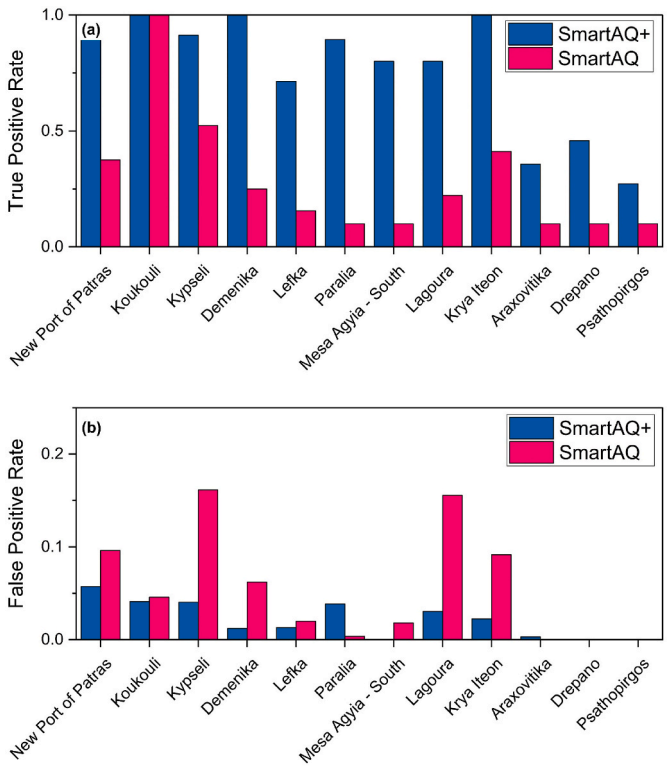


Fig. 8. True Positive (a) and False Positive (b) Rates of SmartAQ+ and SmartAQ in predicting days when the average $PM_{2.5}$ concentration exceeded the limit of 25 μg m^{-3} at multiple locations.

to SmartAQ. At the most polluted locations (Kypseli and Lefka), SmartAQ+ better captured high $PM_{2.5}$ concentrations and their temporal variations, though it slightly underestimated levels at Lefka. During a major cooking event in Patras, SmartAQ+ improved spatial predictions but underestimated concentrations in the outskirts. Both models performed well in less polluted areas. The improved present-time fields and the better identification of days above the European daily $PM_{2.5}$ limit support operational uses in urban air quality management. City services can use the 1 × 1 km^2 maps to target advisories and responses at the neighborhood scale and to inform early alerts for vulnerable populations. The exceedance detection skill can assist regulatory compliance checks and the same fields can guide the design and expansion of sensor networks by revealing persistent gaps in coverage.

The integration of real-time sensor data and land-use features into SmartAQ+ improved the SmartAQ predictions of $PM_{2.5}$ concentration fields. However, when comparing the two systems directly, it is essential to consider that SmartAQ relies on chemical transport modelling without real-time measurements or historical data.

SmartAQ+ performance depends heavily on the availability of sensor input. In remote areas like Nafpaktos and Ovrya, where the number of nearby low-cost sensors was limited, the model relied on SmartAQ's predictions and inherited its biases. This finding is further supported by the SHAP analysis, where the SmartAQ prediction was the dominant feature in data-sparse zones. A practical use of SmartAQ+ is to identify locations where new sensors would yield the largest reduction in uncertainty and to guide short term deployments that supply the references needed to evaluate and refine the corrections that are learned in sensor covered areas and then projected to unsampled cells.

The decision to constrain the model to seven sensor inputs per grid cell, including the Platani background site as the eighth sensor, was made to prevent a small group of nearby devices from dominating the feature space. We did not examine alternative configurations for the number of neighboring sensors or the 4 km distance threshold and future work should assess the effect of fewer sensors or adaptive radii on performance under different coverage regimes.

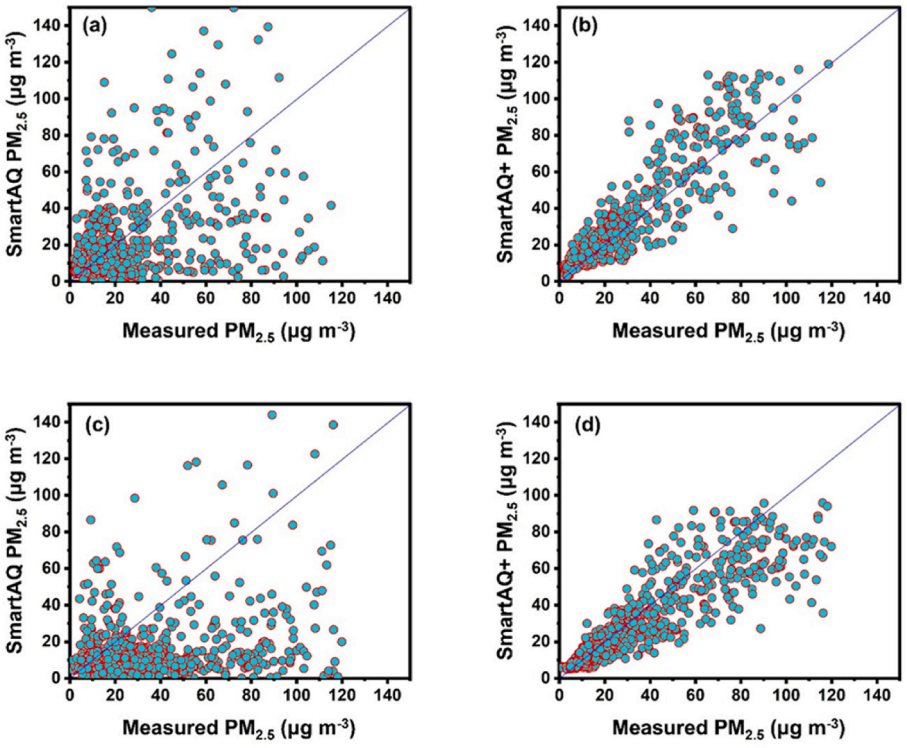


Fig. 9. Hourly $PM_{2.5}$ concentrations at Kypseli (training site) for (a) SmartAQ and (b) SmartAQ+, and Lefka (training site) for (c) SmartAQ and (d) SmartAQ+ during days when the daily average concentration was above 25 μg m^{-3}.

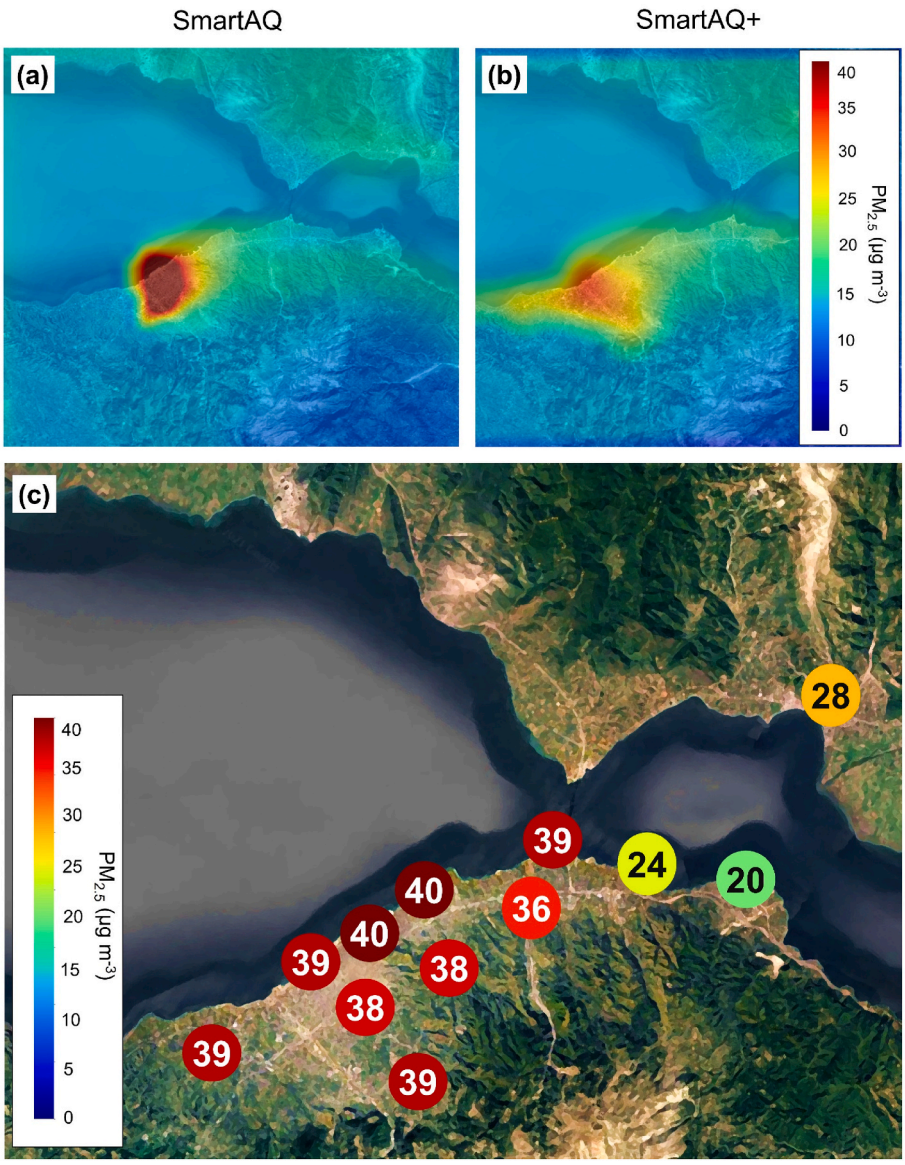


Fig. 10. Average $PM_{2.5}$ on February 16, 2023 as predicted by a) SmartAQ+, b) SmartAQ and c) measured by multiple Purple Air sensors. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Even after applying standard corrections, low-cost sensors can still drift or react differently under extreme pollution levels. That residual noise might mislead the machine-learning model, especially during sudden spikes, so it is worth exploring more dynamic calibration or online bias correction in future work. Training on two years of Patras data and testing on the third could mean the model learned some city-specific weather or seasonal patterns. To be sure SmartAQ+ works elsewhere (or under different climate years), it would help to train and validate across multiple cities or longer time spans. The algorithm is expected to transfer best to cities with similar emission patterns and concentration ranges. Application in regions with substantially higher pollution can lead to larger bias because extrapolation beyond the training domain is not reliable. Preliminary tests in Athens, the capital of Greece, indicate that the model yields reasonable present-time estimates there. These tests are exploratory and will be analyzed in future work.

We evaluated general exceedance counts and a couple of cooking-related spikes but have not yet assessed how SmartAQ+ handles rare weather extremes. Future tests should include those events to confirm the model stays reliable when conditions are most challenging.

While SmartAQ+ performs well in present-time $PM_{2.5}$ estimation, its architecture inherently ties it to historical observations and short-term trends. Attempting to extend its predictions to longer lead times could likely result in propagation of biases from SmartAQ due to errors in the underlying meteorological forecasts, emissions, or simulation of processes. Furthermore, adapting the ML framework to account for other pollutants, predictions of which are already produced by SmartAQ, would significantly broaden its utility in air quality forecasting. Lastly, implementing source apportionment techniques within the ML pipeline might suggest corrections to the SmartAQ source apportionment results.

CRediT authorship contribution statement

Ioannis D. Apostolopoulos: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Evangelia Siouti:** Writing – review & editing, Writing – original draft, Software, Methodology, Data curation. **George Fouskas:** Writing – review & editing, Investigation, Conceptualization. **Spyros N. Pandis:** Writing – review & editing, Supervision, Conceptualization.

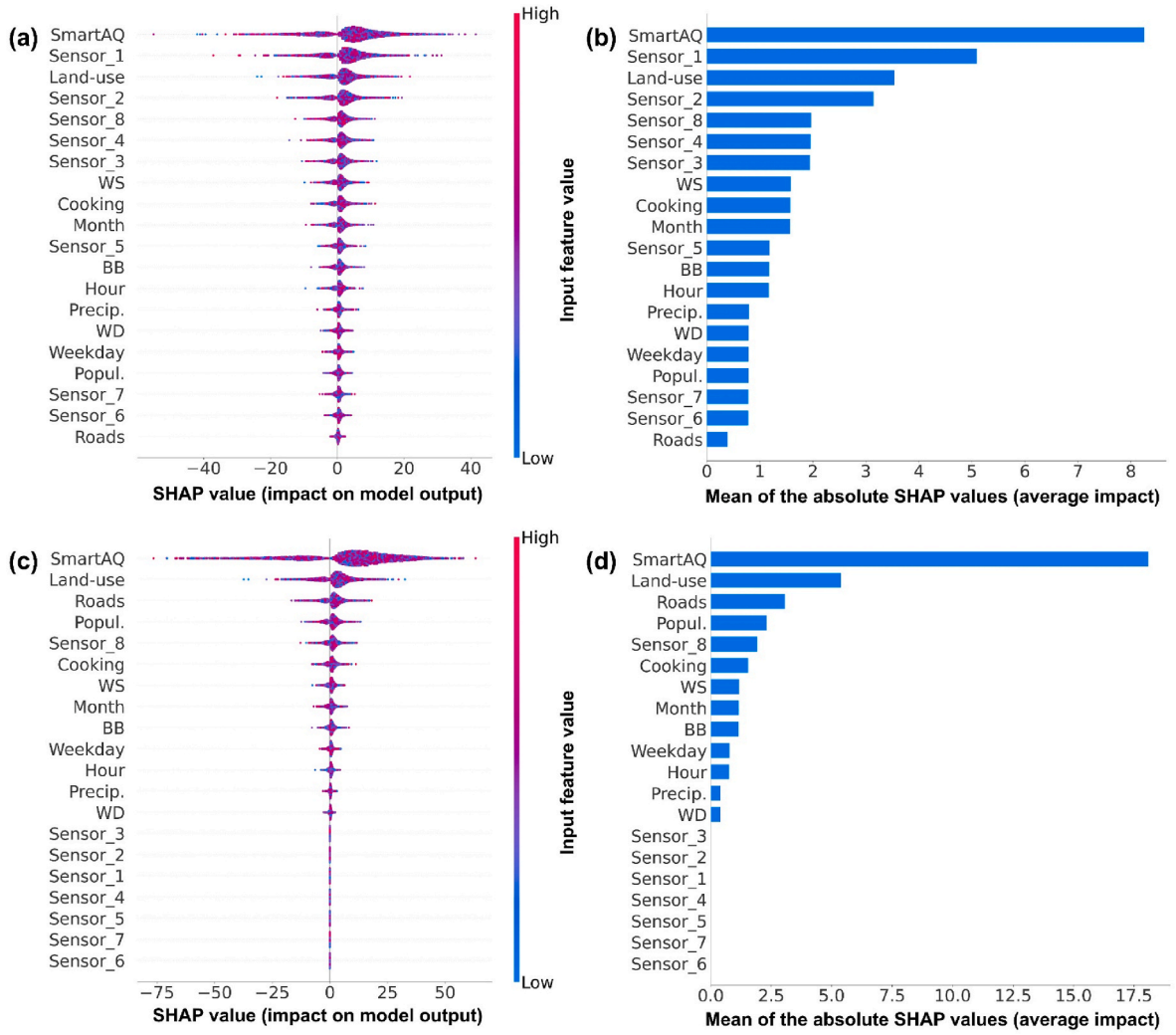


Fig. 11. Distribution of the SmartAQ+ SHAP values per input feature and input instance and mean of the absolute SHAP values for each input feature separating the grid cells into those with at least one sensor nearby (a,b) and remote ones (c,d) using data from 2023.

Funding

This research was funded by the Bodossaki Foundation and the Social and Cultural Affairs Welfare Foundation (KIKPE) (project PANSEN).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2025.121665.

Data availability

Data will be made available on request.

References

Allan, J.D., Williams, P.I., Morgan, W.T., Martin, C.L., Flynn, M.J., Lee, J., Nemitz, E., Phillips, G.J., Gallagher, M.W., Coe, H., 2010. Contributions from transport, solid fuel burning and cooking to primary organic aerosols in two UK cities. Atmos. Chem. Phys. 10, 647–668. https://doi.org/10.5194/acp-10-647-2010.

Arowosegbe, O.O., Röösli, M., Künzli, N., Saucy, A., Adebayo-Ojo, T.C., Schwartz, J., Kebalepile, M., Jeebhay, M.F., Dalvie, M.A., De Hoogh, K., 2022. Ensemble averaging using remote sensing data to model spatiotemporal PM_{10} concentrations in sparsely monitored South Africa. Environ. Pollut. 310, 119883. https://doi.org/10.1016/j.envpol.2022.119883.

Bai, Y., Li, Y., Wang, X., Xie, J., Li, C., 2016. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmos. Pollut. Res. 7, 557–566. https://doi.org/10.1016/j.apr.2016.01.004.

Bi, J., Knowland, K.E., Keller, C.A., Liu, Y., 2022. Combining machine learning and numerical simulation for high-resolution $PM_{2.5}$ concentration forecast. Environ. Sci. Technol. 56, 1544–1556. https://doi.org/10.1021/acs.est.1c05578.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

De Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N., Kloog, I., 2018. Modelling daily $PM_{2.5}$ concentrations at high spatio-temporal resolution across Switzerland. Environ. Pollut. 233, 1147–1154. https://doi.org/10.1016/j.envpol.2017.10.025.

De Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E.A., Strassmann, A., Puhan, M., Röösli, M., Stafoggia, M., Kloog, I., 2019. Predicting fine-scale daily NO_2 for 2005–2016 incorporating OMI satellite data across Switzerland. Environ. Sci. Technol. 53, 10279–10287. https://doi.org/10.1021/acs.est.9b03107.

Delle Monache, L., Alessandrini, S., Djalalova, I., Wilczak, J., Knievel, J.C., Kumar, R., 2020. Improving air quality predictions over the United States with an analog ensemble. Weather Forecast. 35, 2145–2162. https://doi.org/10.1175/WAF-D-19-0148.1.

Dinkelacker, B.T., Garcia Rivera, P., Marshall, J.D., Adams, P.J., Pandis, S.N., 2023. High-resolution downscaling of source resolved $PM_{2.5}$ predictions using machine

learning models. Atmos. Environ. 310, 119967. https://doi.org/10.1016/j.atmosenv.2023.119967.
European Parliament, 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).
Fang, L., Jin, J., Segers, A., Liao, H., Li, K., Xu, B., Han, W., Pang, M., Lin, H.X., 2023. A gridded air quality forecast through fusing site-available machine learning predictions from RFSML v1.0 and chemical transport model results from GEOS-Chem v13.1.0 using the ensemble Kalman filter. Geosci. Model Dev. (GMD) 16, 4867–4882. https://doi.org/10.5194/gmd-16-4867-2023.
Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P.I., Geron, C., 2006. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). Atmos. Chem. Phys. 6, 3181–3210. https://doi.org/10.5194/acp-6-3181-2006.
Guenther, A.B., Jiang, X., Heald, C.L., Sakulyanontvittaya, T., Duhl, T., Emmons, L.K., Wang, X., 2012. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. Geosci. Model Dev. (GMD) 5, 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012.
Hamill, T.M., Whitaker, J.S., 2006. Probabilistic quantitative precipitation forecasts based on Reforecast analogs: theory and application. Mon. Weather Rev. 134, 3209–3229. https://doi.org/10.1175/MWR3237.1.
Hua, W., Lou, S., Huang, X., Xue, L., Ding, K., Wang, Z., Ding, A., 2024. Diagnosing uncertainties in global biomass burning emission inventories and their impact on modeled air pollutants. Atmos. Chem. Phys. 24, 6787–6807. https://doi.org/10.5194/acp-24-6787-2024.
Kaltsonoudis, C., Florou, K., Kodros, J.K., Jorga, S.D., Vasilakopoulou, C.N., Baliaka, H. D., Matrali, A., Aktypis, A., Georgopoulou, M.P., Nenes, A., Pandis, S.N., 2025. Significant contributions of fresh and aged biomass burning organic aerosol from residential burning in a wintertime urban environment. Atmos. Environ. 343, 121018. https://doi.org/10.1016/j.atmosenv.2024.121018.
Kaltsonoudis, C., Kostenidou, E., Louvaris, E., Psichoudaki, M., Tsiligiannis, E., Florou, K., Liangou, A., Pandis, S.N., 2017. Characterization of fresh and aged organic aerosol emissions from meat charbroiling. Atmos. Chem. Phys. 17, 7143–7155. https://doi.org/10.5194/acp-17-7143-2017.
Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X., Sachdeva, S., 2019. Evaluation of different machine learning approaches to forecasting $PM_{2.5}$ Mass concentrations. Aerosol Air Qual. Res. 19, 1400–1410. https://doi.org/10.4209/aaqr.2018.12.0450.
Kaya, K., Gündüz Öğüdücü, Ş., 2020. Deep Flexible Sequential (DFS) model for air pollution forecasting. Sci. Rep. 10, 3346. https://doi.org/10.1038/s41598-020-60102-6.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30, 3146–3154.
Kerckhoffs, J., Hoek, G., Portengen, L., Brunekreef, B., Vermeulen, R.C.H., 2019. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. Environ. Sci. Technol. 53, 1413–1421. https://doi.org/10.1021/acs.est.8b06038.
Koo, Y.-S., Kwon, H.-Y., Bae, H., Yun, H.-Y., Choi, D.-R., Yu, S., Wang, K.-H., Koo, J.-S., Lee, Jae-Bum, Choi, M.-H., Lee, Jeong-Beom, 2023. A development of $PM_{2.5}$ forecasting System in South Korea using chemical transport modeling and machine learning. Asia-Pac. J. Atmospheric Sci. 59, 577–595. https://doi.org/10.1007/s13143-023-00314-8.
Kosmopoulos, G., Salamalikis, V., Pandis, S.N., Yannopoulos, P., Bloutsos, A.A., Kazantzidis, A., 2020. Low-cost sensors for measuring airborne particulate matter: field evaluation and calibration at a South-Eastern european site. Sci. Total Environ. 748, 141396. https://doi.org/10.1016/j.scitotenv.2020.141396.
Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., Denier Van Der Gon, H., 2022. CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. Earth Syst. Sci. Data 14, 491–515. https://doi.org/10.5194/essd-14-491-2022.
Kukkonen, J., Olsson, T., Schultz, D.M., Baklanov, A., Klein, T., Miranda, A.I., Monteiro, A., Hirtl, M., Tarvainen, V., Boy, M., Peuch, V.-H., Poupkou, A., Kioutsioukis, I., Finardi, S., Sofiev, M., Sokhi, R., Lehtinen, K.E.J., Karatzas, K., San José, R., Astitha, M., Kallos, G., Schaap, M., Reimer, E., Jakobs, H., Eben, K., 2012. A review of operational, regional-scale, chemical weather forecasting models in Europe. Atmos. Chem. Phys. 12, 1–87. https://doi.org/10.5194/acp-12-1-2012.
Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N., Baldé, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K.V., McTeer, M.A., Murray, C.J.L., Ndahimananjara, J.D., Perera, F., Potočnik, J., Preker, A.S., Ramesh, J., Rockström, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A., Stewart, R.B., Suk, W. A., Van Schayck, O.C.P., Yadama, G.N., Yumkella, K., Zhong, M., 2018. The Lancet Commission on pollution and health. Lancet 391, 462–512. https://doi.org/10.1016/S0140-6736(17)32345-0.
Lanz, V.A., Alfarra, M.R., Baltensperger, U., Buchmann, B., Hueglin, C., Prévôt, A.S.H., 2007. Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra. Atmos. Chem. Phys. 7, 1503–1522. https://doi.org/10.5194/acp-7-1503-2007.
Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. Atmos. Chem. Phys. 11, 7991–8002. https://doi.org/10.5194/acp-11-7991-2011.
Li, L., Wu, J., 2021. Spatiotemporal estimation of satellite-borne and ground-level NO_2 using full residual deep networks. Remote Sens. Environ. 254, 112257. https://doi.org/10.1016/j.rse.2020.112257.
Liu, B., Tan, X., Jin, Y., Yu, W., Li, C., 2021. Application of RR-XGBoost combined model in data calibration of micro air quality detector. Sci. Rep. 11, 15662. https://doi.org/10.1038/s41598-021-95027-1.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Lv, L., Wei, P., Li, J., Hu, J., 2021. Application of machine learning algorithms to improve numerical simulation prediction of $PM_{2.5}$ and chemical components. Atmos. Pollut. Res. 12, 101211. https://doi.org/10.1016/j.apr.2021.101211.
Ma, J., Yu, Z., Qu, Y., Xu, J., Cao, Y., 2020. Application of the XGBoost machine learning method in $PM_{2.5}$ prediction: a case Study of Shanghai. Aerosol Air Qual. Res. 20, 128–138. https://doi.org/10.4209/aaqr.2019.08.0408.
Malings, C., Knowland, K.E., Pavlovic, N., Coughlin, J.G., King, D., Keller, C., Cohn, S., Martin, R.V., 2024. Air quality estimation and forecasting via data fusion with uncertainty quantification: theoretical framework and preliminary results. J. Geophys. Res. Mach. Learn. Comput. 1, e2024JH000183. https://doi.org/10.1029/2024JH000183.
Monahan, E.C., Spiel, D.E., Davidson, K.L., 1986. A model of marine aerosol generation via whitecaps and wave disruption. In: Monahan, Edward C., Niocaill, G.M. (Eds.), Oceanic Whitecaps, Oceanographic Sciences Library. Springer, Netherlands, Dordrecht, pp. 167–174. https://doi.org/10.1007/978-94-009-4668-2_16.
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.
O'Dowd, C.D., Langmann, B., Varghese, S., Scannell, C., Ceburnis, D., Facchini, M.C., 2008. A combined organic-inorganic sea-spray source function. Geophys. Res. Lett. 35, 2007GL030331. https://doi.org/10.1029/2007GL030331.
Pappa, A., Kioutsioukis, I., 2021. Forecasting particulate pollution in an urban area: from copernicus to Sub-km Scale. Atmosphere 12, 881. https://doi.org/10.3390/atmos12070881.
Pappa, A., Siouti, E., Pandis, S.N., Kioutsioukis, I., 2023. High-resolution WRF forecasts in the SmartAQ system: evaluation of the meteorological forcing used for PMCAMx predictions in an urban area. Atmos. Res. 296, 107041. https://doi.org/10.1016/j.atmosres.2023.107041.
Prasad, K., Gorai, A.K., Goyal, P., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmos. Environ. 128, 246–262. https://doi.org/10.1016/j.atmosenv.2016.01.007.
Requia, W.J., Di, Q., Silvern, R., Kelly, J.T., Koutrakis, P., Mickley, L.J., Sulprizio, M.P., Amini, H., Shi, L., Schwartz, J., 2020. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. Environ. Sci. Technol. 54, 11037–11047. https://doi.org/10.1021/acs.est.0c01791.
Siouti, E., Kilafis, K., Kioutsioukis, I., Pandis, S.N., 2023a. Simulation of the influence of residential biomass burning on air quality in an urban area. Atmos. Environ. 309, 119897. https://doi.org/10.1016/j.atmosenv.2023.119897.
Siouti, E., Skyllakou, K., Kioutsioukis, I., Ciarelli, G., Pandis, S.N., 2021. Simulation of the cooking organic aerosol concentration variability in an urban area. Atmos. Environ. 265, 118710. https://doi.org/10.1016/j.atmosenv.2021.118710.
Siouti, E., Skyllakou, K., Kioutsioukis, I., Patoulias, D., Apostolopoulos, I.D., Fouskas, G., Pandis, S.N., 2023b. Prediction of the concentration and source contributions of $PM_{2.5}$ and gas-phase pollutants in an urban area with the SmartAQ forecasting System. Atmosphere 15, 8. https://doi.org/10.3390/atmos15010008.
Siouti, E., Skyllakou, K., Kioutsioukis, I., Patoulias, D., Fouskas, G., Pandis, S.N., 2022. Development and application of the SmartAQ high-resolution air quality and source apportionment forecasting System for European urban areas. Atmosphere 13, 1693. https://doi.org/10.3390/atmos13101693.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Liu, Z., Berner, J., Wang, W., Powers, J.G., Duda, M.G., Barker, D.M., Huang, X.-Y., 2019. A Description of the Advanced Research WRF Model Version 4. UCAR/NCAR. https://doi.org/10.5065/1DFH-6P97.
Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J., 2019. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environ. Int. 124, 170–179. https://doi.org/10.1016/j.envint.2019.01.016.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958.
Tang, D., Zhan, Y., Yang, F., 2024. A review of machine learning for modeling air quality: overlooked but important issues. Atmos. Res. 300, 107261. https://doi.org/10.1016/j.atmosres.2024.107261.
Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ. 409, 1266–1276. https://doi.org/10.1016/j.scitotenv.2010.12.039.
Wagstrom, K.M., Pandis, S.N., Yarwood, G., Wilson, G.M., Morris, R.E., 2008. Development and application of a computationally efficient particulate matter apportionment algorithm in a three-dimensional chemical transport model. Atmos. Environ. 42, 5650–5659. https://doi.org/10.1016/j.atmosenv.2008.03.012.
Wolf, T., Pettersson, L.H., Esau, I., 2020. A very high-resolution assessment and modelling of urban air quality. Atmos. Chem. Phys. 20, 625–647. https://doi.org/10.5194/acp-20-625-2020.

World Health Organization, 2018. How air pollution is destroying our health [WWW Document]. World Health Organ. URL. https://www.who.int/news-room/spotlight/how-air-pollution-is-destroying-our-health, 9.9.24.

Xu, M., Jin, J., Wang, G., Segers, A., Deng, T., Lin, H.X., 2021. Machine learning based bias correction for numerical chemical transport models. Atmos. Environ. 248, 118022. https://doi.org/10.1016/j.atmosenv.2020.118022.

Yu, W., Li, S., Ye, T., Xu, R., Song, J., Guo, Y., 2022. Deep ensemble machine learning framework for the estimation of $PM_{2.5}$ concentrations. Environ. Health Perspect. 130, 037004. https://doi.org/10.1289/EHP9752.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012. Real-time air quality forecasting, part I: history, techniques, and current status. Atmos. Environ. 60, 632–655. https://doi.org/10.1016/j.atmosenv.2012.06.031.

Zhou, Y., Chang, F.-J., Chang, L.-C., Kao, I.-F., Wang, Y.-S., Kang, C.-C., 2019. Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting. Sci. Total Environ. 651, 230–240. https://doi.org/10.1016/j.scitotenv.2018.09.111.