ACS ES&T Air

Open Access

pubs.acs.org/estair

Article

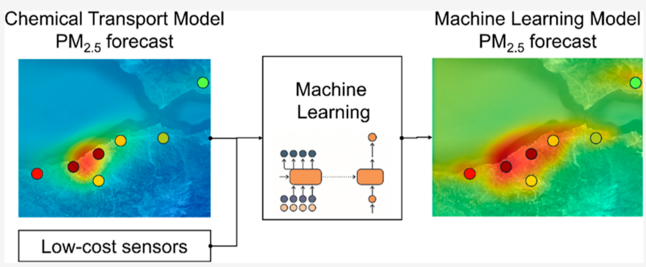
A Machine Learning High-Resolution $PM_{2.5}$ Forecast Model (SmartAQ^{2+}) Fusing Chemical Transport Model Predictions and Low-Cost Sensor Measurements

Ioannis D. Apostolopoulos, Evangelia Siouti, George Fouskas, and Spyros N. Pandis*

Cite This: https://doi.org/10.1021/acsestair.6c00113

Read Online

ABSTRACT: We introduce SmartAQ^{2+}, a data fusion model combining a chemical transport model-based system (SmartAQ), low-cost $PM_{2.5}$ sensors, and machine learning (ML) to improve 3-day high-resolution $PM_{2.5}$ forecasts in urban areas. The ML model was trained using 2021−2023 data from Patras, Greece and evaluated in the same city with data from 2024. SmartAQ^{2+} had better seasonal, monthly, and hourly performance than SmartAQ, reducing its error by a factor of 2. Across the 31 Patras monitoring sites, SmartAQ^{2+} had good-to-excellent (fractional error ≤50%) monthly average performance in 21 sites in winter and spring, 17 in summer, and 22 in autumn. For exceedances of the EU daily mean threshold of 25 $\mu g\ m^{-3}$, SmartAQ^{2+} correctly predicted 23 of 54 exceedance days (43%). Its performance was similar for forecasts issued one and 2 days in advance and deteriorated by 30% after 48 h. Including cooking emissions in SmartAQ and as a separate input feature in SmartAQ^{2+} reduced daily mean error by 30% at areas with high restaurant density. Sensitivity analysis showed highest reliance on sensors in the urban core and increased reliance on SmartAQ and static spatial features toward the suburbs. Without retraining, the model showed preliminary evidence of spatial transferability to Athens.



KEYWORDS: *machine learning, data fusion, chemical transport model, fine particle matter, air quality forecasting*

1. INTRODUCTION

Chemical transport models (CTMs) have been essential tools for forecasting air pollution, source apportionment, exposure assessment, and evaluation of emission control strategies effectiveness.[1−3] They simulate air quality by accounting explicitly for the chemical and physical processes in the atmosphere and can quantify the contribution of pollution sources, whether local or regional, as well as simulate the formation of secondary pollutants,[4] without relying on past data. CTMs require comprehensive information on emissions from sources such as industry, transportation, and biomass burning. However, emission inventories are often incomplete, outdated, or uncertain, which can lead to inaccuracies and biases.[5] Additionally, CTMs inherit the uncertainties of meteorological forecasts. Air quality in a specific area depends, among others, on weather conditions such as temperature, wind, vertical mixing, and precipitation. Therefore, inaccuracies in weather predictions affect the reliability of the corresponding air quality predictions.

On the other hand, statistical models use historical air quality records and meteorological conditions and are computationally efficient, because they do not explicitly represent pollutant emissions or atmospheric chemistry.[6,7] However, these models assume that the future will behave like the past and have weaknesses when conditions fall outside the historical range. They are also constrained by data quality and lack of physical understanding.

Machine learning (ML) models have gained attention in air pollution forecasting due to their ability to represent complex relationships between air pollutant concentrations and diverse predictors. These ML-based predictors use air quality measurements, meteorological predictions, land-use and traffic-related information, and other environmental information as inputs.[8−11] In contrast to CTMs, ML methods are data-driven and can adapt to empirical patterns without explicitly simulating physical or chemical processes. ML models developed in the past include Support Vector Machines, Artificial Neural Networks,[12−16] Random Forests,[17,18] Ensemble methods,[19,20] and Deep Neural Networks.[21]

ML-based models have several limitations. They are constrained by their reliance on historical data, with performance largely determined by data quality, quantity, and representative-

Received: March 18, 2026
Revised: June 6, 2026
Accepted: June 8, 2026

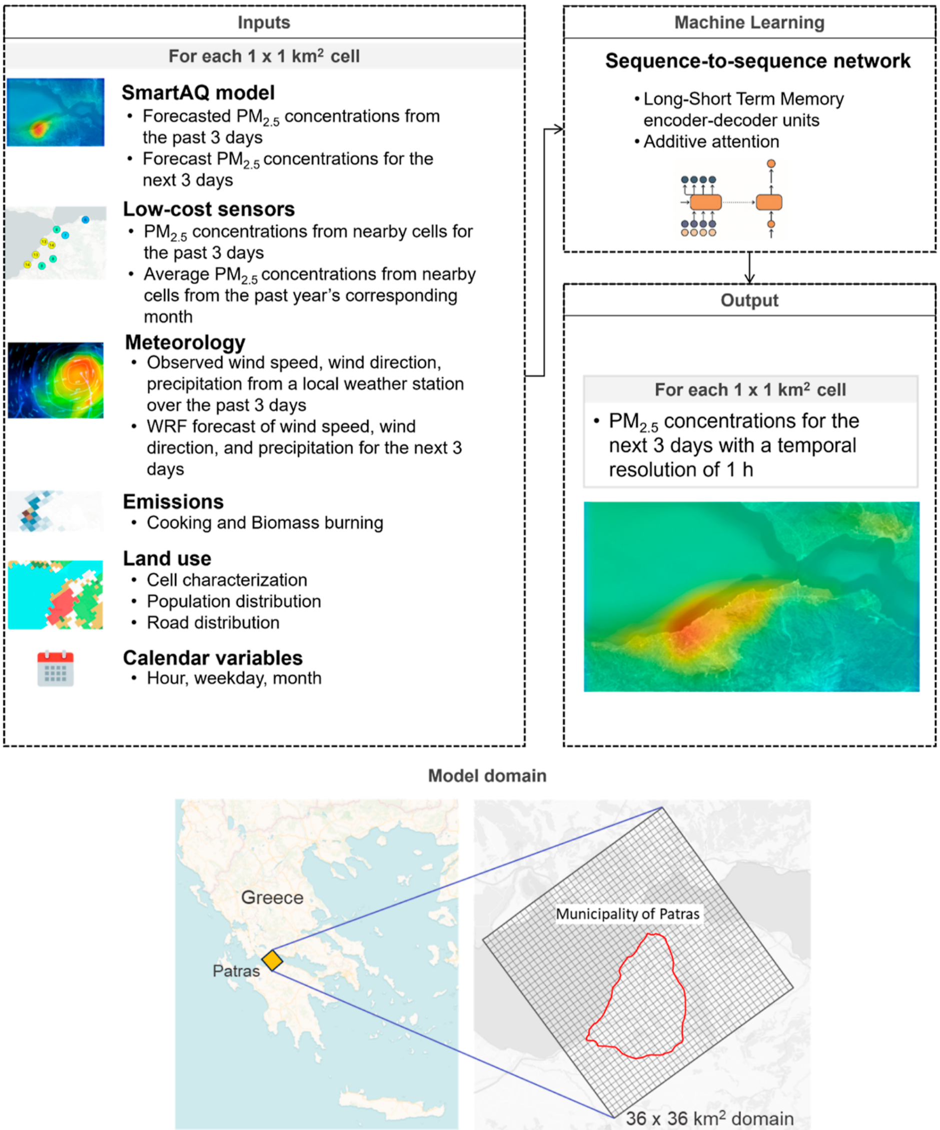


Figure 1. SmartAQ^{2+} pipeline.

ness. Their accuracy often declines under novel conditions such as abrupt meteorological changes, new emission sources, or unprecedendent events. ML models treat inputs abstractly, overlooking the physical and chemical processes that drive atmospheric dynamics, and as a result, they are prone to overfitting and may exhibit limited generalizability. Moreover, increasing ML model complexity can reduce interpretability and raise computational costs without commensurate improvements in predictive accuracy.[22]

Researchers are increasingly exploring hybrid frameworks that combine the strengths of CTMs and ML through data fusion techniques. Recent research involves the fusion of CTMs' output with satellite and ground observations[23] with the aid of ML. One use of such models is the reanalysis of past CTM simulations. For example, Tang et al.[24] used a high-resolution random forest approach to estimate the past daily ground-level concentrations of $PM_{2.5}$, O_3, and NO_2 at 1 km resolution over South Korea. Inputs to the ML model included meteorological reanalysis, satellite aerosol optical depth (AOD), and CTM outputs. Under a 10-fold cross-validation procedure, the achieved root mean squared errors were 5.5 $\mu g\ m^{-3}$ ($PM_{2.5}$), 5.5 ppb (O_3), and 4.7 ppb (NO_2). Hu et al.[25] developed a historically reconstructed global data set of daily $PM_{2.5}$ concentrations resulting from wildfires (2000−2023) using the GEOS-Chem CTM simulations and corrected the observed biases using the XGBoost algorithm. They trained annual XGBoost models on around 9500 ground monitoring sites worldwide, using simulated $PM_{2.5}$, meteorological reanalysis, and land cover as predictors, to correct systematic errors in CTM outputs. The bias-corrected total $PM_{2.5}$ was then combined with fire-to-total ratios resulting from the GEOS-Chem model to estimate the past fire contributions to the total $PM_{2.5}$ concentrations.

Another use of such models is to produce short-term air quality forecasts (e.g., 24 h) in real-time conditions. Xing et al.[26] developed a physically constrained deep-learning fusion framework (DeepMMF) that integrates CTM simulations, satellite retrievals, and ground measurements to estimate present-time surface NO_2 concentrations over the continental U.S. The approach pretrains deep learning surrogate models (DeepCTM) on CTM outputs to capture physically consistent relationships between emissions, meteorology, and concen-

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX

trations, then fine-tunes the model using satellite and in situ observations. A variational autoencoder (VAE) deep learning algorithm with emission constraints is employed to simultaneously adjust concentrations and infer emissions, ensuring physical consistency and mitigating sample-imbalance issues that bias conventional ML methods. The framework achieved an average RMSE for NO_2 of 1.45 ppb, when compared to the Air Quality System surface monitors and demonstrated how embedding physical laws into ML can improve both prediction and interpretability. In other studies,[27,28] ML has been used to improve short-term (6−48 h) $PM_{2.5}$ forecasts by correcting CTM biases using past observed data from regulatory measurements. Raffuse et al.[29] introduced Rapidfire (relatively accurate particulate information derived from inputs retrieved easily), an R package designed to rapidly generate spatially resolved estimates of wildfire smoke exposure for the next 24 h in near real-time. The model integrates regulatory monitors, satellite observations, meteorological models, smoke prediction models, and low-cost sensor measurements using a random forest regression framework. The package was applied to major California wildfire events from 2017−2021, having an average RMSE of 13.5 μg m^{-3} between predicted $PM_{2.5}$ and observed with regulatory instruments.

ML-CTM approaches have also been proposed to improve air quality simulations, either by transforming CTM outputs into fine-resolution pollution predictions for the past[30] and the future,[31,32] or by correcting CTM biases. For example, Lv et al.,[33] applied support vector regression and random forest to reduce bias in $PM_{2.5}$ component forecasts, with high accuracy but limited urban-scale spatial testing. Malings et al.[34] introduced a modular, uncertainty-aware NO_2 prediction framework using satellite data and CTM outputs (GEOS-CF) that achieved 5 km scale forecasts. Fang et al.[23] 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 a high computational cost.

Most CTM bias-correction studies rely primarily on sparse regulatory monitors. Air quality models operating in coarse resolution do not necessarily need dense ground observations for bias correction. In high spatial resolution (e.g., 1 × 1 km^2), however, correcting CTM biases and errors with limited ground stations providing reference concentrations is challenging. The use of calibrated low-cost sensor measurements has not been adequately explored in recent literature. Low-cost sensor measurements can be used both for ML training and as supporting input.

The present study introduces a hybrid CTM-ML approach that fuses multiple data sources, including CTM outputs, measurements of $PM_{2.5}$ from low-cost sensors, meteorology, and emission data to provide a high-resolution (1 × 1 km^2) 3-day $PM_{2.5}$ concentration forecast at an urban area and its surroundings in Western Greece. SmartAQ^{2+} integrates dense low-cost $PM_{2.5}$ sensor observations directly into the forecasting architecture, enabling CTM bias correction in areas with $PM_{2.5}$ observations while relying on CTM-driven predictions in sensor-sparse areas of the simulation domain.

In a recent study,[35] we proposed an XGBoost model that improves the $PM_{2.5}$ concentration fields produced by a CTM for the present time using the same type of data inputs. Here, we extend our methodology to provide a 3-day $PM_{2.5}$ concentration forecast. Because the goal of the proposed model is not to provide a single $PM_{2.5}$ value for each simulation cell but a 72 h forecast, a sequence-to-sequence architecture was selected. The proposed ML model requires past and future CTM and meteorological drivers along with $PM_{2.5}$ measurements from low-cost sensors. We also investigate the feasibility of implementing the trained ML algorithm in another urban area, the capital city of Greece, Athens, without performing model retraining with historical data from Athens.

2. METHODS

2.1. Study Areas

The SmartAQ system, which was based on the PMCAMx CTM, was implemented in Europe at a horizontal resolution of 36 × 36 km^2, while the urban domains had a resolution of 1 × 1 km^2. The outer European grid covered an area of 5400 × 5832 km^2, whereas the urban domains covered a 36 × 36 km^2 area for Patras and a 72 × 72 km^2 area for Athens. Two intermediate grids were also applied, with resolutions of 12 × 12 km^2 and 3 × 3 km^2, both centered on each city of interest. All domains included 14 vertical layers.

2.2. Machine Learning Model Inputs

2.2.1. SmartAQ System. Figure 1 presents an overview of the ML model's inputs and outputs. The SmartAQ forecasting system provides three-day hourly 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. Also, it estimates the sources of all pollutants. Six models predicting anthropogenic, biogenic, and marine emissions, pollutant concentrations, major pollutant sources (biomass burning, long-range transport, transportation, ships, etc.), and weather conditions are utilized.[36,37] The WRF model[38] is used for the computation of meteorological fields (e.g., cloud cover, precipitation, temperature, and humidity) required for air quality predictions. The MEGAN (Model of Emissions of Gases and Aerosols from Nature) model is used to estimate gas and aerosol emissions for natural and terrestrial ecosystems, integrating WRF meteorological data with land use information.[39,40] Marine emissions, including sea salt and organics, are calculated using the O'Dowd and Monahan algorithms,[41,42] which rely on WRF-predicted wind speeds over the sea surface. The TNO emission inventory[43] provides the anthropogenic emissions. Modeling of the air pollution within the target domain is performed by the PMCAMx chemical transport model, while the particulate source apportionment technology (PSAT) algorithm[44] determines the contributions of individual sources to pollutant concentrations. Additional details are available in Siouti et al.[45]

2.2.2. Low-Cost Sensors. A low-cost sensor network consisting of 31 Purple Air devices (PurpleAir PA-II) provided reference $PM_{2.5}$ concentration at fixed locations in the study area (Figure S1). Fifteen devices were in urban computational cells, 13 in suburban cells, and three in rangeland cells. We used measurements from January 1, 2021, to December 31, 2023, for model training and from January 1 to December 31, 2024, for evaluation. All measurements were averaged over 1 h intervals. We used the average $PM_{2.5}$ concentration of the two identical Plantower PMS5003 sensors of each device. In cases where one of the PMS5003 sensors in the same device was reporting concentrations above 700 μg m^{-3} or below 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 calibrated as described in the study by Kosmopoulos et al.[46] for the same area.

2.2.3. Land-Use and Emissions. The ML model's input includes information for each grid cell regarding its land use, population, and major emission sources (Figure 1). The United States Geological Survey (USGS) geographical data sets for topography and land use were utilized to classify the land use within each modeling domain. Land is categorized into sea, urban, agricultural, rangeland, forest, and uncategorized areas based. The population distribution is calculated based on data provided by Eurostat.[47] The data pertain to 2021 and are available in high-resolution.

Residential biomass burning emissions at high-spatial resolution of 1 × 1 km^2 for the Patras area have been estimated by Siouti et al.,[36] based on the density of houses in the area. Cooking emissions are spatially distributed based on the density of the restaurants in each 1 × 1 km^2 grid cell of the modeling domain.[48] For Athens, no cooking emissions were considered, while biomass burning emissions were estimated by Siouti et al.,[49] similar to Patras.

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

2.2.4. Meteorology. In addition to the WRF forecasts used by the SmartAQ system, SmartAQ^{2+} uses hourly observations of wind speed, direction, and rainfall, from the University of Patras meteorological station. Measurements for the previous 3 days before the day that the forecast was issued are used as inputs for the ML model.

2.3. Machine Learning Algorithm

The ML algorithm produces a 72 h forecast of hourly $PM_{2.5}$ for each computational cell of the domain. It receives four sets of input (Table S1). The first set of inputs includes $PM_{2.5}$ simulation results from the SmartAQ model, low-cost sensor $PM_{2.5}$ measurements, meteorology (wind speed, wind direction, precipitation), and calendar variables (month, weekday, hour) for the past 3 days starting from and ending at midnight. The second set of inputs includes a 72 h forecast of $PM_{2.5}$ concentrations provided by the SmartAQ model, meteorological variables provided by the WRF (wind speed, wind direction, precipitation), and calendar variables. The third set of inputs includes land use, cooking and biomass burning emissions, population distribution, and percentage of total road surface. The fourth input set includes the monthly averaged observed and SmartAQ-predicted $PM_{2.5}$ of the past year's same month at the same locations where $PM_{2.5}$ measurements from the low-cost sensor network were available. These variables do not provide information about specific days or events within the forecasted period; they only supply a long-term average from one year earlier.

The ML model is an encoder-decoder with Bahdanau attention layers.[50] The encoder is the part of the network that reads the past and builds an internal summary of recent behavior. The decoder is the part that uses this summary, together with information about the future, to generate the forecast hour by hour. Both parts are implemented with long short-term memory (LSTM) layers. An LSTM is a type of neural network designed for time series, which processes the data in time order and keeps an internal memory of what it has seen.

The training objective is to minimize the mean absolute error (MAE) between the ML-predicted $PM_{2.5}$ concentrations and measurements from the $PM_{2.5}$ low-cost sensors. Detailed information regarding the model's architecture, parameters and hyper-parameters is provided in the supplement. The LSTM model is suitable for air-quality forecasting, because it can capture multiscale temporal dependencies such as diurnal cycles and weekday-weekend shifts exhibited by $PM_{2.5}$.

The particular ML algorithm was selected because the task requires the translation of a recent multivariate history into a multistep future $PM_{2.5}$ trajectory. The 72 h input window contains short-term persistence, diurnal variability, weekly patterns, meteorological effects, and CTM-driven forecast signals, all of which are temporally ordered and interdependent. The encoder-decoder structure is well suited to this setting, as it first summarizes the recent evolution of the system and then generates the 72 h forecast horizon step by step. The attention mechanism further improves this formulation by allowing the decoder to selectively use different parts of the input window at each forecast step, rather than relying only on the final encoder state.

2.4. Sensor Selection

For each computational cell, the ML model uses readings from the eight closest sensors plus a complementary background sensor (located in Platani). Any sensor farther than 4 km is left out, ensuring the inputs are geographically relevant. If fewer than 8 sensors are within 4 km, a smaller number of sensors is used. Focusing on nearby sensors helps the model capture real-time conditions in the target cell and account for local environmental or operational effects, while reducing the chance that distant data will bias predictions. In practice, this lets SmartAQ^{2+} generate estimates in remote areas with little or no sensor coverage by relying more on the SmartAQ predictions and the background site.

2.5. Model Operation and Outputs

The ML model runs for each computational cell. Its output is a 3-day $PM_{2.5}$ concentration time series at an hourly resolution. Application of the ML model for all grid cells results in the construction of the 2D map for each hour of the forecast. The ML model is designed to operate once a day at a predefined time (16.00), which is right after the SmartAQ model has completed its daily run. In addition, the ML model provides the relative importance of each given input to the specific forecast, using the SHapley Additive exPlanations (SHAP) analysis,[51] which ranks the ML input features by how much they move a prediction away from a baseline.

2.6. Performance Assessment

We trained SmartAQ^{2+} on Patras data for 3 years, from January 2021 to December 2023, and evaluated it on a held-out test year (January to December 2024). This baseline configuration of SmartAQ^{2+} included all available input features, including grid-cell cooking emissions. For the 2024 Patras test, we evaluated the model's three-day forecasting capability, quantifying performance separately for 1, 2, and 3-day lead times to capture how performance degrades with the prediction horizon.

To assess robustness when cooking activity is unknown, we repeated the Patras 2024 evaluation under a "no-cooking" scenario. Specifically, cooking-emission inputs were withheld at prediction time, and the SmartAQ predictions passed into SmartAQ^{2+} were likewise generated without access to cooking emissions. This test evaluates the effect of a missing source in the inventory used.

Finally, to assess cross-city transferability, we applied the Patras-trained SmartAQ^{2+} in Athens for a short pilot for one month (September 2025), without additional retraining and without cooking emission inputs (these features were not available).

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

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N} |P_i - O_i| \quad (1)$$

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

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

where, N is the total number of measurements, P_i is the predicted concentration and O_i is the corresponding reference concentration. Based on Morris et al.,[52] $PM_{2.5}$ model performance for daily average values is considered excellent for FBIAS ≤ ±15% and FERROR ≤ ±35%, good for FBIAS ≤ ±30% and FERROR ≤ ±50%, average for FBIAS≤ ±60% and FERROR ≤± 75%, while there are fundamental problems in the modeling system for higher FBIAS and FERROR.

EU has recently defined a daily mean $PM_{2.5}$ limit of 25 $\mu g\ m^{-3}$, allowing at most 18 exceedance days per calendar year,[53] effective after 2030. We evaluated how well SmartAQ and SmartAQ^{2+} identified days when $PM_{2.5}$ surpassed this threshold anywhere in the urban modeling domain. For each site with a reference sensor, we compared how many times the models' daily averages and the sensors' daily averages exceeded 25 $\mu g\ m^{-3}$. This assesses each system's accuracy and reliability in detecting exceedances. We defined a True Positive (TP) where both observed and predicted daily means were >25 $\mu g\ m^{-3}$, a False Positive (FP) when the model exceeded the threshold but observations did not, a True Negative (TN) when both were ≤25 $\mu g\ m^{-3}$, and False Negative (FN) when observations exceeded the threshold, but the model did not.

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX

3. RESULTS

3.1. Model Performance in Patras

3.1.1. Performance in Predicting Average Seasonal and Monthly Patterns. We first used the monthly averaged $PM_{2.5}$ concentrations to evaluate the performance of the SmartAQ^{2+} during each season of 2024 and based on forecasts issued 1 day in advance. Figure 2 illustrates the FERROR and FBIAS of SmartAQ^{2+} per season at each measurement site.

Seasonal analysis revealed distinct spatial patterns in SmartAQ^{2+} performance. In winter, the model achieved "Excellent" or "Good" performance at 21/31 sites with an average monthly FERROR of 35%, including all sites near the center of the city. The model overestimated or underestimated $PM_{2.5}$ at some suburb sites (Koukouli, Sinora, Paralia) by 50% with an average FERROR of 55%. For spring and summer, the model achieved 'Excellent' ratings at more than half of the evaluation sites with an average FERROR of 20%. For several suburban sites the model overestimated $PM_{2.5}$ concentrations by 45% with an average FERROR of 45%. SmartAQ^{2+} performance during autumn was characterized by a systematic shift toward positive bias (+20%) for some suburb sites, while for the majority of sites near the city center the model's performance was excellent (FERROR = 35%, FBIAS = ±10%).

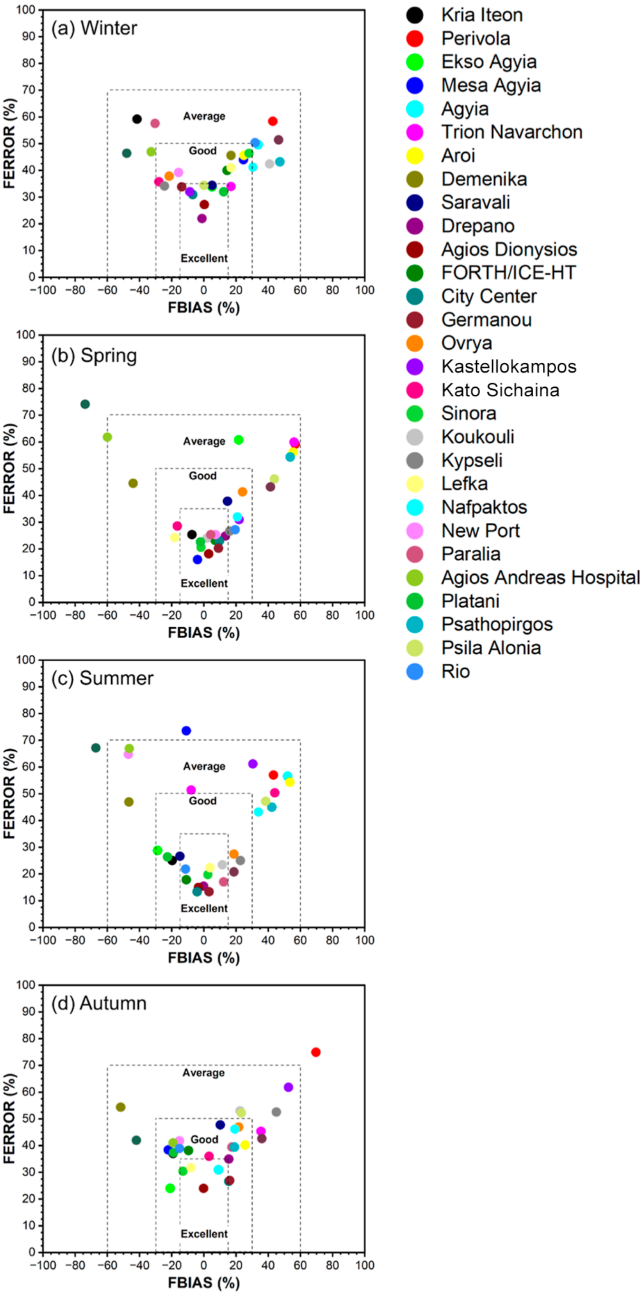


Figure 2. SmartAQ^{2+} evaluation using the average FERROR (%) versus FBIAS (%) during (a) winter, (b) spring, (c) summer, and (d) autumn for Patras.

Figure 3 shows average $PM_{2.5}$ predictions of the SmartAQ^{2+} model for all months of 2024, based on forecasts issued 1 day in advance. In winter months, $PM_{2.5}$ concentrations at areas within the city center and its suburbs were higher, due to intense biomass burning during the afternoon and night.[54] SmartAQ^{2+} captured this pollution pattern with an average FERROR of 30%. From May to October, the average monthly $PM_{2.5}$ concentration within the city ranged from 6 to 10 $\mu g\ m^{-3}$, based on the low-cost sensor measurements. SmartAQ^{2+} had an average FERROR of 25% and its FBIAS was ±17%.

We quantified the model's behavior for each simulation cell in relation to the cell's land use over the year using daily average $PM_{2.5}$ concentrations. Complete performance metrics are provided in Tables S2−S4.

In winter, average measured concentrations were about 18 $\mu g\ m^{-3}$ at urban sites and 4 $\mu g\ m^{-3}$ at nonurban sites, while SmartAQ^{2+} predicted roughly 17 and 5 $\mu g\ m^{-3}$ (FERROR of 5% and 22%, respectively). In spring, measured concentrations were approximately 8 $\mu g\ m^{-3}$ in the urban core and 7 $\mu g\ m^{-3}$ at the remaining sites, and SmartAQ^{2+} predicted 9 and 8 $\mu g\ m^{-3}$, with FERROR of 12% for both site types. In autumn, urban concentrations were around 9 $\mu g\ m^{-3}$ and were reproduced very well by SmartAQ^{2+} (5% FERROR), whereas nonurban concentrations were near 5 $\mu g\ m^{-3}$ and were overestimated at about 7 $\mu g\ m^{-3}$ (FERROR of 33%).

During summer, when measured $PM_{2.5}$ levels were lower overall (about 7 $\mu g\ m^{-3}$ at urban sites and 5−6 $\mu g\ m^{-3}$ at nonurban sites), the average of SmartAQ^{2+} daily FERRORs increases to about 42% at urban locations and 25% at nonurban locations, mainly because of underestimation errors at neighboring stations such as Ekso Agyia, Kato Sichaina, and Mesa Agyia. For example, the average ME in Kato Sichaina during the summer was 4.5 $\mu g\ m^{-3}$ (45% FERROR) and the MB was −2 $\mu g\ m^{-3}$ (−20% FBIAS).

3.1.2. Performance in Predicting Daily $PM_{2.5}$ Limit Exceedance. We evaluated the SmartAQ^{2+} model in detecting exceedance events by comparing its 1 day-in-advance predictions with sensor measurements at all sites during 2024. We have an exceedance event if any of the urban and suburban sensors' daily average $PM_{2.5}$ exceeded the limit. The total number of days with at least one sensor exceeding the average daily $PM_{2.5}$ limit was 54 (Table 1). SmartAQ^{2+} identified correctly 23 (43%) and missed 31 (57%) $PM_{2.5}$ exceedances, while having 12 false positives and 299 true negatives. From the 31 missed days, 15 were during summer, 6 during autumn, and 10 during winter. During summer, the 15 missed exceedances were due to local $PM_{2.5}$ spikes in the port of Patras (New Port), affecting one sensor. They are likely due to sources like maritime and heavy-duty vehicle emissions. The average daily $PM_{2.5}$ concentrations based on sensor measurements during the exceedance days were between 27 and 42 $\mu g\ m^{-3}$, whereas the predicted $PM_{2.5}$ concentrations were between 19 and 24 $\mu g\ m^{-3}$. For 13 of the 31 missed days, the predicted $PM_{2.5}$ concentration

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX

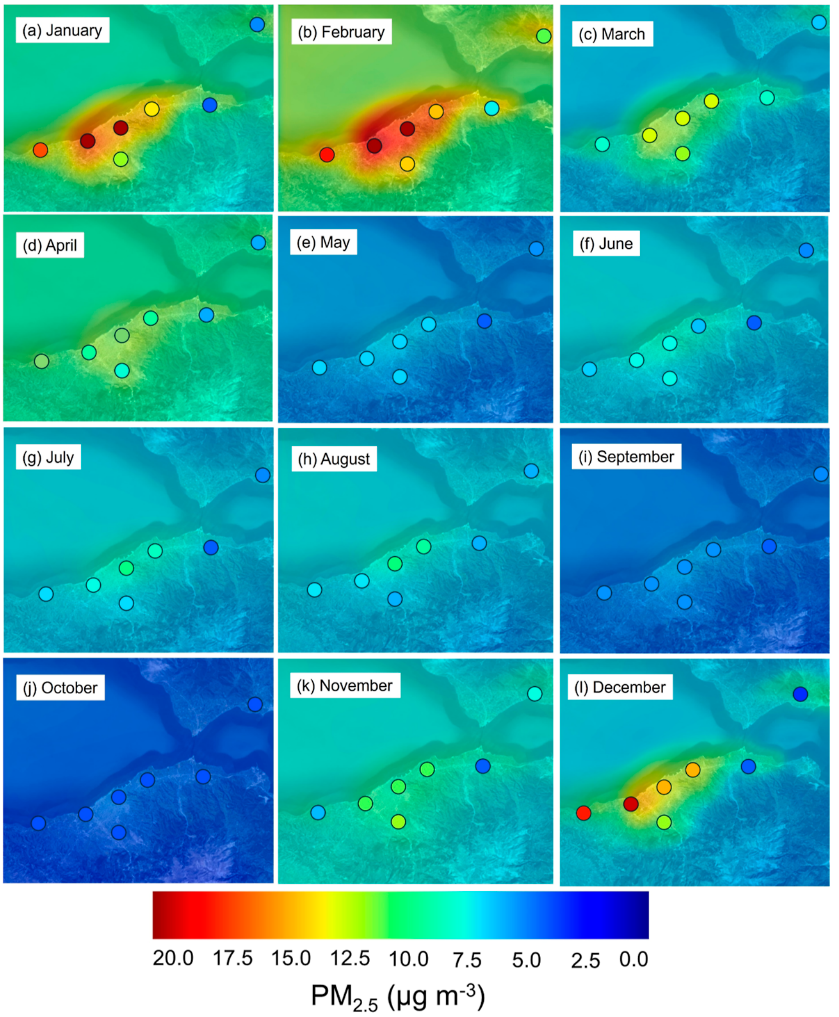


Figure 3. Predicted monthly average $PM_{2.5}$ concentrations by SmartAQ^{2+} for Patras domain. Colored circles denote monthly average $PM_{2.5}$ sensor measurements at representative locations.

Table 1. $PM_{2.5}$ Limit Exceedance (>25 μg m^{-3}) Occurrences During 2024 and Performance Metrics of the SmartAQ and SmartAQ^{2+} Models

	smartAQ^{2+}			smartAQ
	1 day ahead	2 days ahead	3 days ahead	1 day ahead
true positives	23	23	14	18
true negatives	299	300	298	280
false positives	12	11	13	31
false negatives	31	31	40	36

was close to the threshold (>22 μg m^{-3}). For days considered false positives, the average daily $PM_{2.5}$ concentrations based on sensor measurements were between 9 and 24 μg m^{-3}, whereas the model predicted 25–30 μg m^{-3}.

We selected the nine most polluted locations (New Port, Kria Iteon, Lefka, Trion Navarchon, Sinora, Paralia, Demenika, Koukouli, and Kypseli), which exceeded the $PM_{2.5}$ limit for 5 days or more (Table S5). The total number of exceedance events at these nine sites together was 144. SmartAQ^{2+} identified correctly 64 (44%) and had 80 false negatives. On the other hand, it had 56 false positive cases (47% of the predicted as positive). For example, the most polluted site was at the new port of Patras (32 days of limit exceedance). SmartAQ^{2+} identified correctly 9 days, missed 23 by an average of 15 μg m^{-3} (60%), and identified incorrectly 15 days. For these sites, we also examined the daily averaged $PM_{2.5}$ measurements and model predictions during the year (Figure 4). Several false negatives (missed exceedances) were due to errors of a few μg m^{-3} with the corresponding SmartAQ^{2+} predictions in the 18–24 μg m^{-3} range. False positives typically occurred when the observed $PM_{2.5}$ concentrations were already near the threshold (18–24 μg m^{-3}) and the SmartAQ^{2+} model overestimated by 5–8 μg m^{-3}.

3.1.3. Performance in Predicting Hourly Concentrations. SmartAQ^{2+} performance for hourly $PM_{2.5}$ concentrations was assessed based on forecasts issued 1 day in advance. Figure 5 illustrates the observed hourly $PM_{2.5}$ concentrations compared to the MB of SmartAQ^{2+} for winter, spring, summer, and autumn. In winter, hourly $PM_{2.5}$ concentrations above 30 μg m^{-3} occur mostly during afternoon and night hours at some locations and tend to be underestimated by SmartAQ^{2+} (MB = −10 μg m^{-3}, FBIAS = −30%). For $PM_{2.5}$ concentrations lower than 30 μg m^{-3} in winter, SmartAQ^{2+} performs well during the day (ME = 3 μg m^{-3}, FERROR = 27%) and night (ME = 4 μg m^{-3}, FERROR = 14%). In the spring, the observed hourly $PM_{2.5}$ concentrations range between 2 and 10 μg m^{-3} with few exceptions due to local events. SmartAQ^{2+} performs well with hourly bias less than 2.5 μg m^{-3} in most cases (Figure 5b). During the summer, SmartAQ^{2+} tends to underpredict $PM_{2.5}$

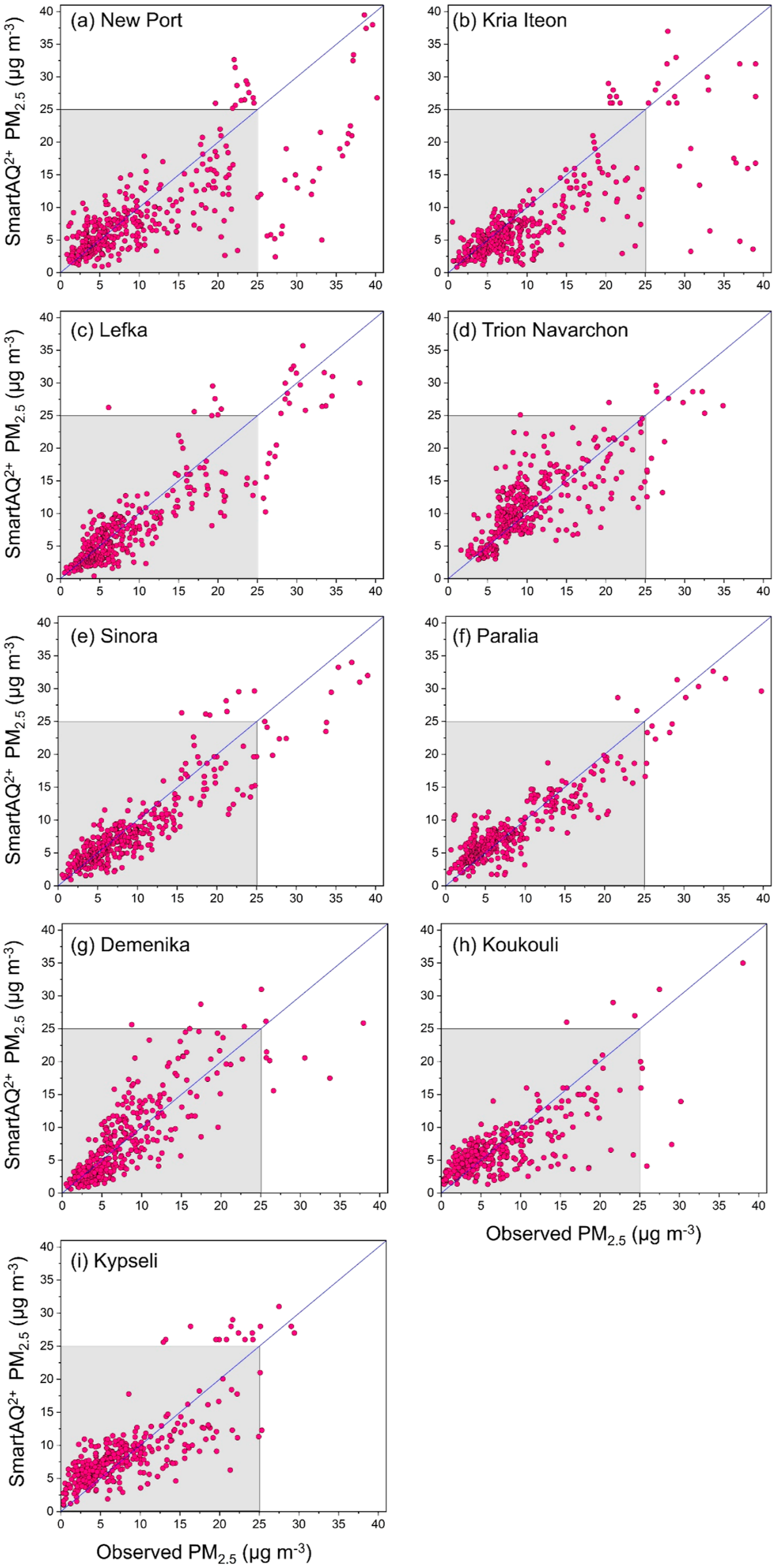


Figure 4. Daily $PM_{2.5}$ concentrations at the nine most polluted Patras sites during days when the daily average concentration was above 25 μg m^{-3}.

concentrations by 30% (ME = 3.5 μg m^{-3}). The fractional error of SmartAQ^{2+} during autumn remained less than 38%, with the mean error varying from 2 μg m^{-3} during the day to 4 μg m^{-3} during the night.

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX

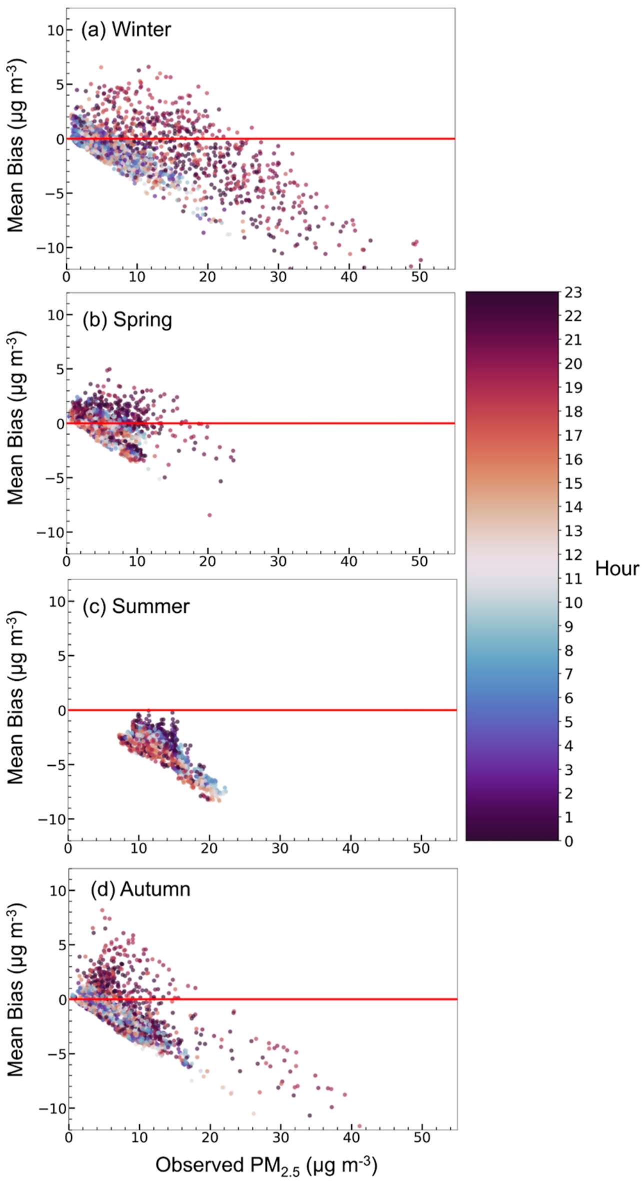


Figure 5. Hourly observed $PM_{2.5}$ concentrations compared to SmartAQ^{2+} MB for (a) winter, (b) spring, (c) summer, and (d) autumn. Please note the different scales in all graphs.

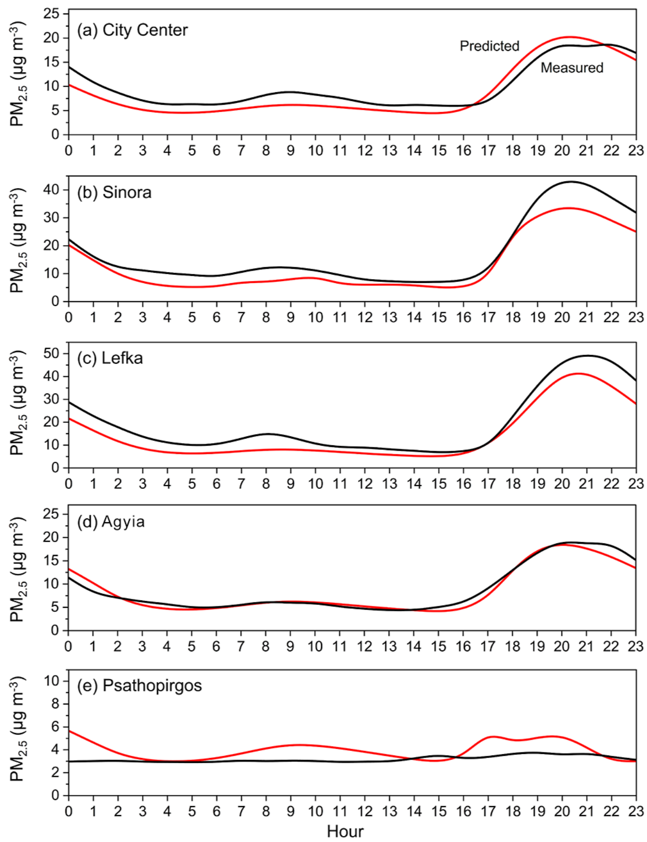


Figure 6. Average diurnal $PM_{2.5}$ profiles for February 2024 at (a) City Center, (b) Sinora, (c) Lefka, (d) Agyia, and (e) Psathopirgos. Different scales are used.

Figure 6 illustrates the average diurnal $PM_{2.5}$ profiles for the most polluted month of 2024 (February) at 5 selected sites (3 urban, 1 suburban, 1 background). Four of the sites (City Center, Sinora, Lefka, Agyia) are in Patras, where the majority of the study's sensors exist. Rio is located near Patras (~5 km from the center) and Psathopirgos is a background site 15 km from the city center. Elevated $PM_{2.5}$ concentrations were measured by the sensors at all five selected urban and suburban sites after 17:00, due to biomass burning, affecting most of the city. SmartAQ^{2+} captured these events with 35% FERROR and a mean R^2 (based on hourly averages) of 0.7.

3.1.4. Comparison with SmartAQ. Using forecasts issued 1 day in advance and seasonal mean $PM_{2.5}$ concentrations, SmartAQ^{2+} had lower errors compared to SmartAQ by an average of 50% (Figure S12). Overall, SmartAQ^{2+} reduced the persistent bias patterns seen in SmartAQ from +22% to +5%. For example, in winter, the $PM_{2.5}$ sensors measured approximately 18 $\mu g\ m^{-3}$ in the urban core and 4 $\mu g\ m^{-3}$ elsewhere. SmartAQ overpredicted $PM_{2.5}$ in the urban core (22 $\mu g\ m^{-3}$, 20% FERROR), whereas SmartAQ^{2+} was a lot closer (17 $\mu g\ m^{-3}$, 5% FERROR). In spring, SmartAQ^{2+} improved nonurban performance, decreasing SmartAQ's FERROR to 12% from 33%.

In terms of predicting the $PM_{2.5}$ daily limit exceedances based on forecasts issued 1 day in advance, SmartAQ^{2+} was better than SmartAQ. Over 54 exceedance days, SmartAQ^{2+} had 23 true positives and SmartAQ 18. Also, SmartAQ^{2+} reduced the false positives of SmartAQ from 31 to 12. At frequently polluted sites, SmartAQ^{2+} detected the exceedance days with an accuracy of 44% and SmartAQ with 16%. At the most polluted site (New Port), SmartAQ^{2+} identified 8 exceedances that SmartAQ missed.

3.1.5. Effect of Cooking Emissions on Model Performance in Patras. SmartAQ^{2+} and SmartAQ were run with and without inclusion of cooking emissions. We selected two sites for determining the effect of cooking emission information on the performance of SmartAQ^{2+} and SmartAQ and used the average daily $PM_{2.5}$ concentrations for the analysis. The first site is Trion Navarchon, an area with high cooking emissions[48] near the city center. The second site is Sinora, which has low cooking emissions and its distance from the city center is approximately 2 km.

In Trion Navarchon, SmartAQ^{2+} and SmartAQ benefited from the inclusion of cooking emissions with an average ME reduction of 30 and 40%, respectively (Figure S8). When inspecting the average of the rest of the sites, SmartAQ improved its performance in terms of ME by 20% and SmartAQ^{2+} by 15%. There were some sites (e.g., Sinora), where only the SmartAQ model improved its performance by approximately 10% when running with cooking emissions.

If cooking emissions are not available during deployment of SmartAQ^{2+} in other cities the errors of SmartAQ^{2+} are expected to increase by 30% at sites with heavy cooking activity and the citywide average performance of SmartAQ^{2+} might decrease by 15%.

3.1.6. Model Performance in Patras for Different Forecast Horizons. We quantified the performance of SmartAQ^{2+} using daily average $PM_{2.5}$ concentrations for winter (a period with high $PM_{2.5}$ concentrations) and summer (low concentrations) based on forecasts issued one, two, and 3 days in advance. Figure 7 compares the average predicted $PM_{2.5}$ for January and July 2024 based on SmartAQ^{2+} forecasts issued one, two, and 3 days in advance.

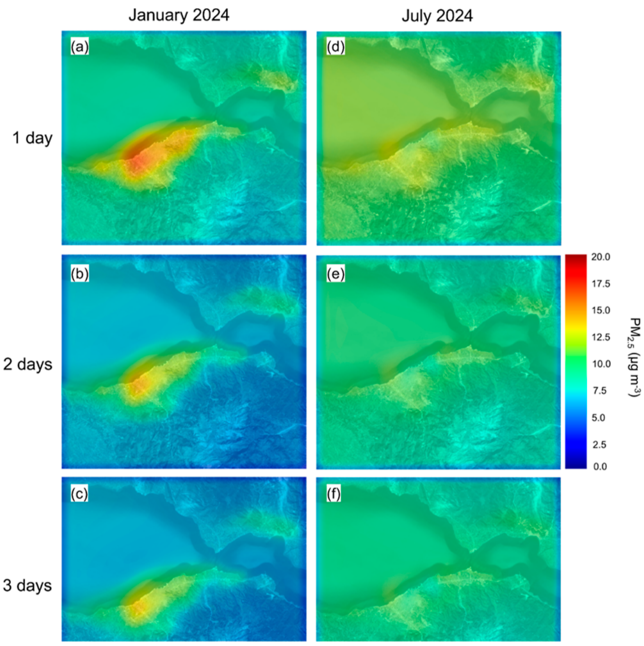


Figure 7. Average SmartAQ^{2+} predicted $PM_{2.5}$ concentrations for January 2024 and July 2024, based on forecasts issued (a,d) 1 day in advance, (b,e) 2 days in advance, and (c,f) 3 days in advance for the Patras domain.

In winter, the average SmartAQ^{2+} FERROR across sites rises from 40% (Day 1) to 41% (Day 2) and 43% (Day 3). The highest increase was at Kastellokampos (from 38% to 45%). The high concentration area in Figure 7a marks the urban core, while the suburbs lie to the right and left. In January, across forecasts issued one, two and 3 days in advance, the predicted $PM_{2.5}$ concentrations are lower in the urban core, the inner suburbs (such as Agios Andreas Hospital), and background sites (Figure S9) but remain practically the same at the outer suburbs (such as Kastellokampos). Hence, despite the overall ME of SmartAQ^{2+} during winter changing slightly between Day 1, 2, and 3, the ME in the urban core increased by 2 $\mu g\ m^{-3}$ and 3 $\mu g\ m^{-3}$ for Days 2 and 3, while remained approximately the same at the rest of the sites (Figure S9). In summer, the SmartAQ^{2+} average daily FERROR for all sites rises from 39% (Day 1) to 40% (Days 2 and 3). Average ME and FERROR increases were the same for urban and nonurban sites.

We then assessed the performance of SmartAQ^{2+} in predicting daily $PM_{2.5}$ limit exceedance for the second and third forecasted day (Table 1). True positive events based on forecasts issued 2 days in advance were 23, one less than those issued 1 day in advance. The false positive alarms remained 10 and the false negative alarms were 48. The SmartAQ^{2+} performance based on forecasts issued 3 days in advance was worse by approximately 35% than the first day (15 true positives, 13 false positives, 281 true negatives, and 56 false negatives).

3.2. Model Performance in Athens

We evaluated the performance of SmartAQ^{2+} in a pilot study for Athens, the largest Greek city, using measurements at 6 selected sites equipped with low-cost $PM_{2.5}$ sensors. These sites were Thiseio (center of Athens), Pagkrati (urban core), Piraeus (port of Athens), Haidari (suburb), Petroupoli (suburb), and Vrilissia (urban background). We selected these sites to represent urban, suburban, and urban background areas of Athens. Figure 8 illustrates the average predicted $PM_{2.5}$ concentrations by SmartAQ^{2+} for September 2025 in the Athens domain.

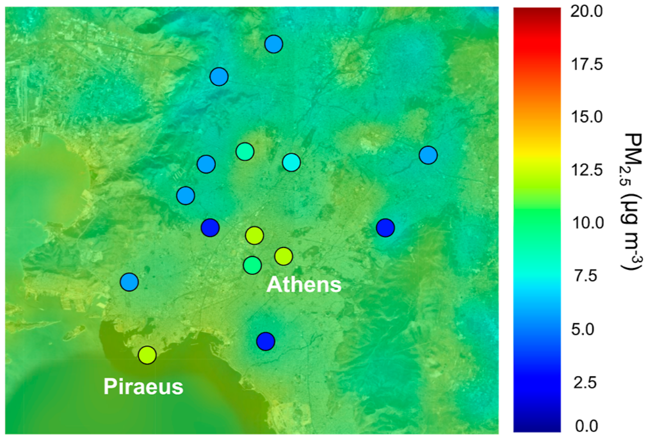


Figure 8. Predicted monthly average $PM_{2.5}$ concentrations by SmartAQ^{2+} for the Athens domain for September 2025. Colored circles denote monthly average $PM_{2.5}$ sensor measurements.

$PM_{2.5}$ sensors in the city center and at the port measured higher average monthly $PM_{2.5}$ concentrations (9−11 $\mu g\ m^{-3}$) compared with the outskirts of the city (4−6 $\mu g\ m^{-3}$). Similar concentrations were predicted by the SmartAQ^{2+} model, which estimated 10−12 $\mu g\ m^{-3}$ in the city center and 5−8 $\mu g\ m^{-3}$ in the surrounding areas. SmartAQ^{2+} was excellent for 4 sites (Thiseio, Haidari, Petroupoli, Vrilissia) and good for two (Pagkrati, Piraeus), based on daily averaged predicted and observed $PM_{2.5}$ concentrations (Figure S10).

The average SmartAQ^{2+} ME using daily concentration values for September 2025 for the six selected sites was 2 $\mu g\ m^{-3}$ (41%) and the MB was 0.3 $\mu g\ m^{-3}$ (7%). Hourly $PM_{2.5}$ concentrations are available in Figure S11.

3.3. Feature Importance

To interpret how SmartAQ^{2+} uses its inputs, we computed the SHAP values[51] for every 1 × 1 km^2 cell across the 36 × 36 km^2 domain of Patras. SHAP attributes each deviation from a reference prediction to specific inputs, yielding a consistent measure of feature influence. We evaluated how importances shift with local data availability by splitting cells into three groups: urban core, inner, and outer suburbs.

The relative importance of sensors is the highest (21−32%) among the other input features for sites in the urban core and the inner suburbs regardless of the season (Table S6). In the outer suburbs and background sites, sensor importance declines (9–16%). This behavior is expected, since many areas have limited sensor coverage and the model is forced to rely on other inputs.

The SmartAQ forecast for the next 3 days is consistently more important than the past three-day SmartAQ forecast, regardless of proximity to sensors and of the season. This is likely because, when processing the previous 3 days, the ML model can directly leverage the observed sensor readings to capture $PM_{2.5}$ patterns. In contrast, for the upcoming 3 days it must depend on the SmartAQ forecast as the primary driver.

Land use and emissions play the third most important role. Together, their relative importance is 16–20% in sites with sensor coverage and 20–28% in peripheral or background sites. Calendar variables, i.e., the month, weekdays, and hours of the forecasted days, have consistent importance, regardless of the site (around 10% during summer and 15% during winter). During winter, the relative importance of calendar variables becomes the second strongest predictor in the urban and close suburban core (around 17%). The ML model might have captured a periodic pattern of $PM_{2.5}$ concentrations during winter in the specific cells, which is related to the seasonality of biomass burning.

4. DISCUSSION

SmartAQ^{2+} combines physics-guided forecasts from SmartAQ (CTM) with real-time observations from low-cost sensors to improve 3 day hourly $PM_{2.5}$ concentration prediction at 1×1 km^2 resolution in urban areas. In contrast to our previously developed SmartAQ+ system, which provides only the current $PM_{2.5}$ concentration fields, SmartAQ^{2+} operates on temporal sequences providing multiple-day forecasts. It uses 72 h of past data, encodes the evolving pollution state, and projects it forward for 72 h. This requires the model to learn temporal dynamics (e.g., the build-up and decay of evening biomass-burning episodes, weekend–weekday emission shifts). The encoder–decoder architecture with attention allows SmartAQ^{2+} to selectively weight different past time steps when generating each future hour.

SmartAQ^{2+} improved seasonal, monthly, and hourly performance of SmartAQ across the Patras domain in 2024, by reducing seasonal ME by 50% and MB by 77%, monthly ME and MB by 35 and 55%, and hourly ME and MB by 35 and 25%, compared to its corresponding CTM. SmartAQ^{2+} better represented winter evening increases associated with biomass burning than SmartAQ but tended to underpredict the highest winter peaks at certain urban sites.

In 2024, SmartAQ^{2+} correctly identified 23 of 54 citywide $PM_{2.5}$ exceedance days (43%), significantly outperforming SmartAQ by doubling the true positives while reducing false alarms by over 60%. This predictive performance remained at a 2 day lead time before degrading by Day 3. Nearly half of the citywide missed exceedances (13 of 31) were "near-misses" within 3 μg m^{-3} of the threshold (25 μg m^{-3}). Its most important misses were at the new port of Patras, where it failed to capture 23 of 32 exceedance events related to maritime and heavy-duty vehicle traffic. The corresponding emissions in this area are probably not well captured by the used emissions inventory. False positives were primarily overestimations near the threshold, while the observed $PM_{2.5}$ concentrations were already elevated (18–24 μg m^{-3}).

SmartAQ^{2+} performance based on daily average $PM_{2.5}$ concentrations was retained for lead times of one and 2 days, with significant performance degradation (roughly 30%) after the second forecasted day. This behavior is related to objective errors in CTM and meteorology forecasts for lead times larger than 24 h. Custom ML algorithm loss functions that handle short-time and long-time errors differently could partially mitigate this issue.

Data fusion can reduce systematic CTM bias, but it cannot fully compensate for missing air pollution sources. Where a source is explicitly represented, SmartAQ^{2+} can learn how to adjust its contribution using local observations. When a source is absent, the ML model may still reproduce average patterns through correlated predictors, but it becomes less reliable during source-dominated episodes. For example, including cooking emission information in SmartAQ and as an input feature in SmartAQ^{2+} reduced daily mean error at high restaurant density areas by roughly 30%. If such information is absent during deployment, localized performance degradation is likely. The SHAP analysis indicates that SmartAQ^{2+} adapts to the available information, relying more on sensors in densely monitored areas and more on SmartAQ and spatial predictors in peripheral cells. This behavior is encouraging for operational use, but SHAP values should not be read as causal explanations, especially because several inputs are correlated. The results mainly provide a diagnostic of how the model distributes trust among its inputs across the domain, but in-depth ablation studies are needed to fully disentangle the contribution of each model component. These will be the topic of future model assessments.

The application of the ML model to Athens suggests that some relationships learned in Patras, such as the correction of broad CTM bias using recent observations and spatial predictors, may be reusable across Greek urban environments. Without retraining and with missing cooking inputs, the Patras-trained SmartAQ^{2+} achieved excellent to good performance at urban and suburban sites in Athens in September 2025, based on daily $PM_{2.5}$ concentration averages. However, the Athens evaluation was shorter and relied on fewer monitored locations, so the evidence should be interpreted as encouraging, but preliminary.

Sensor influence on the SmartAQ^{2+} predictions was highest in the urban core and close suburbs where the low-cost sensor network was dense and declined toward the periphery (sparser placement of low-cost sensors) where the model appropriately relied more on the CTM output and static spatial features. This is a desired behavior of an ML system that must perform over heterogeneous data density. Sensitivity to low-cost sensor density and placement is required in future analysis.

The findings are consistent with recent literature on hybrid CTM-ML frameworks for air quality forecasting. Earlier studies have shown that ML can downscale or correct CTM outputs for the present or future using a combination of meteorology, land use, satellite proxies, and regulatory measurements.[34,35] The present work adds an important element, which is the inclusion of low-cost sensor observations directly into a forecasting architecture rather than using only regulatory monitors or satellite products. Part of the strong monthly and seasonal performance of SmartAQ^{2+} might derive from the inclusion of the previous-year's same-month observed and SmartAQ-predicted $PM_{2.5}$ as input. This is particularly relevant for urban environments where small scale gradients and evening peaks from local sources can be difficult to capture.

While SmartAQ^{2+} improved $PM_{2.5}$ forecast accuracy, it essentially corrected the effects of biases in emission inventories and meteorological forecasts propagated in the CTM rather than their underlying causes. Because the ML model is trained and

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX

optimized for the current CTM behavior, any future updates to emission inventories or major changes in urban activity could shift the CTM output $PM_{2.5}$ concentrations that are not previously seen by the ML model. Periodic retraining is, therefore, mandatory to maintain reliability.

Extending the evaluation to additional Greek and Mediterranean cities with varying sensor densities would strengthen evidence for transfer learning and help define minimal observational requirements for stable performance. SmartAQ already provides source contributions via PSAT (e.g., cooking, biomass burning, long-range transport, traffic, marine). SmartAQ^{2+} currently corrects total $PM_{2.5}$ only, yet policy design depends on source attribution. Future research should correct the source apportionment, also.

A direct comparison against transformer-based or other architectures has been beyond the scope of this work. The quantification of the relative performance of these approaches is an interesting topic for future work.

■ ASSOCIATED CONTENT

Data Availability Statement

The study data are available upon request directed to S.N.P. (spyros@chemeng.upatras.gr).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsestair.6c00113.

> Detailed description of the machine learning algorithm; complete list of model input features (Table S1); seasonal FERROR and FBIAS per monitoring site (Tables S2 and S3); observed and predicted $PM_{2.5}$ concentrations during January and July 2024 (Table S4); $PM_{2.5}$ limit exceedance occurrences and corresponding model performance (Table S5); SHAP-based input feature importance (Table S6); low-cost sensor network of Patras (Figure S1); seasonal, daily, and hourly model evaluation at multiple sites (Figures S2−S7); effect of cooking emissions on model performance (Figure S8); performance of forecasts issued one to 3 days in advance (Figure S9); Athens application results (Figures S10 and S11); seasonal comparison of SmartAQ^{2+} and SmartAQ (Figure S12) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Spyros N. Pandis − *Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Patras 26504, Greece; Department of Chemical Engineering, University of Patras, Patras 26504, Greece;* orcid.org/0000-0001-8085-9795; Email: spyros@chemeng.upatras.gr

Authors

Ioannis D. Apostolopoulos − *Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Patras 26504, Greece*

Evangelia Siouti − *Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Patras 26504, Greece; Department of Chemical Engineering, University of Patras, Patras 26504, Greece*

George Fouskas − *Institute of Chemical Engineering Sciences, Foundation for Research and Technology Hellas, Patras 26504, Greece*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsestair.6c00113

Funding

This research was funded by the Bodossaki Foundation and the Social and Cultural Affairs Welfare Foundation (KIKPE) (project PANSEN). The open access publishing of this article is financially supported by HEAL-Link.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) 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. A review of operational, regional-scale, chemical weather forecasting models in Europe. *Atmos. Chem. Phys.* **2012**, *12*, 1−87.

(2) Delle Monache, L.; Alessandrini, S.; Djalalova, I.; Wilczak, J.; Knievel, J. C.; Kumar, R. Improving air quality predictions over the United States with an analog ensemble. *Weather Forecast.* **2020**, *35*, 2145−2162.

(3) Kaya, K.; Gündüz Öğüdücü, S.,. Deep Flexible Sequential (DFS) model for air pollution forecasting. *Sci. Rep.* **2020**, *10*, 3346.

(4) Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmos. Environ.* **2012**, *60*, 632−655.

(5) Hua, W.; Lou, S.; Huang, X.; Xue, L.; Ding, K.; Wang, Z.; Ding, A. Diagnosing uncertainties in global biomass burning emission inventories and their impact on modeled air pollutants. *Atmos. Chem. Phys.* **2024**, *24*, 6787−6807.

(6) Hamill, T. M.; Whitaker, J. S. Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Mon. Weather Rev.* **2006**, *134*, 3209−3229.

(7) Pappa, A.; Kioutsioukis, I. Forecasting particulate pollution in an urban area: From Copernicus to sub-km scale. *Atmosphere* **2021**, *12*, 881.

(8) Lee, H. J.; Liu, Y.; Coull, B. A.; Schwartz, J.; Koutrakis, P. A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations. *Atmos. Chem. Phys.* **2011**, *11*, 7991−8002.

(9) De Hoogh, K.; Saucy, A.; Shtein, A.; Schwartz, J.; West, E. A.; Strassmann, A.; Puhan, M.; Röösli, M.; Stafoggia, M.; Kloog, I. Predicting fine-scale daily NO_2 for 2005−2016 incorporating OMI satellite data across Switzerland. *Environ. Sci. Technol.* **2019**, *53*, 10279−10287.

(10) 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. Ensemble averaging using remote sensing data to model spatiotemporal PM_{10} concentrations in sparsely monitored South Africa. *Environ. Pollut.* **2022**, *310*, 119883.

(11) Tang, D.; Zhan, Y.; Yang, F. A review of machine learning for modeling air quality: Overlooked but important issues. *Atmos. Res.* **2024**, *300*, 107261.

(12) Voukantsis, D.; Karatzas, K.; Kukkonen, J.; Räsänen, T.; Karppinen, A.; Kolehmainen, M. 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.* **2011**, *409*, 1266−1276.

(13) Bai, Y.; Li, Y.; Wang, X.; Xie, J.; Li, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* **2016**, *7*, 557−566.

(14) Prasad, K.; Gorai, A. K.; Goyal, P. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmos. Environ.* **2016**, *128*, 246−262.

(15) Zhou, Y.; Chang, F.-J.; Chang, L.-C.; Kao, I.-F.; Wang, Y.-S.; Kang, C.-C. Multi-output support vector machine for regional multi-step-ahead $PM_{2.5}$ forecasting. *Sci. Total Environ.* **2019**, *651*, 230−240.

(16) Karimian, H.; Li, Q.; Wu, C.; Qi, Y.; Mo, Y.; Chen, G.; Zhang, X.; Sachdeva, S. Evaluation of different machine learning approaches to forecasting $PM_{2.5}$ mass concentrations. *Aerosol Air Qual. Res.* **2019**, *19*, 1400−1410.

(17) De Hoogh, K.; Héritier, H.; Stafoggia, M.; Künzli, N.; Kloog, I. Modelling daily $PM_{2.5}$ concentrations at high spatio-temporal resolution across Switzerland. *Environ. Pollut.* **2018**, *233*, 1147−1154.

(18) 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. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013−2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170−179.

(19) Requia, W. J.; Di, Q.; Silvern, R.; Kelly, J. T.; Koutrakis, P.; Mickley, L. J.; Sulprizio, M. P.; Amini, H.; Shi, L.; Schwartz, J. An ensemble learning approach for estimating high spatiotemporal resolution of ground-level ozone in the contiguous United States. *Environ. Sci. Technol.* **2020**, *54*, 11037−11047.

(20) Yu, W.; Li, S.; Ye, T.; Xu, R.; Song, J.; Guo, Y. Deep ensemble machine learning framework for the estimation of $PM_{2.5}$ concentrations. *Environ. Health Perspect.* **2022**, *130*, 037004.

(21) Li, L.; Wu, J. Spatiotemporal estimation of satellite-borne and ground-level NO_2 using full residual deep networks. *Remote Sens. Environ.* **2021**, *254*, 112257.

(22) Kerckhoffs, J.; Hoek, G.; Portengen, L.; Brunekreef, B.; Vermeulen, R. C. H. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. *Environ. Sci. Technol.* **2019**, *53*, 1413−1421.

(23) Fang, L.; Jin, J.; Segers, A.; Liao, H.; Li, K.; Xu, B.; Han, W.; Pang, M.; Lin, H. X. 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.* **2023**, *16*, 4867−4882.

(24) Tang, B.; Stanier, C. O.; Carmichael, G. R.; Gao, M. Ozone, nitrogen dioxide, and $PM_{2.5}$ estimation from observation-model machine learning fusion over S. Korea: Influence of observation density, chemical transport model resolution, and geostationary remotely sensed AOD. *Atmos. Environ.* **2024**, *331*, 120603.

(25) Hu, Y.; Tian, C.; Yue, X.; Lei, Y.; Cao, Y.; Xu, R.; Guo, Y. Global high-resolution fire-sourced $PM_{2.5}$ concentrations for 2000−2023. *Earth Syst. Sci. Data* **2025**, *17*, 3741−3756.

(26) Xing, J.; Baek, B. H.; Li, S.; Wang, C.-T.; Song, G.; Ma, S.; Zheng, S.; Liu, C.; Tong, D.; Woo, J.-H.; Liu, T.-Y.; Fu, J. S. A physically constrained deep-learning fusion method for estimating surface NO_2 concentration from satellite and ground monitors. *Environ. Sci. Technol.* **2024**, *58*, 21218−21228.

(27) Xu, M.; Jin, J.; Wang, G.; Segers, A.; Deng, T.; Lin, H. X. Machine learning based bias correction for numerical chemical transport models. *Atmos. Environ.* **2021**, *248*, 118022.

(28) Koo, Y.-S.; Kwon, H.-Y.; Bae, H.; Yun, H.-Y.; Choi, D.-R.; Yu, S.; Wang, K.-H.; Koo, J.-S.; Lee, J.-B.; Choi, M.-H.; Lee, J.-B. A development of $PM_{2.5}$ forecasting system in South Korea using chemical transport modeling and machine learning. *Asia-Pac. J. Atmospheric Sci.* **2023**, *59*, 577−595.

(29) Raffuse, S.; O'Neill, S.; Schmidt, R. A model for rapid $PM_{2.5}$ exposure estimates in wildfire conditions using routinely available data: rapidfire v0.1.3. *Geosci. Model Dev.* **2024**, *17*, 381−397.

(30) Dinkelacker, B. T.; Garcia Rivera, P.; Marshall, J. D.; Adams, P. J.; Pandis, S. N. High-resolution downscaling of source resolved $PM_{2.5}$ predictions using machine learning models. *Atmos. Environ.* **2023**, *310*, 119967.

(31) Bi, J.; Knowland, K. E.; Keller, C. A.; Liu, Y. Combining machine learning and numerical simulation for high-resolution $PM_{2.5}$ concentration forecast. *Environ. Sci. Technol.* **2022**, *56*, 1544−1556.

(32) Bertrand, J.-M.; Meleux, F.; Ung, A.; Descombes, G.; Colette, A. Technical note: Improving the European air quality forecast of the Copernicus Atmosphere Monitoring Service using machine learning techniques. *Atmos. Chem. Phys.* **2023**, *23*, 5317−5333.

(33) Lv, L.; Wei, P.; Li, J.; Hu, J. Application of machine learning algorithms to improve numerical simulation prediction of $PM_{2.5}$ and chemical components. *Atmos. Pollut. Res.* **2021**, *12*, 101211.

(34) Malings, C.; Knowland, K. E.; Pavlovic, N.; Coughlin, J. G.; King, D.; Keller, C.; Cohn, S.; Martin, R. V. Air quality estimation and forecasting via data fusion with uncertainty quantification: Theoretical framework and preliminary results. *J. Geophys. Res. Mach. Learn. Comput.* **2024**, *1*, No. e2024JH000183.

(35) Apostolopoulos, I. D.; Siouti, E.; Fouskas, G.; Pandis, S. N. 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. *Atmos. Environ.* **2026**, *364*, 121665.

(36) Siouti, E.; Kilafis, K.; Kioutsioukis, I.; Pandis, S. N. Simulation of the influence of residential biomass burning on air quality in an urban area. *Atmos. Environ.* **2023**, *309*, 119897.

(37) Siouti, E.; Skyllakou, K.; Kioutsioukis, I.; Patoulias, D.; Apostolopoulos, I. D.; Fouskas, G.; Pandis, S. N. 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* **2024**, *15*, 8.

(38) 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. *A Description of the Advanced Research WRF Model Version 4*. Version 4; University Corporation for Atmospheric Research, 2019..

(39) Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P. I.; Geron, C. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* **2006**, *6*, 3181−3210.

(40) Guenther, A. B.; Jiang, X.; Heald, C. L.; Sakulyanontvittaya, T.; Duhl, T.; Emmons, L. K.; Wang, X. 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.* **2012**, *5*, 1471−1492.

(41) Monahan, E. C.; Spiel, D. E.; Davidson, K. L. A model of marine aerosol generation via whitecaps and wave disruption. In *Oceanic Whitecaps' 'Oceanographic Sciences Library*; Monahan, E. C., Niocaill, G. M., Eds.; Springer: Netherlands, Dordrecht, 1986; pp 167−174.

(42) O'Dowd, C. D.; Langmann, B.; Varghese, S.; Scannell, C.; Ceburnis, D.; Facchini, M. C. A combined organic-inorganic sea-spray source function. *Geophys. Res. Lett.* **2008**, *35*, 2007GL030331.

(43) Kuenen, J.; Dellaert, S.; Visschedijk, A.; Jalkanen, J.-P.; Super, I.; Denier Van Der Gon, H. CAMS-REG-v4: a state-of-the-art high-resolution European emission inventory for air quality modelling. *Earth Syst. Sci. Data* **2022**, *14*, 491−515.

(44) Wagstrom, K. M.; Pandis, S. N.; Yarwood, G.; Wilson, G. M.; Morris, R. E. Development and application of a computationally efficient particulate matter apportionment algorithm in a three-dimensional chemical transport model. *Atmos. Environ.* **2008**, *42*, 5650−5659.

(45) Siouti, E.; Skyllakou, K.; Kioutsioukis, I.; Patoulias, D.; Fouskas, G.; Pandis, S. N. Development and application of the SmartAQ high-resolution air quality and source apportionment forecasting system for European urban areas. *Atmosphere* **2022**, *13*, 1693.

(46) Kosmopoulos, G.; Salamalikis, V.; Pandis, S. N.; Yannopoulos, P.; Bloutsos, A. A.; Kazantzidis, A. Low-cost sensors for measuring airborne particulate matter: Field evaluation and calibration at a South-Eastern European site. *Sci. Total Environ.* **2020**, *748*, 141396.

(47) European Commission, 2021. GISCO Population distribution. Available at: https://ec.europa.eu/eurostat/web/gisco/geodata/population-distribution/population-grids (access date 4 July 2026).

(48) Siouti, E.; Skyllakou, K.; Kioutsioukis, I.; Ciarelli, G.; Pandis, S. N. Simulation of the cooking organic aerosol concentration variability in an urban area. *Atmos. Environ.* **2021**, *265*, 118710.

(49) Siouti, E.; Skyllakou, K.; Patoulias, D.; Athanasopoulou, E.; Mihalopoulos, N.; Kuenen, J.; Pandis, S. N. High resolution source-resolved $PM_{2.5}$ spatial distribution and human exposure in a large urban area. *Atmos. Environ.* **2025**, *355*, 121277.

(50) Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, ARXIV.1409.0473.
(51) Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. *arXiv* **2017**, *30*, arXiv.1705.07874.
(52) Morris, R. E.; McNally, D. E.; Tesche, T. W.; Tonnesen, G.; Boylan, J. W.; Brewer, P. Preliminary evaluation of the community multiscale air quality model for 2002 over the southeastern United States. *J. Air Waste Manag. Assoc.* **2005**, *55*, 1694−1708.
(53) European Parliament. 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), 2024. Available at: https://eur-lex.europa.eu/eli/dir/2024/2881 (accessed on 4 July 2026).
(54) 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. Significant contributions of fresh and aged biomass burning organic aerosol from residential burning in a wintertime urban environment. *Atmos. Environ.* **2025**, *343*, 121018.

https://doi.org/10.1021/acsestair.6c00113
ACS EST Air XXXX, XXX, XXX−XXX