ACS ES&T Air

pubs.acs.org/estair

Article

Machine Education Approach for Generating Accurate NO_2 and $PM_{2.5}$ Pollution Maps in Israel

Published as part of ACS ES&T Air special issue "Elevating Atmospheric Chemistry Measurements and Modeling with Artificial Intelligence".

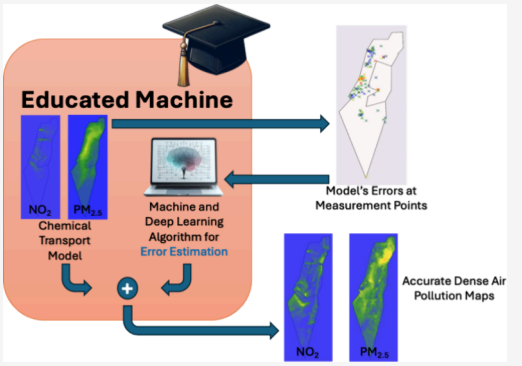
Avitay Geltman, Ilan Levy, and Barak Fishbain*

Cite This: ACS EST Air 2025, 2, 1411–1425

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Dense pollution maps are essential for understanding and reducing air pollution. However, pollution measurements are often sparse. Common methods to address this gap are chemistry transport models (CTMs) and air pollution interpolation models. Nonetheless, many of these models have poor performances and intrinsic systematic biases. Here, a machine education approach, which integrates a CTM and in situ measurements with Artificial Intelligence (AI) techniques, is proposed for generating dense pollution maps with enhanced accuracy. Specifically, the CHIMERE CTM is combined with a Neural Network, eXtreme Gradient Boosting (XGBoost), or Random Forest AI methods. The results show that the educated machine models significantly improved predictions of air pollution levels compared to the standalone CTM. The machine education model combining a neural network with CHIMERE performed best among all models, followed closely by the educated model with XGBoost, while the Random Forest model trailed. Relative to CHIMERE, the proposed models achieved reductions of up to 51.34% and 50.54% in the root mean square error (RMSE) and mean absolute error (MAE), respectively, for NO_2. For $PM_{2.5}$, the models demonstrated relative reductions of up to 40.08% and 36.54%, respectively. CHIMERE exhibited an inherent systematic underestimation bias, characterized by mean fractional bias (MFB) values of 0.509 and 0.725 for NO_2 and $PM_{2.5}$, respectively. Our models successfully eliminated this bias. Furthermore, promising dense air pollution maps were generated using our models.



KEYWORDS: *Artificial Intelligence, Machine Education, Machine Learning, Deep Learning, Air Pollution, CHIMERE Model, Neural Networks, Prediction Accuracy*

1. INTRODUCTION

Air pollution poses a threat to human health via both acute and chronic effects, causing harm to a wide range of systems and organs. Particulate matter (PM) is an air pollutant that consists of small particles suspended in the air. These particles vary in size and composition, and their sources are both natural and anthropogenic.[1] Exposure to $PM_{2.5}$ (particulate matter with a diameter of 2.5 μm or less) is a global predominant cause of premature mortality and a variety of serious health conditions, responsible for the death of millions of individuals annually.[2,3] This is particularly relevant to populous developing countries with extensive exposure of the population to the pollution of growing industries. An additional harmful air pollutant is nitrogen dioxide (NO_2). Short-term exposure to NO_2 is linked to respiratory morbidity, adverse cardiovascular effects, and premature mortality, while long-term exposure is linked to additional harms such as diabetes, poorer birth outcomes, and cancer.[4,5] Thus, mitigating air pollution, and particularly these two pollutants, has a critical role in sustaining the modern way of life, and measuring and monitoring air pollution is the first step toward maintaining public health.

Nowadays, air pollution is regularly monitored by Air Quality Monitoring stations (AQMs). While other means for measuring air pollution exist,[6] AQMs are still considered to be the gold standard, and the most common tool for measuring air pollution. Despite being accurate and reliable, a major drawback of the AQM stations is their sparse deployment due to their bulkiness, high cost, and professional maintenance requirements.[7] This limits the AQM network's ability to adequately capture highly resolved air pollutant spatial variations. Therefore, numerous efforts of air pollution modeling have been made over the years, many of which

Received: July 26, 2024
Revised: May 3, 2025
Accepted: May 5, 2025
Published: June 12, 2025

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411–1425

rely on mathematical, statistical, physical, and chemical theories, such as Chemistry Transport Models (CTMs).[8] With that, while the core equations for meteorological modeling have been well-established for many years, those for atmospheric chemistry are still evolving and being researched.[9] Accordingly, CTMs have inherent limitations and systematic biases, as they cannot account for all the complex and obscure phenomena and mechanisms that affect the propagation and dispersion of pollutants, such as topography, surroundings, gaseous flow, and chemical reactions.[10,11] Therefore, several studies have attempted to correct CTM outputs using various mathematical and statistical methods. One of those studies, utilizing an analog ensemble bias correction approach, was integrated into the operational U.S. National Oceanic and Atmospheric Administration (NOAA) National Air Quality Forecasting Capability (NAQFC) system.[12] The goal was to reduce biases and improve the Community Multiscale Air Quality (CMAQ)[13] model predictions of $PM_{2.5}$ over the contiguous United States. The analog ensemble approach assumes that, in stable climate conditions, model forecast errors in past similar weather events (or analogs) can be used to statistically adjust and correct current forecasts of the model.[14] Thus, in this case, an analog is essentially a past weather event with meteorological conditions similar to the current forecast scenario. For the search for analogs, three meteorological variables, which were found to be strong predictors, were used, as well as CMAQ $PM_{2.5}$ estimations. All four variables were assigned equal weights for the analog search procedure. This method was limited to a small number of variables. Thus, a big portion of the variables were neglected, meaning that some of the inherently biased behavior of the predictive model was left unaddressed. Furthermore, a significant number of samples did not have proper analogs, leading to corrections that were based on unsatisfactory resemblance between scenarios. Additionally, samples were associated with analog ensembles, and hence, the degrees of freedom were decreased, which impaired the model's ability to effectively generalize on extensive real-world data. Another study presented a different adaptation of the analog ensemble for forecasting $PM_{2.5}$ values by applying Kalman Filter (KF) correction to an analog ensemble weighted mean.[15] An inherent weakness of the KF is that it assumes that both the system and observation model equations are linear, which is not realistic in many real-life situations, especially in air pollution modeling, where chemical reactions are often characterized by nonlinearity. Additionally, in conventional KF, both process and observation noises are assumed to be white, Gaussian, additive, and mutually uncorrelated, which is another assumption to be placed under great doubt in air pollution modeling. Moreover, the methodology was tested using 12 months of data. It is strongly advised that air pollution forecasting applications utilize predictors covering a period of over a year, to account for seasonal factors that have a significant impact on air pollution levels.[10]

Neural networks (NNs) are a type of machine learning model inspired by the human brain. In air pollution forecasting, due to their ability to capture complex and nonlinear relationships between various factors such as weather conditions, emissions, and geographical features, they are used for providing accurate forecasts, aiding in pollution control and public health management. In the context of air pollution, researchers predicted averages of $PM_{2.5}$ concentrations on the US−Mexico border through three different topologies of neural networks: multilayer perceptron, radial basis function, and square multilayer perceptron.[16] In a different study, a Random Forest (RF) model with SHapley Additive explanation (SHAP) was used to identify key drivers of severe haze pollution, focusing on feature attribution and interpretability.[17] In contrast, our educated-machine approach aims to correct systematic biases in CHIMERE and generate dense pollution maps, prioritizing predictive accuracy over causal analysis. In a 2017 study, a fully connected feedforward NN approach was implemented as a means of air pollution forecasting and was compared with CMAQ.[18] Their model made use of meteorological variables and monitored $PM_{2.5}$ observations in the state of New York as predictors for predicting the $PM_{2.5}$ of the following day. To account for the seasonal variations, the month was also used as an input variable to the neural network. The authors made use of previously monitored $PM_{2.5}$ observations as inputs to their neural network but did not incorporate valuable information hidden within the bias of the CMAQ model. Moreover, only 5 meteorological variables were utilized for the training of the neural network model, which limits the ability to comprehend complex patterns. With the extensive utilization of NNs, the above body of work utilized NNs for air pollution prediction (forecasting), without the integration of any CTM. A recent work by Ajdour et al. utilized NNs and discrete wavelet transform to predict concentrations of a different air pollutant, ozone.[19] This work showed several limitations: the geographical scope was limited to the city of Casablanca, Morocco; the inputs to their NN architecture were solely CHIMERE's CTM ozone outputs and the ground-truth observations; and the training set data spanned across a short period of time of about 6 weeks, while a single week was used as the test set. Mhawish et al. integrated satellite-derived aerosol optical depth, columnar water vapor, and various meteorological variables to develop linear mixed effect and RF models for predicting daily $PM_{2.5}$ concentrations across the Indo-Gangetic Plain in South Asia.[20] However, their framework did not incorporate a CTM, thus isolating the physical and chemical processes from the modeling task.

CHIMERE is a multiscale chemistry transport model for atmospheric composition analysis and forecast developed by the Laboratoire de Météorologie Dynamique and other contributors.[21] CHIMERE is designed to estimate atmospheric pollutant concentrations over local to continental scales; CHIMERE conducts physicochemical simulations, incorporating meteorological inputs, pollutant emission fluxes, and chemical boundary conditions. CHIMERE has been widely used globally to model pollution levels across scales ranging from megacities to continental regions. In Israel, CHIMERE is deployed by the Israeli Ministry of Environmental Protection for estimating pollutant concentrations in the atmosphere throughout Israel. However, comparisons with AQM station measurements consistently reveal its spatially limited performance across diverse geographical regions.[22−24] To address the shortcomings of CHIMERE, a surrogate metamodel for CHIMERE, rather than correcting its output, has been suggested.[25] This essentially disassociates the problem from its physicochemical context, underscoring the need for a method that not only improves CHIMERE's performance and accuracy but also preserves its connection to the physics governing pollution systems. While machine learning has become a powerful tool for data analysis, its reliance on mathematical algorithms without understanding the underlying physical processes limits its flexibility across domains. To

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

bridge this gap, Kendler et al.[26] have introduced the machine-education approach, coupling Machine and Deep Learning (M&DL) architectures with CTMs, providing a solution that leverages both data-driven accuracy and physical interpretability.

Machine education differs from conventional machine learning by incorporating physical knowledge from models such as CTMs to guide the machine learning models in their learning process. This integration leverages domain-specific insights, such as pollutant dispersion dynamics, emissions, and meteorological influences, to help the machine learning models focus on correcting estimation errors rather than learning pollutant levels from scratch. This renders the educated machine accurate, and its generalization capabilities outperform classical machines.

Here, we utilize an educated machine approach for generating dense air pollution maps, where the machine capitalizes on a CTM and M&DL algorithms. Specifically, CHIMERE is used as the predictive CTM and NN, XGBoost, and Random Forest are considered for the M&DL component of the machine. The educated machine employs a coarse-to-fine approach by integrating CHIMERE's estimations with M&DL algorithms. The pollutant level is estimated by superimposing CHIMERE's output with the M&DL model error prediction. Specifically, CHIMERE's estimation is adjusted by adding the predicted error to it (Educated Machine Predicted Concentration = CHIMERE Base Estimation + Predicted CHIMERE Error), thereby refining the output to align more closely with the ground-truth measurements and enhancing accuracy and generalization by capturing CHIMERE's weaknesses alongside complex relationships in the data.

The dynamic range of the target space influences the complexity of the model, the adaptability of the network, and the speed at which the model is trained.[27−29] Therefore, the goal is to keep the dynamic range of the target space as small as possible. To keep the target space as small as possible, as mentioned above, we use the M&DL models to estimate CHIMERE's error, rather than the AQMs' measurements. However, it is important to note that the error, in this case, is on the order of magnitude of the actual pollutant values. Thus, while theory supports using the error rather than the actual value, the difference in this study is marginal.

The educated machines are trained using a comprehensive data set that includes CHIMERE's concentration estimates for several pollutants, AQM station in situ measurements, meteorological, geographical, and topographical variables, and temporal information. It is worth noting that CHIMERE's simulations themselves heavily rely on emission inventories, ensuring that emissions data is fed to our models indirectly. Here, our focus is directed toward the following harmful and concerning pollutants: $PM_{2.5}$ and NO_2 with the rationale of addressing both gaseous and particulate species.

2. METHODOLOGY

2.1. Data. Two data sources were used in this study. The first source is Israel's ambient AQM network, comprising over 200 stations operated by the Ministry of Environmental Protection, municipal air quality associations, and industry. All stations are ISO 17025 certified, ensuring routine maintenance and strict data quality assurance. For this study, 115 stations measuring NO_2 and 60 stations measuring $PM_{2.5}$ were included. The second source is CHIMERE's operational forecast output from 2018 to 2020, provided by the Ministry of Environmental Protection. The CHIMERE model (v2007) was run on a nested grid configuration, with the finest resolution offering 3 km hourly forecasts over Israel. Forecasts included 2 days worth of data plus a 12 h spin-up, totaling 60 h. Further details on CHIMERE's operational runs are available in a 2020 study.[30] For $PM_{2.5}$, the model output includes three components: biogenic, anthropogenic, and total particulate matter. This study uses the total $PM_{2.5}$, representing all contributions, comparable to the AQM in situ measurements. Meteorological variables were derived from the WRF (Weather Research and Forecasting) model, used as input to CHIMERE. CHIMERE's error was calculated by subtracting hourly CHIMERE forecasts from corresponding hourly AQM observations.

The data sets for NO_2 and $PM_{2.5}$ contained many features (see Supporting Information for the complete list), following is the partial list:

- CHIMERE's estimations for the following pollutants concentrations: carbon monoxide (CO), nitrogen oxide (NO), NO_2, ozone (O_3), PM_{10}, $PM_{2.5}$, and sulfur dioxide (SO_2).
- CHIMERE's estimation error at AQM station locations.
- Meteorological variables related to temperature, wind, relative humidity, air density, cloud water content and attenuation, precipitation, soil moisture, shortwave radiation, vertical diffusivity, etc.
- Geographical and topographical variables such as longitude, latitude, altitude, etc.
- Hour of the day.

The estimation error was set to be the target variable to be learned by the models, while all other variables were set to be inputs to our models.

The NO_2 data set contained 39,148 samples, dating from January 1, 2018, to December 23, 2020. For the equivalent period, the $PM_{2.5}$ data set contained 18,826 samples due to missing data. Each sample contained the above-mentioned variables for a certain AQM station at a certain timestamp. For instance, on January 1, 2018, at 6 AM, 53 different AQM stations produced NO_2 measurements; hence, the data set contains 53 samples for this timestamp. Similarly, at 11 AM that day, 83 stations recorded measurements, resulting in 83 samples in the data set. Typically, for each day there are 1−2 different timestamps at random varying times, for which there are samples. The units of measurement for the pollutant concentrations are ppb (parts per billion) and $\mu g/m^3$ for NO_2 and $PM_{2.5}$, respectively.

The complete data set was randomly divided into two subsets: a training set containing 75% of the data, which was used to develop and train the models, and a test set containing the remaining 25%, which was used to evaluate the models' performance on previously unseen data. This standard division ensures an unbiased assessment of the models' predictive capabilities while maintaining sufficient data for effective training. Although the training set was already created through a random split of the data set, it was further shuffled as a safeguard to ensure complete randomness and diversity within the data. This step is particularly relevant for Neural Networks (NNs), which are trained iteratively on batches of data and benefit from randomized exposure during training. The test set, however, was not shuffled, as performance evaluation is

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

conducted on the entire data set, and shuffling would not impact the results.

For NO_2, training and test sets consisted of 29,361 and 9,787 samples, respectively. For $PM_{2.5}$, training and test sets consisted of 14,119 and 4,707 samples, respectively. Both the training and test sets were standardized, providing a mean of 0 and a standard deviation of 1 for all explanatory variables. The target (or response) variable, i.e., the estimation error, was left unstandardized.

Another aspect of interest was to evaluate the significance of employing a random split for dividing the data into train and test sets, as opposed to a temporal split strategy. A temporal split involves partitioning the data based on chronological order, such as using the initial 25% of observations as the test set and the subsequent 75% as the training set or vice versa. The chosen allocation was to have the initial 25% of observations as the test set and the subsequent 75% as the training set rather than employing the reverse split, wherein the initial 75% serves as the training set and the last 25% as the test set. This decision was driven by the fact that the final 25% of the data corresponded to the onset of the COVID-19 epidemic surge, a period characterized by significant disruptions and anomalies in air pollution patterns. By incorporating the COVID-19 period into the training set, its impact was distributed across a larger data set, reducing its influence and ensuring more reliable and unbiased model evaluation. Conversely, including this biased and atypical period in the test set would disproportionately influence the evaluation of the models, given that the test set is three times smaller than the train set. Thereafter, the models underwent training and testing based on this split.

Feature selection algorithms have been applied to identify the optimal subset of features for the task of learning CHIMERE's estimation errors. No improvement has been observed, and thus all features were used for the analysis. The feature selection algorithms and the results are reported in the Supporting Information.

2.2. Educated Machines. Educated machines are based on the integration of CTM into a machine learning framework. To accomplish that, several methodologies have been reported. A 2017 study suggested training the network using both real and synthetic data derived from a physical model.[31] Physics-Informed Neural Networks (PINNs) leverage governing physical equations in the network's training phase, in which the network must fit observed data while reducing Partial Differential Equation (PDE) residuals.[32] A recent study suggested an approach where preprocessing was applied on video sequences to extract relevant features, thus alleviating the need to find these features from the raw data by the network.[33] The coarse to fine approach was also presented. To this end, researchers used the European Center for Medium-Range Weather Forecasts (ECMWF) Object-Oriented Prediction System, as a primary, coarse, estimation, and then a NN for refining the estimation.[34] Similarly, Farchi et al. presented data assimilation methodologies for the primary estimation, which were then fine-tuned by a fully connected neural network.[34,35]

The use of educated machines essentially guides the NN in the solution process, rendering simpler the problem to be solved by the NN itself. This, in return, allows for (i) training the network with data that does not necessarily need to be large nor complete,[36] (ii) shorter training times[27,29] and (iii) the use of simpler network architectures.[28,29] Thus, with some knowledge about the physical characteristics of the problem and some form of training data (even sparse and incomplete), educated machines may be used for finding an optimal solution with high fidelity.

Here we present the use of the coarse-to-fine approach, consisting of the CHIMERE dispersion model combined with either NN, eXtreme Gradient Boosting (XGBoost), or Random Forest (RF). NNs have been chosen due to their ability to capture complex nonlinear relationships and learn intricate patterns in data. Through this, NNs are capable of learning and recognizing patterns in data and can be used for a variety of tasks, including predictive modeling.[37] XGBoost and RF algorithms were chosen for their ensemble learning capabilities and their ability to handle diverse data characteristics. These algorithms excel in capturing both linear and nonlinear relationships, enabling effective modeling of complex patterns in the estimation error. Moreover, their ensemble nature fosters enhanced model robustness and reduces the risk of overfitting, resulting in improved generalization and predictive performance.

2.3. Neural Networks. Capitalizing on the advantages of educated machines, a simple fully connected linear neural network was implemented with similar architecture for both NO_2 and $PM_{2.5}$, as it was found to provide the best results for both pollutants. A manual search was conducted for the network's hyperparameters tuning. The search included variations of number of hidden layers, number of neurons in each layer, batch size, activation function, learning rate, loss function, and optimizer. As part of the training process, the loss on the test set was tracked at the end of every training epoch, as well as the training loss. For each epoch in which the network produced a validation loss lower than the minimal validation loss obtained up to that step, that model was saved and the validation loss as well as the epoch number were recorded. Two stop criteria were defined for the training process; the first was when the training process reached the predefined last epoch. The second was when validation loss over the last 40 epochs did not go below the current minimal validation loss obtained, meaning that the best model has already been achieved. The latter criterion was defined for computational efficiency. The networks' chosen architecture was 3 hidden layers, 200 neurons in each hidden layer, input layer neurons as the number of features, i.e., 40 neurons, output layer of 1 single neuron (the target variable), learning rate of 0.001, batch size of 64, Adam optimizer, MSE loss function, and ReLU activation function.

2.4. Decision Tree Based Algorithms. *2.4.1. Decision Trees.* Decision Trees (DTs) are a fundamental machine learning algorithm that constructs hierarchical decision structures to solve classification or regression problems.[38] They recursively partition the input data based on feature values, creating a tree-like structure where each internal node represents a decision based on a specific feature and each leaf node represents a final prediction or outcome. DTs are characterized by their simplicity, interpretability, and ability to handle both categorical and numerical features. They are particularly effective at capturing complex decision boundaries and interactions among features. One of the key advantages of DTs is their ability to handle both linear and nonlinear relationships in the data. However, they can be prone to overfitting, especially when the tree becomes deep and complex. To mitigate this, ensemble methods, such as Random Forest or XGBoost are often employed.

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

2.4.2. Random Forest. Random Forest is an ensemble learning method that builds upon the foundation of DTs.[39] Ensemble methods are designed to improve the performance of models by combining multiple weaker learners (models) to create a single stronger learner. The forests are composed of multiple individual DTs, where each tree is constructed using a different subset of the training data and a random subset of features. Each DT in the RF independently makes predictions, and the final prediction is determined by aggregating the results from all trees. RF harnesses the power of ensemble learning to improve predictive accuracy and mitigate the shortcomings of individual DTs by combining the predictions of diverse trees. This provides robustness against overfitting and enhances generalization. Moreover, the randomness injected during training helps capture different aspects of the data, resulting in a more comprehensive modeling of complex relationships.

To establish a profound RF model, an approach of grid search with *k*-fold cross validation[40] was utilized for the models' hyperparameter tuning, both for NO_2 and $PM_{2.5}$. The goal was to find the combination of hyperparameters that resulted in the best performance of the model on the validation set. Once the model is trained and evaluated using all the possible hyperparameter combinations, the model with the best performance is selected and used to make predictions on new data. The grid search hyperparameters chosen for the construction of the RF model were the following: maximum depth of a tree, number of features to consider when looking for the best split at each node, number of trees constructing the forest, minimum number of samples required to split an internal node, and minimum number of samples required to be at a leaf node. The best performing model, i.e. the lowest MSE achieved over the *k*-fold cross validation, was a RF with the following hyperparameters: for $PM_{2.5}$, maximum depth of 50 nodes for each tree, the amount of features to consider for a node split being the square root of the total amount of features, 150 trees constructing the forest, minimum of 5 samples required to split an internal node, minimum of 1 sample required to be at a leaf node; for NO_2, maximum depth of 50 nodes for each tree, number of features to consider for a node split being all features, 150 trees constructing the forest, minimum of 2 samples required to split an internal node, minimum of 4 samples required to be at a leaf node.

2.4.3. eXtreme Gradient Boosting: XGBoost. XGBoost is an advanced gradient boosting framework that leverages the strengths of DTs to deliver enhanced predictive performance.[41] Similar to RF, XGBoost utilizes an ensemble approach but with a different strategy called boosting. Boosting involves sequentially adding DTs to the ensemble, where each subsequent tree aims to correct the errors made by the preceding trees. By iteratively refining the model, XGBoost adapts and learns from previous mistakes, ultimately converging to a highly accurate and robust prediction model. XGBoost employs gradient-based optimization for precise adjustments during each boosting iteration, ensuring superior convergence and efficiency. Furthermore, XGBoost utilizes regularization techniques to control the model complexity and mitigate overfitting, enhancing generalization capabilities.

A 3-fold cross validation was chosen for finding XGBoost hyperparameters for both pollutants. The grid search hyperparameters chosen for the construction of the XGBoost model were as follows: maximum depth of a tree; η, a regularization parameter that shrinks feature weights in each boosting step; subsample ratio of the training instances, for example, a chosen subsample ratio of 0.5 means that the XGBoost would randomly sample half of the training data prior to growing trees; subsample ratio of features when constructing each tree, for example, for a chosen subsample ratio of 0.5, XGBoost will randomly select 50% of the features to build each tree and the remaining 50% of the features will be ignored; γ, minimum loss reduction required to make a further partition on a leaf node of the tree; α, L_1 regularization term added to the loss function; λ, L_2 regularization term added to the loss function; number of trees constructing the XGBoost model; number of parallel trees constructed during each iteration; and booster type. Regarding the latter, two types of boosters were compared: Gradient-boosting tree, gbtree; and Dropouts meet Multiple Additive Regression Trees, dart. The main difference between the two boosters is that the dart booster utilizes dropout regularization by introducing randomness in the training process by dropping out random trees during training, which helps to reduce overfitting and improve generalization performance. In contrast, the gbtree booster builds trees sequentially without introducing randomness.

The best performing model was an XGBoost with the following hyperparameters: for $PM_{2.5}$, maximum depth of 8 nodes for each tree, η of 0.1, 1.0 subsample ratio of training instances, 0.75 subsample ratio of features, γ of 0, α of 0.5, λ of 1.0, 200 trees constructing the model, 1 tree constructed during each iteration, booster type of dart; for NO_2, maximum depth of 8 nodes for each tree, η of 0, 0.7 subsample ratio of training instances, 1.0 subsample ratio of features, γ of 5, α of 0, λ of 1.0, 100 trees constructing the model, 3 trees constructed parallelly during each iteration, booster type of dart.

2.5. Performance Evaluation. To evaluate the overall performance of our models, numerous tests and metrics were utilized:

1. **Mean Absolute Error** (MAE), a commonly used metric for evaluating the performance of regression models. It measures the average absolute difference between the predicted and true values of the target variable. It is given by

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |O_i - E_i| \tag{1}$$

where O_i and E_i denote observed and estimated concentration of sample i, respectively. N is the total number of samples.

2. **Root Mean Square Error** (RMSE), another widely used metric for evaluating performance. It measures the square root of the average of the squared differences between the predicted and true values of the target variable. Compared to the MAE, RMSE penalizes larger errors more heavily due to the squaring operation. It is given by

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (O_i - E_i)^2}{N}} \tag{2}$$

3. **Mean Fractional Bias** (MFB), a dimensionless number estimating a model's underestimation or overestimation of the ground truth, essentially indicating an inherent systematic error. It is given by

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

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

which also translates to

$$\mathrm{MFB} = 2 \times \frac{\mu_O - \mu_E}{\mu_O + \mu_E} \quad (4)$$

where μ_O and μ_E are the overall mean of the observed and estimated concentrations, respectively. MFB values range between −2 and +2, where a value of 0 is ideal and indicates that the model is unbiased, a positive value indicates that the model is underestimating the ground truth values, and a negative value indicates that the model is overestimating the ground truth values.

4. **Cumulative Distribution Function** (CDF), a plot depicting the cumulative frequency of AQM stations for which the model mean estimation error is below the x axis threshold. In other words, the y axis depicts the cumulative frequency of the AQM stations, while the x axis depicts mean estimation errors. Plotting the CDF enables us to understand what portion of the stations fall in a desirable range of mean estimation errors for the various models.
5. **Distribution plot**, visualizing the distribution of estimation errors for both test sets (NO_2 and $PM_{2.5}$) providing a comprehensive overview of the spread and concentration of estimation errors across the entire test sets. Examining the distribution allows for the identification of central tendencies, outliers, and the overall shape of the error distribution. Moreover, it serves as a valuable complement to the CDF, offering additional insights into the overall predictive behavior of the models. Ideally, it would be desirable for a model to exhibit the following characteristics:
    - Most estimation errors should be centered around zero, indicating that, on average, the model's predictions are accurate.
    - A narrower spread of estimation errors (smaller standard deviation), signifying that the model's predictions are consistent and have low variability. In other words, the model is making precise predictions.
    - Minimal occurrences of extreme outliers.
6. **Spatial performance plots**, for visual assessment of the results, two plots, one for each pollutant, visualizing the average absolute estimation errors across all geographically scattered AQM sites throughout Israel and in addition two dense pollution maps covering the entire grid of national boundaries, including locations where no AQM stations are deployed.

The educated machine's performance was compared to simpler alternatives; two additional methods were introduced into our testing. The first method, "Simplistic CHIMERE", involved a straightforward approach, where CHIMERE's estimations were shifted by the mean estimation error observed across all test set observations. The second method employed an Elastic Net model, serving as a regularized linear regression technique, implemented in the interest of affirming the benefit of leveraging M&DL methodologies capable of capturing complex nonlinear relationships within the data. The Elastic Net model was configured with a 0.5 ratio of L_1 and L_2 penalties (weighing them equally), and the constant that multiplies the penalty terms was set to 1.0. To ensure that the results were not dependent on a single random split of the data into training and test sets, five independent random splits were performed. Each split maintained a 75%/25% training/test ratio, and for each split, separate M&DL models were trained and evaluated on their corresponding training and test sets. The performance metrics reported in the study represent the average results across these five splits, providing a robust assessment of the models.

3. RESULTS AND DISCUSSION

The performances of the models based on the metrics of MAE, RMSE, and MFB are shown in Table 1 (NO_2) and Table 2 ($PM_{2.5}$). With respect to NO_2, while CHIMERE displays a clear inherent systematic bias demonstrated by an MFB of 0.509, all evaluated models achieve successful elimination of this phenomenon, with MFBs close to zero. The positive value of the MFB implies CHIMERE's tendency to underestimate the ground truth. The MFB of the simplistic CHIMERE is not reported because it was artificially forced to be zero when the correction of the estimations was introduced merely by the mean estimation error of plain CHIMERE. Overall, CHIMERE demonstrated the poorest performance of all models, highlighting the clear need for its correction. The best performance was assuredly achieved by the NN-based educated machine model, demonstrating both the lowest MAE value as well as the lowest RMSE. Next, in a look at the MAE and RMSE results, the XGBoost (XGB) educated machine model outperformed the RF model. Subsequently, the RF model outperformed the Elastic Net. Interestingly, the simplistic CHIMERE achieved a lower RMSE than plain CHIMERE, but marginally higher MAE. This suggests that this straightforward correction was slightly effective solely at mitigating the impact of larger errors.

Compared to CHIMERE, the NN-, XGBoost-, and RF-based models yield relative reductions of 50.54%, 35.29%, and 30.51% in MAE, respectively, and relative reductions of

Table 1. MAE, RMSE, and MFB Results for NO_2[a]

	MAE [ppb]	RMSE [ppb]	MFB
CHIMERE	6.49(0.04)	12.27(0.57)	0.509(0.005)
Simplistic CHIMERE	6.54(0.06)	11.68(0.60)	
Elastic Net	6.16(0.05)	11.00(0.62)	−0.003(0.004)
RF	4.51(0.15)	8.18(1.02)	−0.002(0.022)
XGB	4.20(0.07)	7.65(0.85)	0.000(0.0290)
NN	3.21(0.04)	5.97(0.72)	0.001(0.004)

[a]Results are presented as mean (STD). Performance rankings appear in a vertical order of worst to best.

Table 2. MAE, RMSE, and MFB results for $PM_{2.5}$[a]

	MAE [$\mu g/m^3$]	RMSE [$\mu g/m^3$]	MFB
CHIMERE	12.15(0.14)	19.71(0.54)	0.725(0.008)
Simplistic CHIMERE	10.03(0.12)	17.00(0.67)	
Elastic Net	9.36(0.11)	13.97(0.13)	0.000(0.009)
RF	8.14(0.15)	12.74(0.31)	−0.043(0.011)
XGB	7.84(0.14)	12.31(0.58)	−0.013(0.022)
NN	7.71(0.13)	11.81(0.11)	0.023(0.014)

[a]Results are presented as mean (STD). Performance rankings appear in a vertical order of worst to best.

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

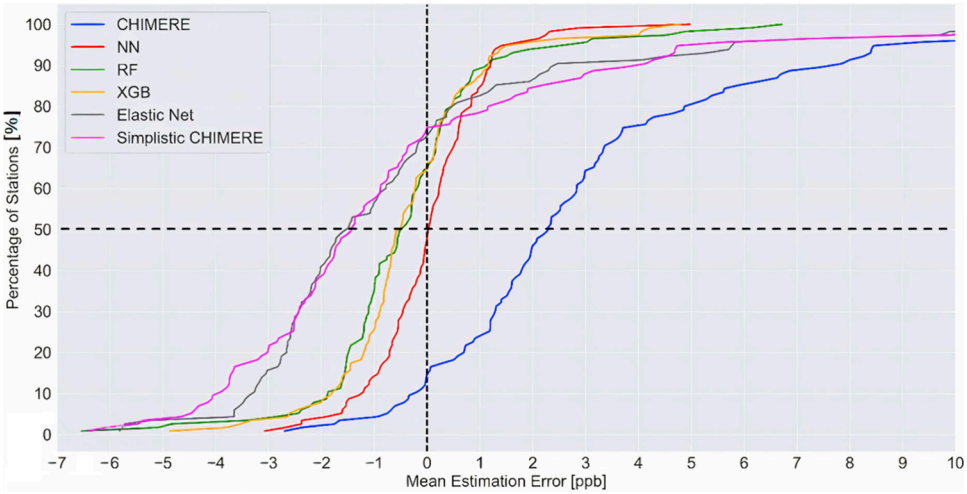


Figure 1. CDF of the percentage of AQM stations for which the models' mean absolute error is below the corresponding threshold on the *x*-axis, constructed of all NO_2 AQM stations.

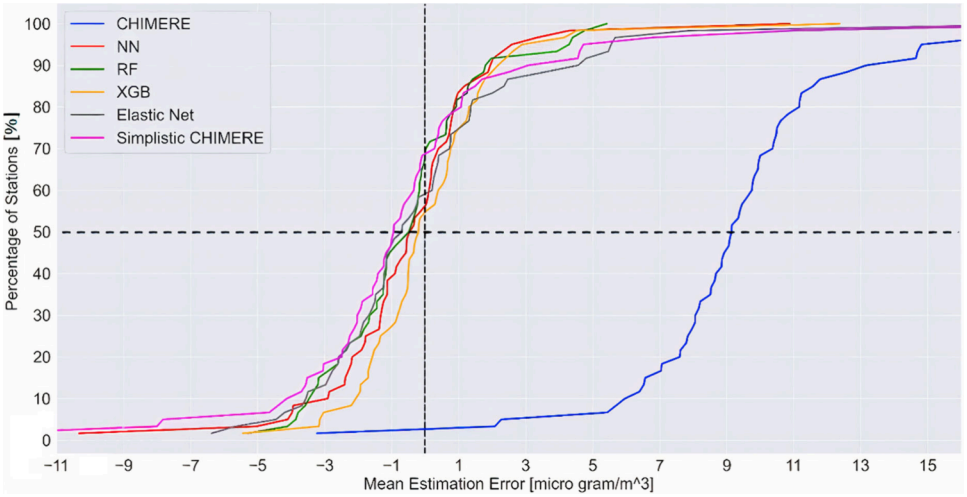


Figure 2. CDF of the percentage of AQM stations for which the models' mean absolute error is below the corresponding threshold on the *X*-axis, constructed of all $PM_{2.5}$ AQM stations

51.34%, 37.65%, and 33.33% in RMSE, respectively. Elastic Net achieved far inferior relative reductions of 5.08% in MAE and 10.35% in RMSE. While simplistic CHIMERE did not demonstrate an improvement in the MAE metric, it achieved a relative reduction of 4.81% in RMSE, which is negligible compared to that of the NN-based model.

Regarding $PM_{2.5}$, the inherent systematic bias of CHIMERE and its tendency to underestimate the ground truth was even more significant, demonstrated by an MFB of 0.725. Despite that, all models successfully resolved the systematic bias with MFB values close to zero. Overall, CHIMERE yet again performed worse than all other models, for all metrics. The NN-educated machine model once more demonstrated its superiority over the other models, although not as prominently as in the case of NO_2, as its results are only slightly better than XGBoost's results. Subsequently, the XGBoost model performed slightly better than the RF-based model. Thereafter, RF outperformed Elastic Net, and simplistic CHIMERE ranked last, as it did in the case of NO_2. Nonetheless, with respect to the $PM_{2.5}$ task, simplistic CHIMERE indeed demonstrated a slight improvement in performance in comparison to plain CHIMERE. Compared to CHIMERE, the NN-, XGBoost-, and RF-based models yielded relative reductions of 36.54%, 35.47%, and 33.00% in MAE, respectively, and relative reductions of 40.08%, 37.54%, and 35.36% in RMSE, respectively. Elastic Net achieved relative reductions of 22.96% in MAE and 29.12% in RMSE. Simplistic CHIMERE achieved relative reductions of 17.45% in MAE and 13.75% in RMSE.

Regarding the comparison between the models' performances when trained using random data splits versus a temporal data split, the former demonstrated a clear advantage. Detailed results and a discussion are provided in the Supporting Information.

Performance in terms of the CDF is illustrated in Figure 1 (NO_2) and Figure 2 ($PM_{2.5}$), and a more thorough analysis is presented in Table 3 (NO_2) and Table 4 ($PM_{2.5}$), depicting, for each educated model, the percentage of AQM stations for which their mean estimation error is below varying thresholds (absolute-wise), using all AQM stations measuring the relevant pollutant (115 and 60 for NO_2 and $PM_{2.5}$, respectively). CHIMERE's previously disclosed systematic bias is evident for both NO_2 and $PM_{2.5}$. In accordance with the aforementioned MFB values, it is notably more substantial for $PM_{2.5}$, as its

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

Table 3. Percentages of AQM Stations for Which the Models' NO_2 Mean Estimation Error Is Less than Varying Thresholds, Absolute-wise ($|\varepsilon| <$ Threshold)

	threshold [ppb]				
	1	3	4	7	10
CHIMERE	20.00	64.35	74.78	88.70	95.65
Simplistic CHIMERE	20.87	66.96	80.00	95.65	97.39
NN	70.43	98.26	99.13	100.00	100.00
RF	53.91	91.30	93.91	100.00	100.00
XGB	61.74	93.04	95.65	100.00	100.00
Elastic Net	26.09	74.78	86.96	95.65	98.26

Table 4. Percentages of AQM Stations for Which the Models' $PM_{2.5}$ Mean Estimation Error Is Less than Varying Thresholds, Absolute-Wise ($|\varepsilon| <$ Threshold)

	threshold [$\mu g/m^3$]				
	1	3	4	7	10
CHIMERE	0.00	2.98	3.98	15.12	67.80
Simplistic CHIMERE	30.00	71.11	80.85	91.67	95.00
NN	45.00	86.67	91.67	96.67	96.67
RF	36.67	76.67	90.00	100.00	100.00
XGB	46.67	88.33	93.33	98.33	98.33
Elastic Net	26.67	75.00	81.67	96.67	98.33

curve is far to the right-hand side of the y axis (Percentage of Stations), meaning it greatly underestimates the ground truth. In fact, the bias is so significant that out of the 60 AQM stations monitoring $PM_{2.5}$, for 59 of them, the mean estimation error was positive. Regarding NO_2, while CHIMERE's CDF curve is much closer to the y axis compared to the case of $PM_{2.5}$, the curve is characterized by a significantly wider stretch, meaning a high variance. With respect to $PM_{2.5}$ the curves of all educated models demonstrate elimination of CHIMERE's severe underestimation bias. While CHIMERE undoubtedly demonstrates the worst performance, the XGBoost-based model exhibits the best performance among all models, reflected in the narrowest curve stretch and the closest intersection with the y axis at the 50%−50% point, meaning an almost equal division of AQM stations having a negative and positive mean estimation error. Ranked second is the NN-based model; thereafter, the RF and Elastic Net demonstrate comparable performance, and simplistic CHIMERE falls short among the others. Nonetheless, there is a high resemblance among the curves of all models, excluding CHIMERE. This observation gives more ground to the previously mentioned suggestion that the properties of the $PM_{2.5}$ data are less characterized by complex nonlinear relationships.

In the case of NO_2, the NN educated machine-based model demonstrates the best performance, intersecting the y axis exactly at the 50%−50% point, while also having the narrowest curve stretch. Subsequently, the RF and XGBoost-based models were second, having a curve stretch almost as narrow as the NN, but with a slight offset into the negative section of the mean estimation error. Thereafter, with inferior performances are the Elastic Net and simplistic CHIMERE, demonstrating a significantly different curve pattern, reflected in a wide curve stretch and a substantial offset into the negative section of the mean estimation error. While the educated machine models display successful narrowing of CHIMERE's curve, implying a meaningful reduction in variance, there is practically none of this effect occurring for the Elastic Net. Simplistic CHIMERE's curve is merely a shift of plain CHIMERE's curve. The evident differences between the curves of the educated models and the Elastic Net, which is a linear model, implies that NO_2 is most probably characterized by complex nonlinear relationships, as previously suggested.

Table 3 (NO_2) and Table 4 ($PM_{2.5}$) present the superior performance, for both pollutants, of the educated models over the Elastic Net and simplistic CHIMERE, which, in turn, outperform plain CHIMERE. For both pollutants, the performances of the NN- and XGBoost-based educated models are quite comparable, while outperforming the RF model. A visualization of a quantile analysis can be seen in Figure 3 and Figure 4, essentially complementing the information presented in Table 3 and Table 4, depicting the percentage of AQM stations for which the models' mean estimation error (absolute-wise) is below a given threshold. Examining each quantile for both pollutants, it is noteworthy that the curve representing CHIMERE consistently lies above the curves of all models. This positioning argues that for any

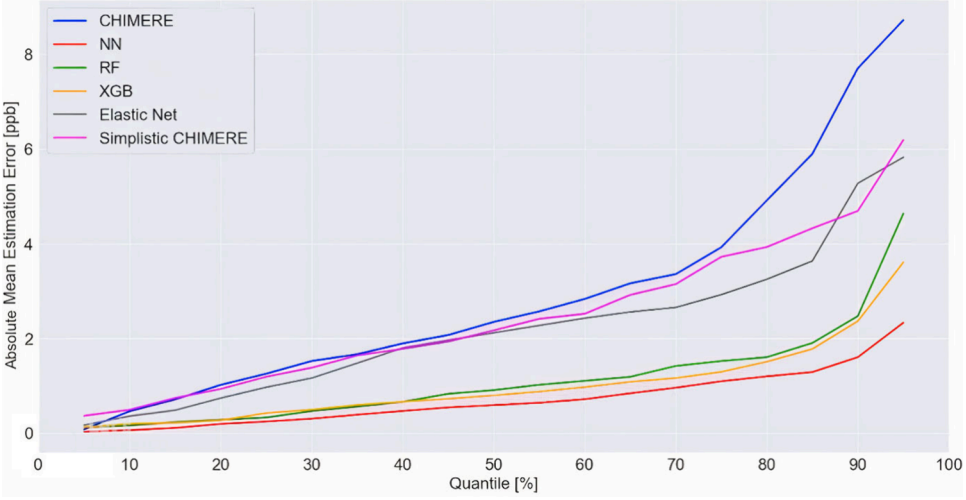


Figure 3. Quantile Analysis of the NO_2 mean estimation error for all AQM stations.

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

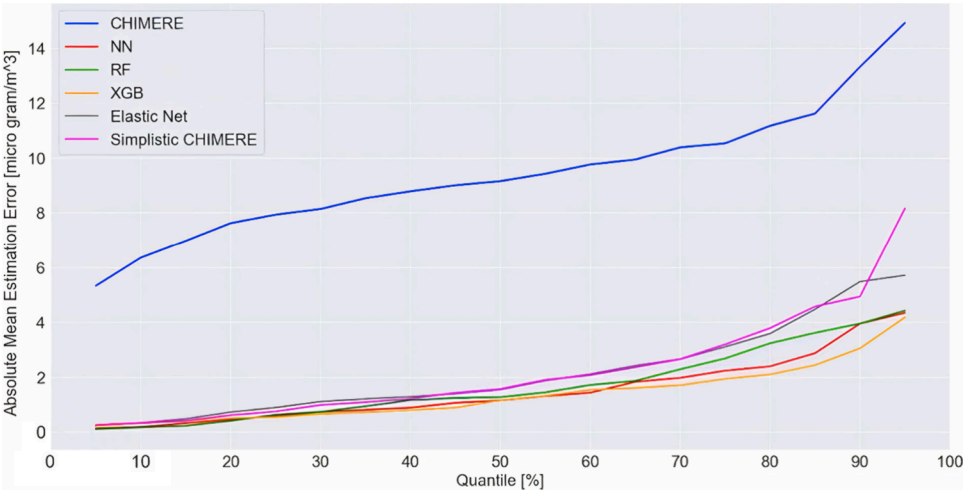


Figure 4. Quantile Analysis of the $PM_{2.5}$ mean estimation error for all AQM stations.

given percentage of AQM stations, CHIMERE exhibits a larger absolute mean estimation error compared to all other models. In practical terms, this suggests that the mean estimation errors for the AQM stations fall within a more constrained range when predicted by the models compared to CHIMERE. Regarding NO_2, the NN-based model outperforms all other models, having the lowest curve throughout all the quantile range. Subsequently, with comparable curves, RF- and XGBoost-based educated models come second. Thereafter, Elastic Net demonstrates a slight advantage over simplistic CHIMERE. Ranked last is plain CHIMERE. Regarding $PM_{2.5}$, the XGBoost educated model has a marginal advantage over the NN-based model, which, in turn, holds a slight edge over the RF-based model. Next, there are both the Elastic Net and the simplistic CHIMERE, with roughly equal curves. As previously mentioned, CHIMERE ranks last.

For all models, Figure 5 and Figure 6 present the distribution of the estimation errors among all test set samples, for NO_2 (9787 samples) and $PM_{2.5}$ (4707 samples), respectively. In addition, for all models, the figures note the mean and standard deviation for the test sets' estimation errors. Upon examination of the figures, for both pollutants, it is evident that the educated models exhibit better performance compared to CHIMERE, and they achieve a reduced standard deviation, a significantly reduced mean estimation error, and fewer outlier occurrences. The Elastic Net succeeds in reducing the mean estimation error substantially; however, its effect on standard deviation reduction is minimal. Hence, overall, the educated models outperform Elastic Net, while CHIMERE ranks last. A discussion concerning simplistic CHIMERE is redundant, as its distribution simply replicates plain CHIMERE's distribution, with a shift along the x-axis. As for the internal ranking of the educated machines, it is task dependent. For NO_2, the NN-based educated model achieves the lowest mean estimation error and standard deviation; hence, it ranks first. Its standard deviation, 6.570 ppb, is just above half that of CHIMERE (12.074 ppb). Thereafter, the XGBoost-based model achieves a standard deviation noticeably lower than the RF-based model; however, its mean estimation error is slightly higher, making this performance comparison nuanced. Regardless, all educated models significantly reduce CHIMERE's 3.698 ppb mean estimation error to close to zero, −0.081 to 0.194 ppb. With respect to $PM_{2.5}$, the XGBoost-based educated machine demonstrated its superiority over the other models, achieving the lowest mean estimation error and standard deviation. Following this, the NN-based model achieves a lower (absolute-wise) mean estimation error than the RF model; however, its standard deviation is higher, making their performances quite comparable (similar to Figure 4). Nonetheless, all educated models substantially reduce CHIMERE's extremely high mean estimation error of 10.088 $\mu g/m^3$ to the range of −0.472 to 0.155 $\mu g/m^3$).

It is important to note that there is an inherent discrepancy between the grid-cell-based nature of the CTM simulations and the localized nature of in situ measurements, particularly for pollutants such as $PM_{2.5}$ and NO_2. This is a recognized challenge in air quality modeling. While CTMs provide valuable insights into large-scale atmospheric processes, they may not always capture the fine-scale variations in pollution levels observed at specific locations. Given the grid-cell-based nature of the CTM, the estimated pollutant levels represent average values for the entire grid cell, presenting an intrinsic limitation of the CTM. As a result, direct comparisons between CTM predictions and in situ measurements are inherently limited and may not accurately reflect localized pollution levels. Nevertheless, as can be observed in Figure 7 and Figure 8, which display the test-set-average absolute estimation error at all AQM sites across the country, our educated machine models demonstrate significant improvements at the local scale of the AQM sites compared to CHIMERE. Spatial error reduction was consistently observed for both pollutants, as evidenced by the predominantly blue-colored markers (indicating lower error values) in the educated machine models' maps compared to CHIMERE's markers.

To evaluate the models' performance in predicting alarming pollution levels, we analyzed the frequency of predictions exceeding the Israeli annual standards within the test set. For NO_2, where the standard is set at 40 $\mu g/m^3$, which is about 21.27 ppb at 25 °C and pressure of 1 atm, the percentages of samples exceeding this threshold were 1.48%, 7.49%, 4.20%, and 5.03% for CHIMERE, NN-, RF-, and XGB-based models, respectively. For $PM_{2.5}$, with a standard of 25 $\mu g/m^3$, the corresponding percentages were 1.10%, 17.12%, 9.98%, and 12.19% for the respective models. The higher percentages achieved by the educated machine models are an expected consequence of CHIMERE's well-established systematic

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

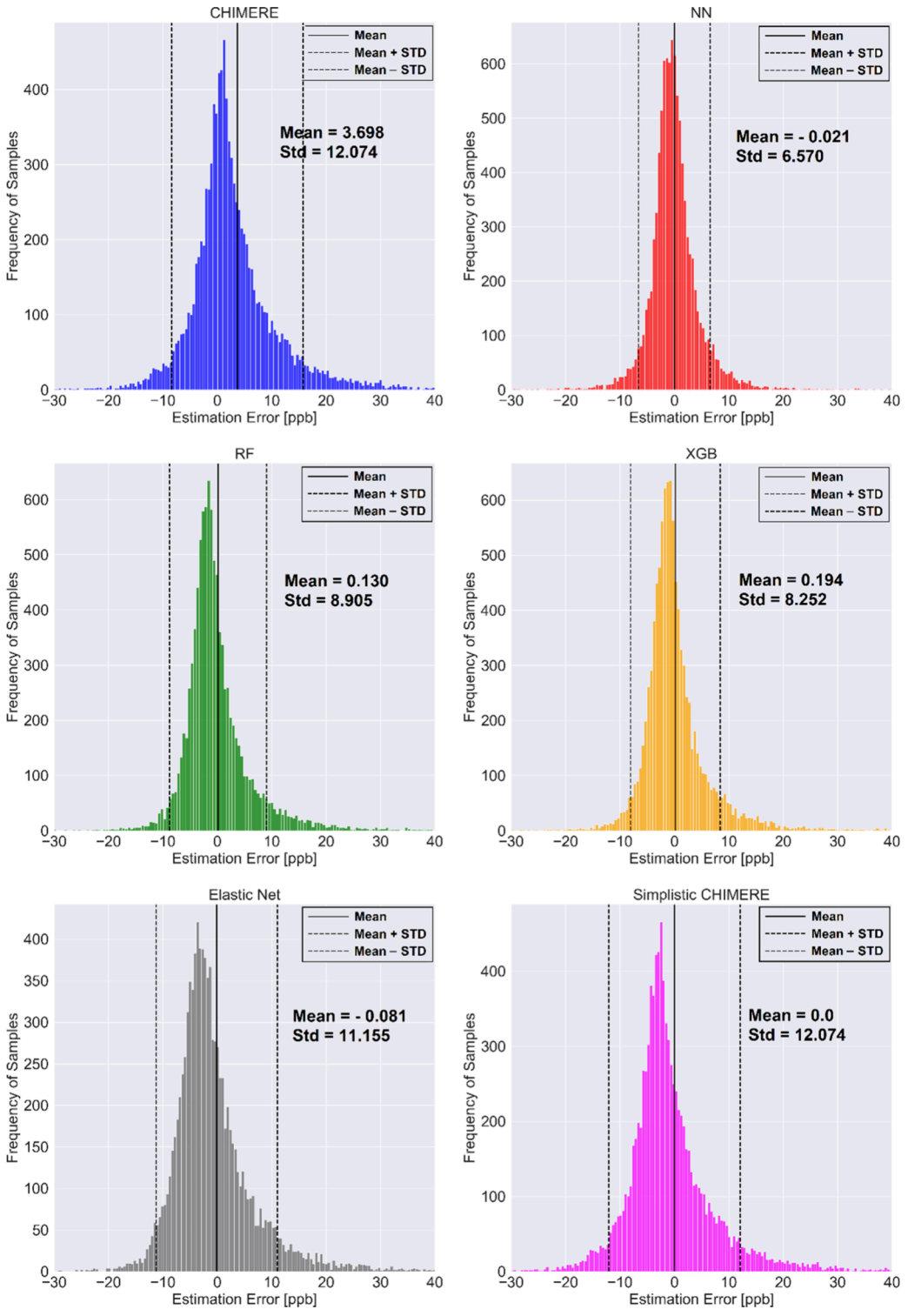


Figure 5. Distribution of all NO_2 estimation errors in the test set.

underestimation bias, which is particularly pronounced for $PM_{2.5}$. These results are particularly significant given the Israeli Ministry of Environmental Protection's mandate to alert the population about impending high-pollution events.

To comprehensively evaluate the predictive capabilities of our educated machine models, dense pollution maps were generated across the country for a randomly selected out-of-sample date (August 21st, 2021), as illustrated in Figure 9 and Figure 10. The maps display predicted NO_2 concentrations for 14:00 and $PM_{2.5}$ concentrations for 9:00, with values outside the country's borders set to zero, serving the objective of providing the population with domestic pollution patterns and values within national jurisdiction. The educated machine models' predictions were constructed by superimposing their predicted CHIMERE errors onto CHIMERE's base estimations.

A notable observation is the distinct spatial patterns exhibited by the different model architectures. For both

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

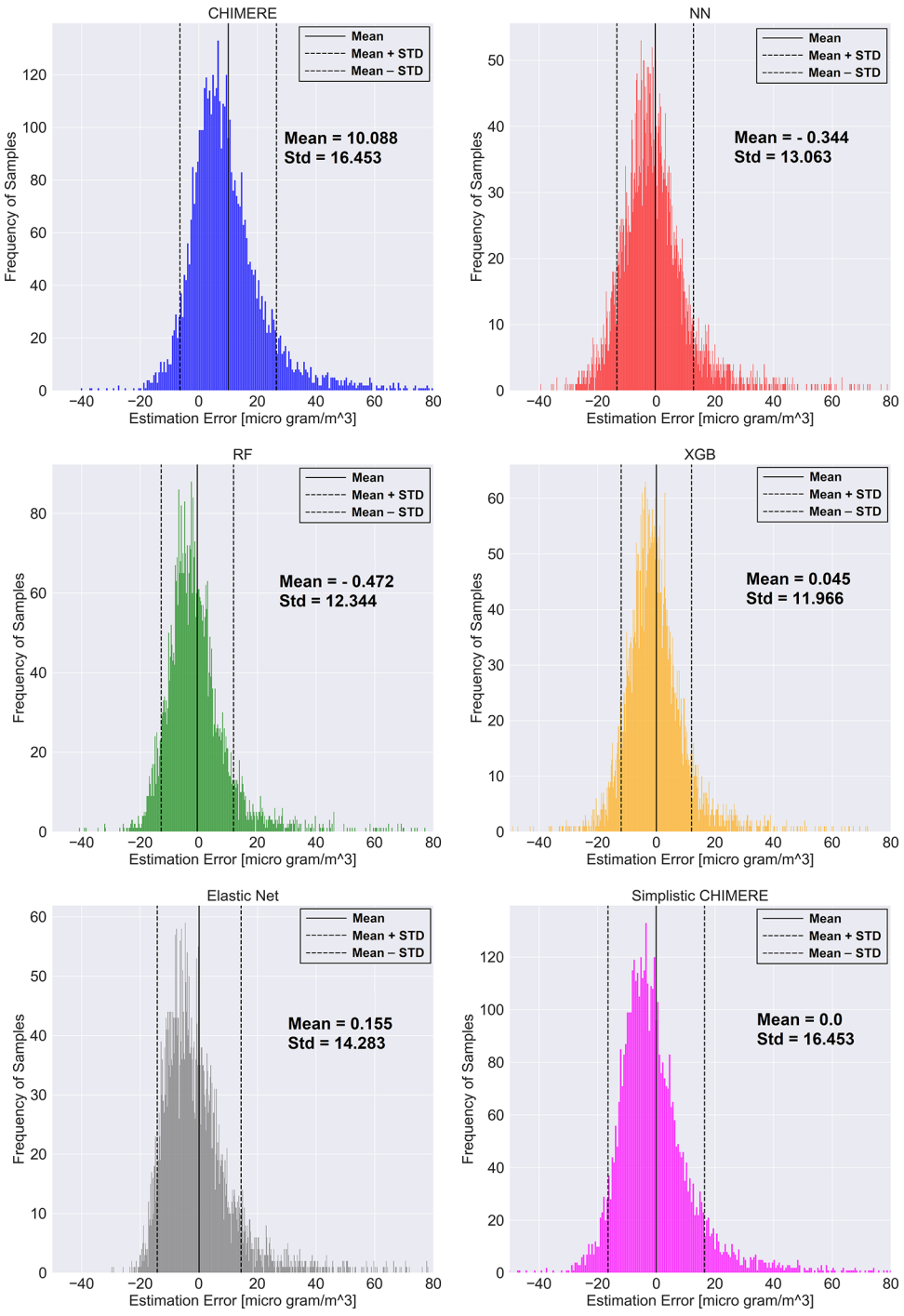


Figure 6. Distribution of all $PM_{2.5}$ estimation errors in the test set.

pollutants, the educated machine models extend the pollution patterns produced by CHIMERE to further regions within the country, where CHIMERE estimates zero levels of pollution. The RF- and XGB-based models demonstrate noticeably similar spatial patterns, consistent with their shared decision-tree ensemble foundation. In contrast, the NN-based model produces more distinctive patterns, particularly for $PM_{2.5}$, showing greater similarity to CHIMERE's spatial patterns, but with enhanced detail. The NN predictions are characterized by smoother transitions between regions, whereas the RF-based model, in particular, exhibits more erratic spatial variability.

All educated machine models consistently predict higher pollutant concentrations compared to CHIMERE across the domain, substantiating their ability to mitigate CHIMERE's documented systematic underestimation bias.

All in all, while CHIMERE relies on robust mathematical, chemical, and physical frameworks, there are inherent limitations in capturing the full complexity of atmospheric phenomena. CTMs struggle with representing intricate and synergistic processes in the atmosphere that are not yet fully understood or easily modeled. Additionally, CHIMERE's performance is constrained by several factors, including uncertainties in emission inventories, inaccuracies in meteoro-

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

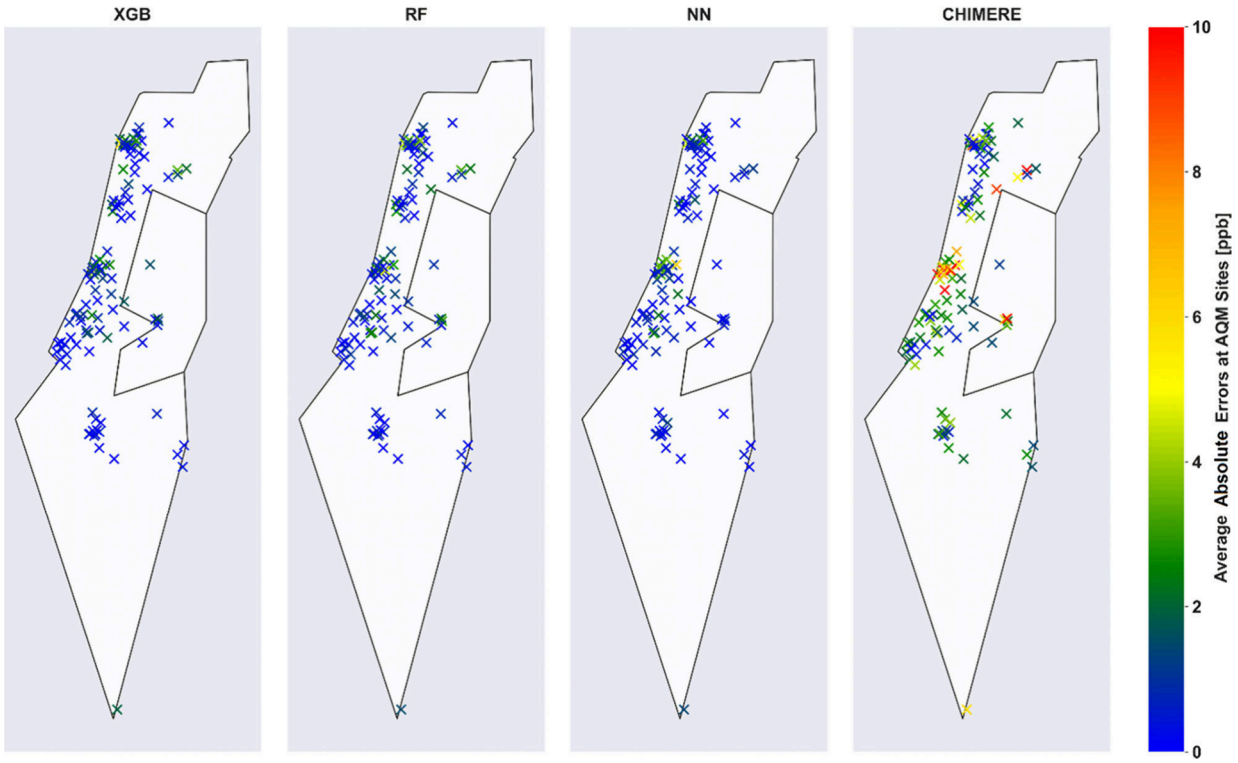


Figure 7. Average NO_2 absolute estimation errors across the test set samples at all AQM sites.

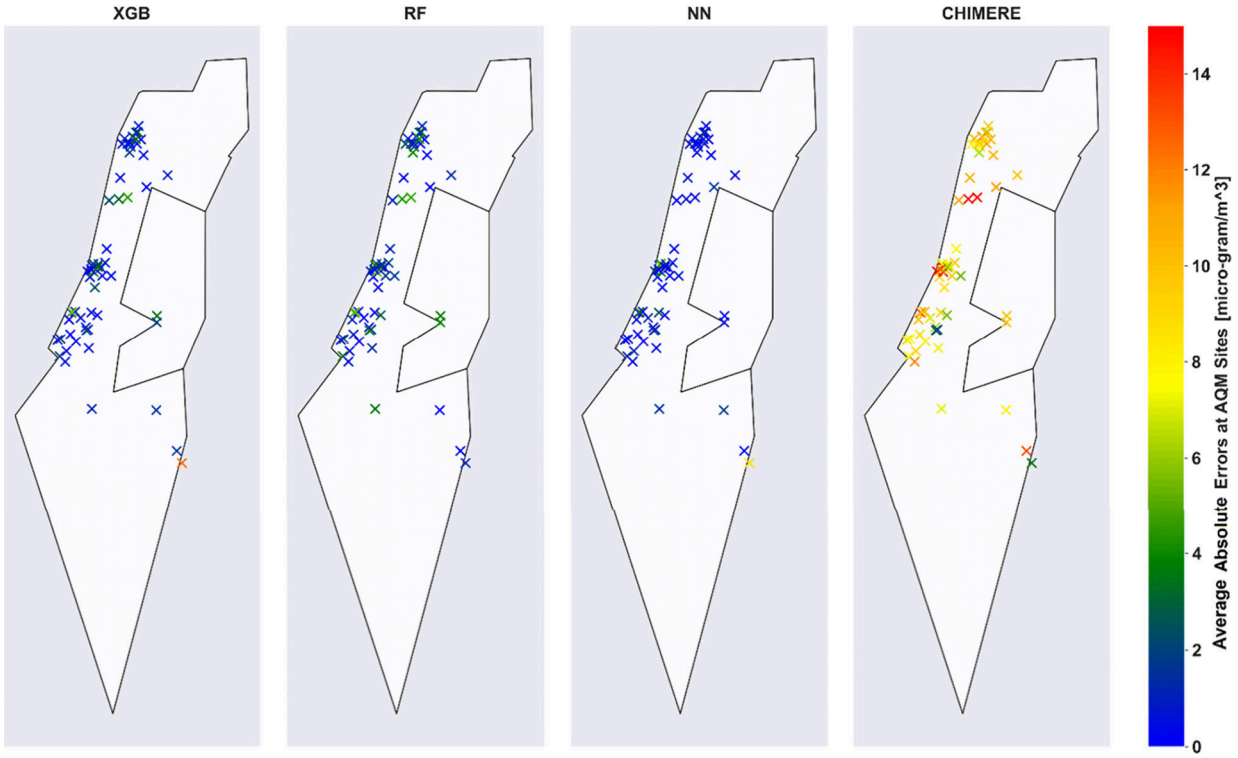


Figure 8. Average $PM_{2.5}$ absolute estimation errors across the test set samples at all AQM sites.

logical inputs, and limitations in representing long-range transport and subgrid processes. Further challenges arise from computational limitations and assumptions made to simplify these complex systems, such as numerical approximations, which inevitably impair the accuracy of the predictions. Representativeness issues with AQM site data may also contribute to the discrepancies. These factors collectively exacerbate CHIMERE's limited performance. Addressing these limitations requires integration of advanced methodologies, such as AI techniques, to complement and enhance the predictive capabilities of CTMs. Based on the

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

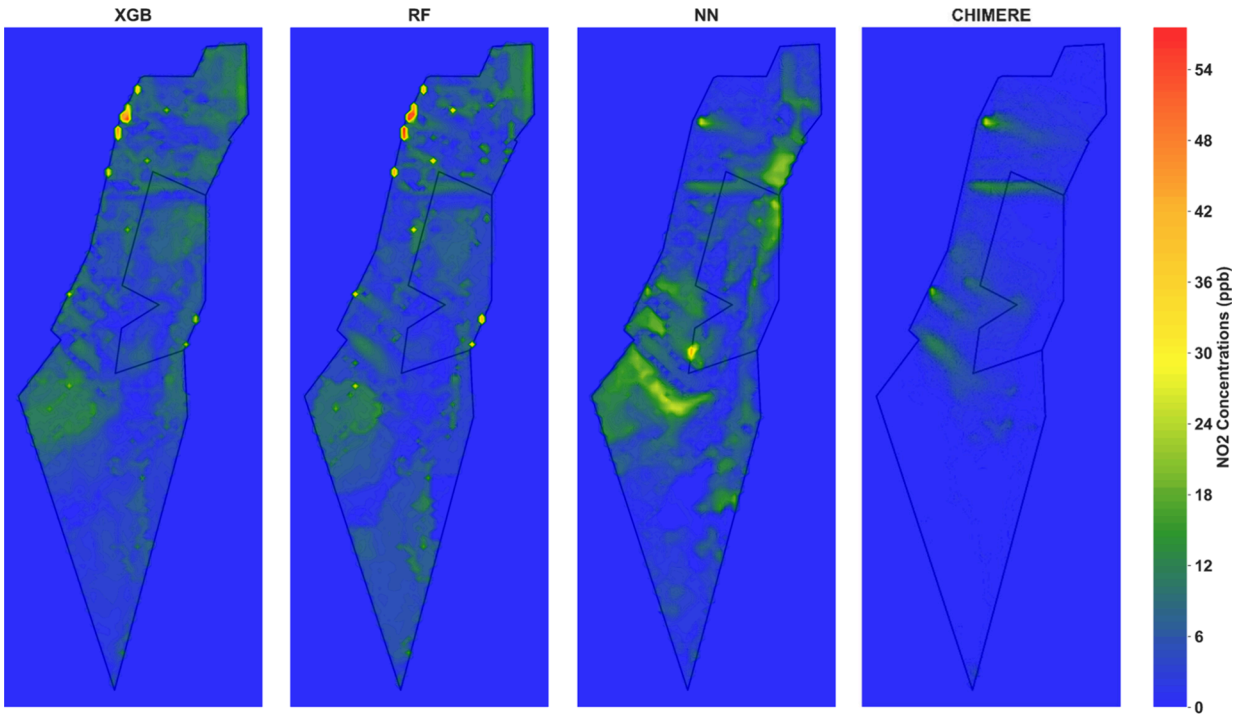


Figure 9. Instantaneous dense NO_2 pollution maps generated across the entire grid of Israel.

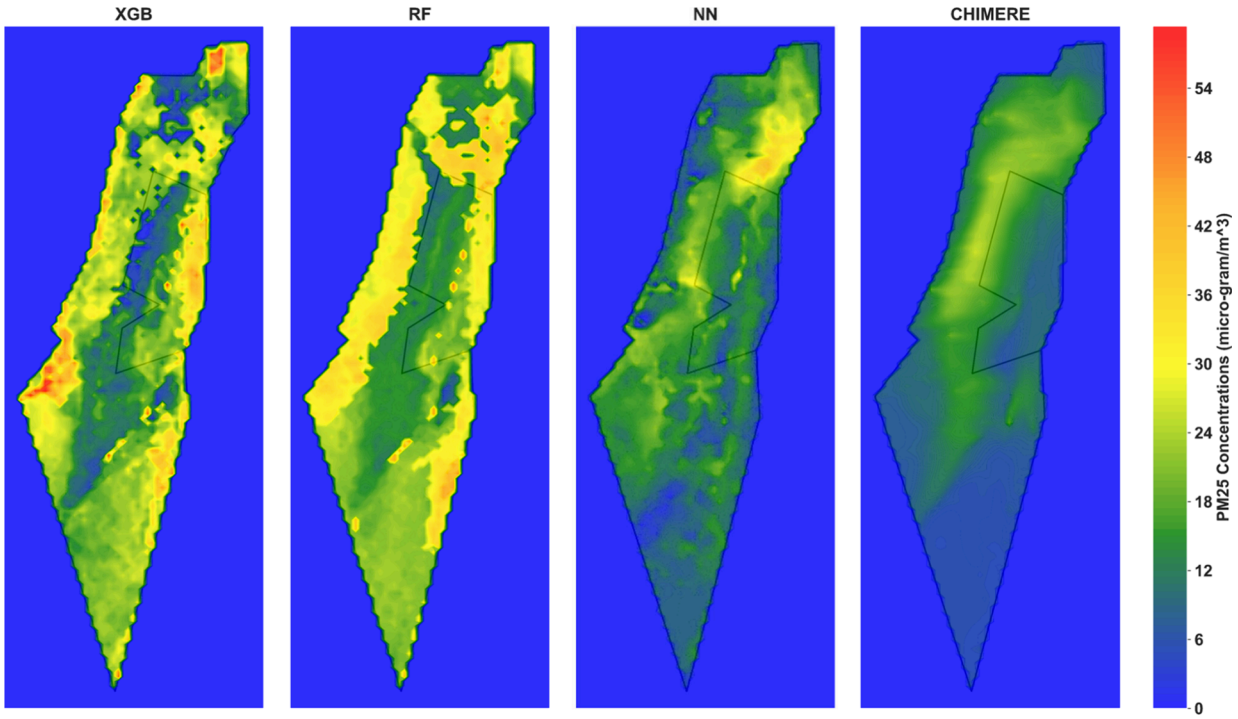


Figure 10. Instantaneous dense $PM_{2.5}$ pollution maps across Israel.

results, the machine education approach has proved a strong candidate for air pollution modeling.

4. CONCLUSIONS

The study introduces the educated machine approach, which integrates a chemistry transport model with Machine and Deep Learning (M&DL) algorithms, for generating dense air pollution maps in Israel. Specifically, the educated machine here consists of the CHIMERE chemistry transport model and either Neural Networks, Random Forest, or XGBoost as the M&DL component in the machine. This method improves data efficiency and accuracy over CHIMERE alone. Neural Networks performed best, with the greatest reductions in estimation errors (MAE, RMSE) and the most plausible dense air pollution maps generated. Using a three-year data set, including meteorological geographical and topographical variables, CHIMERE's estimates, and AQM station measurements, the models addressed biases and improved predictions,

https://doi.org/10.1021/acsestair.4c00178
ACS EST Air 2025, 2, 1411−1425

especially for high-pollution events. Overall, machine learning significantly enhances pollutant estimation accuracy.

Our study presents and affirms the educated machine approach potential, combining CTMs with machine and deep learning methodologies, offering significant benefits to global efforts to mitigate the impact of air pollution on human health and mortality.

Finally, given the fundamental similarities in CTM frameworks and their common challenges, the results achieved in this study hold promise to yield comparable benefits across different modeling systems across a diverse range of geographical regions. Nevertheless, their broader applicability warrants further investigation. To establish the generalizability of these techniques, further efforts should be invested in exploring diverse geographic regions and assessing performance across various chemistry transport models and pollutants. Additionally, the integration of low-cost sensors could extend spatial coverage, enabling validation in unmonitored areas and providing valuable data to further refine model predictions.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsestair.4c00178.

> Detailed comparison of the models' performance when trained using random versus temporal data splits, highlighting the advantage of the former; the application and outcomes of feature selection techniques, which yielded no improvement and led to the inclusion of all features in the final analysis; complete list of features used in the data sets for NO_2, and $PM_{2.5}$ (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Barak Fishbain − *Technion Environmental Informatics Lab (TechEL), Faculty of Civil and Environmental Engineering, The Technion − Israel Institute of Technology, Haifa 320003, Israel;* orcid.org/0000-0003-4211-7445; Email: fishbain@technion.ac.il

Authors

Avitay Geltman − *Technion Environmental Informatics Lab (TechEL), Faculty of Civil and Environmental Engineering, The Technion − Israel Institute of Technology, Haifa 320003, Israel;* orcid.org/0009-0003-8667-0426

Ilan Levy − *Department of Air Quality and Climate Change, The Israeli Ministry of Environmental Protection, Jerusalem 9195024, Israel*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsestair.4c00178

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was partially funded by the Israeli Ministry of Science and Technology, the Israeli Ministry of Environmental Protection and the Israeli Water Authority (grant number 1820020069).

■ REFERENCES

(1) Kampa, M.; Castanas, E. Human Health Effects of Air Pollution. *Environ. Pollut.* **2008**, *151* (2), 362−367.

(2) Apte, J. S.; Marshall, J. D.; Cohen, A. J.; Brauer, M. Addressing Global Mortality from Ambient PM2.5. *Environ. Sci. Technol.* **2015**, *49* (13), 8057−8066.

(3) Araujo, L. N.; Belotti, J. T.; Alves, T. A.; Tadano, Y. de S.; Siqueira, H. Ensemble Method Based on Artificial Neural Networks to Estimate Air Pollution Health Risks. *Environ. Model. Softw.* **2020**, *123*, No. 104567.

(4) Nebenzal, A.; Fishbain, B. *Long-Term Forecasting of Nitrogen Dioxide Ambient Levels in Metropolitan Areas Using the Discrete-Time Markov Model* **2018**, *107*, 175−185.

(5) U.S. Environmental Protection Agency. *Integrated Science Assessment (ISA) for Oxides of Nitrogen − Health Criteria*; U.S. Environmental Protection Agency: Washington, DC, 2016.

(6) Bi, J.; Wildani, A.; Chang, H. H.; Liu, Y. Incorporating Low-Cost Sensor Measurements into High-Resolution PM2.5 Modeling at a Large Spatial Scale. *Environ. Sci. Technol.* **2020**, *54* (4), 2152−2162.

(7) Moltchanov, S.; Levy, I.; Etzion, Y.; Lerner, U.; Broday, D. M. D. M.; Fishbain, B. On the Feasibility of Measuring Urban Air Pollution by Wireless Distributed Sensor Networks. *Sci. Total Environ.* **2015**, *502*, 537−547.

(8) Huijnen, V.; Williams, J.; van Weele, M.; van Noije, T.; Krol, M.; Dentener, F.; Segers, A.; Houweling, S.; Peters, W.; de Laat, J.; Boersma, F.; Bergamaschi, P.; van Velthoven, P.; Le Sager, P.; Eskes, H.; Alkemade, F.; Scheele, R.; Nédélec, P.; Pätz, H.-W. The Global Chemistry Transport Model TM5: Description and Evaluation of the Tropospheric Chemistry Version 3.0. *Geosci. Model Dev.* **2010**, *3* (2), 445−473.

(9) Silva, S. J.; Evans, M. Artificial Intelligence and Machine Learning in Atmospheric Chemistry. *ACS EST Air* **2024**, *1* (5), 330−331.

(10) Cabaneros, S. M.; Calautit, J. K.; Hughes, B. R. A Review of Artificial Neural Network Models for Ambient Air Pollution Prediction. *Environ. Model. Softw.* **2019**, *119*, 285−304.

(11) Tiwary, A.; Williams, I. *Air Pollution | Measurement, Modelling and Mitigation*, 4th ed.; CRC Press, 2018.

(12) Huang, J.; McQueen, J.; Wilczak, J.; Djalalova, I.; Stajner, I.; Shafran, P.; Allured, D.; Lee, P.; Pan, L.; Tong, D.; Huang, H.-C.; DiMego, G.; Upadhayay, S.; Delle Monache, L. Improving NOAA NAQFC PM2.5 Predictions with a Bias Correction Approach. *Weather Forecast* **2017**, *32* (2), 407−421.

(13) Byun, D.; Schere, K. L. Review of the Governing Equations, Computational Algorithms, and Other Components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System. *Appl. Mech. Rev.* **2006**, *59* (2), 51−77.

(14) Hamill, T. M.; Whitaker, J. S. Probabilistic Quantitative Precipitation Forecasts Based on Reforecast Analogs. *Mon. Weather Rev.* **2006**, *134* (11), 3209.

(15) Djalalova, I.; Delle Monache, L.; Wilczak, J. PM2.5 Analog Forecast and Kalman Filter Post-Processing for the Community Multiscale Air Quality (CMAQ) Model. *Atmos. Environ.* **2015**, *108*, 76−87.

(16) Ordieres, J. B.; Vergara, E. P.; Capuz, R. S.; Salazar, R. E. Neural Network Prediction Model for Fine Particulate Matter (PM2.5) on the US−Mexico Border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environ. Model. Softw.* **2005**, *20* (5), 547−559.

(17) Hou, L.; Dai, Q.; Song, C.; Liu, B.; Guo, F.; Dai, T.; Li, L.; Liu, B.; Bi, X.; Zhang, Y.; Feng, Y. Revealing Drivers of Haze Pollution by Explainable Machine Learning. *Environ. Sci. Technol. Lett.* **2022**, *9* (2), 112−119.

(18) Lightstone, S. D.; Moshary, F.; Gross, B. Comparing CMAQ Forecasts with a Neural Network Forecast Model for PM2.5 in New York. *Atmosphere* **2017**, *8* (9), 161.

(19) Ajdour, A.; Adnane, A.; Ydir, B.; Ben hmamou, D.; Khomsi, K.; Amghar, H.; Chelhaoui, Y.; Chaoufi, J.; Leghrib, R. A New Hybrid Models Based on the Neural Network and Discrete Wavelet

Transform to Identify the CHIMERE Model Limitation. *Environ. Sci. Pollut. Res.* **2023**, *30* (5), 13141−13161.
(20) Mhawish, A.; Banerjee, T.; Sorek-Hamer, M.; Bilal, M.; Lyapustin, A. I.; Chatfield, R.; Broday, D. M. Estimation of High-Resolution PM2.5 over the Indo-Gangetic Plain by Fusion of Satellite Data, Meteorology, and Land Use Variables. *Environ. Sci. Technol.* **2020**, *54* (13), 7891−7900.
(21) Menut, L.; Bessagnet, B.; Khvorostyanov, D.; Beekmann, M.; Blond, N.; Colette, A.; Coll, I.; Curci, G.; Foret, G.; Hodzic, A.; Mailler, S.; Meleux, F.; Monge, J.-L.; Pison, I.; Siour, G.; Turquety, S.; Valari, M.; Vautard, R.; Vivanco, M. G. CHIMERE 2013: A Model for Regional Atmospheric Composition Modelling. *Geosci. Model Dev.* **2013**, *6* (4), 981−1028.
(22) Ferreyra, M. F. G.; Curci, G.; Lanfri, M. First Implementation of the WRF-CHIMERE-EDGAR Modeling System Over Argentina. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9* (12), 5304−5314.
(23) Petetin, H.; Sciare, J.; Bressi, M.; Gros, V.; Rosso, A.; Sanchez, O.; Sarda-Estève, R.; Petit, J.-E.; Beekmann, M. Assessing the Ammonium Nitrate Formation Regime in the Paris Megacity and Its Representation in the CHIMERE Model. *Atmospheric Chem. Phys.* **2016**, *16* (16), 10419−10440.
(24) Terrenoire, E.; Bessagnet, B.; Rouïl, L.; Tognet, F.; Pirovano, G.; Létinois, L.; Beauchamp, M.; Colette, A.; Thunis, P.; Amann, M.; Menut, L. High-Resolution Air Quality Simulation over Europe with the Chemistry Transport Model CHIMERE. *Geosci. Model Dev.* **2015**, *8* (1), 21−42.
(25) Bessagnet, B.; Couvidat, F.; Lemaire, V. A Statistical Physics Approach to Perform Fast Highly-Resolved Air Quality Simulations − A New Step towards the Meta-Modelling of Chemistry Transport Models. *Environ. Model. Softw.* **2019**, *116*, 100−109.
(26) Kendler, S.; Mano, Z.; Aharoni, R.; Raich, R.; Fishbain, B. Hyperspectral Imaging for Chemicals Identification: A Human-Inspired Machine Learning Approach. *Sci. Rep. 2022 121* **2022**, *12* (1), 1−10.
(27) Brownlee, J. *Understand the Impact of Learning Rate on Neural Network Performance*. https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/ (accessed 2024-07-12).
(28) Bhavsar, M. A.; Munjani, J. H.; Joshi, M. Target Tracking in WSN Using Dynamic Neural Network Techniques. In *Smart and Innovative Trends in Next Generation Computing Technologies*; Bhattacharyya, P., Sastry, H. G., Marriboyina, V., Sharma, R., Eds.; Springer: Singapore, 2018; pp 771−789, DOI: 10.1007/978-981-10-8660-1_58.
(29) Larremore, D. B.; Shew, W. L.; Restrepo, J. G. Predicting Criticality and Dynamic Range in Complex Networks: Effects of Topology. *Phys. Rev. Lett.* **2011**, *106* (5), No. 058101.
(30) Levy, I.; Karakis, I.; Berman, T.; Amitay, M.; Barnett-Itzhaki, Z. A Hybrid Model for Evaluating Exposure of the General Population in Israel to Air Pollutants. *Environ. Monit. Assess.* **2020**, *192* (1), 4.
(31) Le, T. A.; Baydin, A. G.; Zinkov, R.; Wood, F. Using Synthetic Data to Train Neural Networks Is Model-Based Reasoning. *2017 International Joint Conference on Neural Networks (IJCNN)* **2017**, 3514−3521.
(32) Cuomo, S.; Di Cola, V. S.; Giampaolo, F.; Rozza, G.; Raissi, M.; Piccialli, F. Scientific Machine Learning Through Physics−Informed Neural Networks: Where We Are and What's Next. *J. Sci. Comput.* **2022**, *92* (3), 88.
(33) Feldman, A.; Kendler, S.; Marshall, J.; Kushwaha, M.; Sreekanth, V.; Upadhya, A. R.; Agrawal, P.; Fishbain, B. Urban Air-Quality Estimation Using Visual Cues and a Deep Convolutional Neural Network in Bengaluru (Bangalore), India. *Environ. Sci. Technol.* **2024**, *58* (1), 480−487.
(34) Farchi, A.; Chrust, M.; Bocquet, M.; Laloyaux, P.; Bonavita, M. Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework. *J. Adv. Model. Earth Syst.* **2023**, *15* (9), No. e2022MS003474.
(35) Zhu, J.; Hu, S.; Arcucci, R.; Xu, C.; Zhu, J.; Guo, Y. Model Error Correction in Data Assimilation by Integrating Neural Networks. *Big Data Min. Anal.* **2019**, *2* (2), 83−91.
(36) Arzani, A.; Dawson, S. T. M. Data-Driven Cardiovascular Flow Modelling: Examples and Opportunities. *J. R. Soc. Interface* **2021**, *18* (175), 2021.
(37) Abiodun, O. I.; Jantan, A.; Omolara, A. E.; Dada, K. V.; Mohamed, N. A.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon* **2018**, *4* (11), No. e00938.
(38) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification And Regression Trees*; Routledge, 1984; DOI: 10.1201/9781315139470.
(39) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5−32.
(40) Geisser, S. The Predictive Sample Reuse Method with Applications. *J. Am. Stat. Assoc.* **1975**, *70* (350), 320−328.
(41) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; KDD '16*; Association for Computing Machinery: New York, NY, USA, 2016; pp 785−794, DOI: 10.1145/2939672.2939785.