



Optimizing BenMAP health impact assessment with meteorological factor driven machine learning models

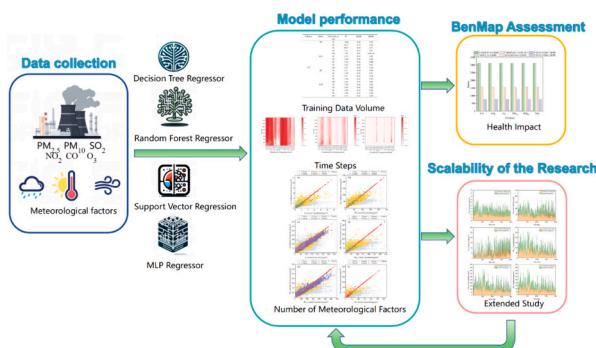
Juncheng Wu, Qili Dai, Shaojie Song*

State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China
CMA-NKU Cooperative Laboratory for Atmospheric Environment-Health Research, Tianjin 300350, China

HIGHLIGHTS

- Utilize BenMAP and ML to improve air pollution assessment.
- Find optimal prediction steps and key meteorological factors.
- High accuracy in CO and O₃ prediction with Decision Tree.
- Demonstrated the generalizability of the method for application to different cities.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Meng Gao

Keywords:
Air pollution
BenMap
Machine learning
Meteorological factors
Prediction accuracy
Health impact assessment

ABSTRACT

This study aims to address accuracy challenges in assessing air pollution health impacts using Environmental Benefits Mapping and Analysis Program (BenMap), caused by limited meteorological factor data and missing pollutant data. By employing data increment strategies and multiple machine learning models, this research explores the effects of data volume, time steps, and meteorological factors on model prediction performance using several years of data from Tianjin City as an example. The findings indicate that increasing training data volume enhances the performance of Random Forest Regressor (RF) and Decision Tree Regressor (DT) models, especially for predicting CO, NO₂, and PM_{2.5}. The optimal prediction time step varies by pollutant, with the DT model achieving the highest R² value (0.99) for CO and O₃. Combining multiple meteorological factors, such as atmospheric pressure, relative humidity, and dew point temperature, significantly improves model accuracy. When using three meteorological factors, the model achieves an R² of 0.99 for predicting CO, NO₂, PM₁₀, PM_{2.5}, and SO₂. Health impact assessments using BenMap demonstrated that the predicted all-cause mortality and specific disease mortalities were highly consistent with actual values, confirming the model's accuracy in assessing health impacts from air pollution. For instance, the predicted and actual all-cause mortality for PM_{2.5} were both 3120; for cardiovascular disease, both were 1560; and for respiratory disease, both were 780. To validate its generalizability, this method was applied to Chengdu, China, using several years of data for training and prediction of PM_{2.5}, CO, NO₂, O₃, PM₁₀, and SO₂, incorporating atmospheric pressure, relative humidity, and

* Corresponding author.

E-mail address: songs@nankai.edu.cn (S. Song).



dew point temperature. The model maintained excellent performance, confirming its broad applicability. Overall, we conclude that the machine learning and BenMap-based methods show high accuracy and reliability in predicting air pollutant concentrations and health impacts, providing a valuable reference for air pollution assessment.

1. Introduction

With the rapid development of industrialization and urbanization, air pollution has become a global environmental issue, posing serious threats to human health and ecosystems (Ducruet et al., 2024). In China, particularly in urban areas with concentrated industry and dense populations, the issue of air pollution is more pronounced. Concentrations of pollutants such as PM_{2.5}, SO₂, and CO often exceed health standards, increasing the likelihood of respiratory diseases, cardiovascular conditions, and other health risks (Ai et al., 2024; Kalisa et al., 2023; Zhong et al., 2023). Meteorological factors, such as temperature, humidity, wind direction and speed, and precipitation, significantly influence the dispersion and transformation of pollutants, thereby further determining the spatiotemporal distribution characteristics of air pollution (de Souza Fernandes Duarte et al., 2024; Li et al., 2024; Ni et al., 2023). Thus, a deep understanding of the current state of air pollution, the relationship between meteorological factors and pollutants, and the assessment of their potential impact on public health is crucial for the development of effective environmental policies and health protection measures.

BenMap (Environmental Benefits Mapping and Analysis Program), as a widely utilized health impact assessment tool, offers robust technical support for evaluating the effects of air pollution on health by integrating information from geographic information systems (GIS), environmental monitoring data, demographic statistics, and disease incidence rates (Chen et al., 2017; Su et al., 2024; Voorhees et al., 2014). However, a primary challenge when using BenMap for health risk assessment is the inaccuracy of pollutant data. Due to the limitations of monitoring station distribution and the constraints on data collection frequency, traditional methods may fail to comprehensively capture the detailed temporal and spatial variations of pollutants, leading to uncertainties in the assessment outcomes (Goudarzi et al., 2021; Nguyen et al., 2023). Furthermore, the accuracy of predicting future health conditions is significantly impacted by the lack of future pollutant emission and concentration data (Aguiar-Gil et al., 2020; Han et al., 2021; Parthum et al., 2017). The absence of such data makes it difficult for assessment models to adapt to potential future environmental changes, limiting our ability to predict and plan for future health risks. Therefore, to enhance the accuracy and forward-looking nature of assessments, there is an urgent need to explore methods that can overcome these limitations. The introduction of machine learning technology offers new opportunities to address these challenges. By leveraging historical data and advanced algorithms, machine learning models are expected to improve the predictive accuracy of pollutant concentration changes and provide a more reliable basis for assessing future health risks.

Machine learning models, particularly those based on meteorological factors, can process and analyze large-scale complex datasets, thereby offering more precise predictions of pollutant concentrations (Chen et al., 2024; Sadiq et al., 2023; Wang et al., 2023). By training models to recognize the complex relationships between pollutants and meteorological factors, we can predict the distribution of pollutants under various meteorological conditions (Verma et al., 2024; Zhang et al., 2022). This approach not only enhances the accuracy of pollutant concentration forecasts but also supplies more reliable input data for health impact assessment tools like BenMap.

This study aims to address accuracy challenges in assessing air pollution health impacts using BenMap, caused by limited meteorological factor data and missing pollutant data. By employing data increment

strategies and multiple machine learning models, this research explores the effects of data volume, time steps, and meteorological factors on model prediction performance using several years of data from Tianjin City, China, as an example. Tianjin was chosen due to the availability and completeness of historical meteorological and air quality data from the Tianjin Municipal Environmental Protection Bureau and the Meteorological Bureau, which allowed for comprehensive training and optimization of our machine learning models. Here we collected air pollution and meteorological data from Tianjin for the period of 2013–2019 and utilized four machine learning models: DT, RF, Support Vector Regression (SVR), and Multilayer Perceptron (MLP) to analyze the relationship between meteorological factors and pollutants (Fig. S1). Special emphasis was placed on cross-validation to ensure the generalizability and predictive accuracy of the models. Through cross-validation, we were able to effectively assess the performance of different models on independent datasets, thereby selecting the best model to forecast potential future changes in pollutant concentrations. By integrating the predictions from these rigorously validated machine learning models with BenMap, we hope to more accurately assess the impact of air pollution on public health and provide policymakers with a more scientific basis for decision-making. Moreover, the outcomes of this study will offer new perspectives and methodologies for research and practice in global air pollution health risk assessment, especially in addressing the challenges of complex data and dynamic changes.

2. Material and methods

2.1. Data sources and preprocessing strategies

The data utilized in this study were derived from the official databases of the Tianjin Municipal Environmental Protection Bureau and the Meteorological Bureau, encompassing a period from 2013 to 2019, with the impact of the COVID-19 pandemic excluded. The dataset included a variety of key meteorological parameters, such as atmospheric temperature, atmospheric pressure, relative humidity, wind speed, total cloud cover, dew point temperature, and precipitation amount, as well as air quality indicators including PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃. Machine learning model workflow (see Fig. 1) was applied in this study. Data from the years 2013 to 2018 were employed as the training set, while data from 2019 were designated as the validation set. Prior to the construction of the model, a series of preprocessing operations were conducted to ensure the quality of the data and the accuracy of the analysis.

The preprocessing procedure initially involved data cleaning using the Pandas library, which excluded records with incomplete information or outliers. For missing values within the dataset, advanced imputation strategies such as K-Nearest Neighbors Imputation or Iterative Regression Imputation were employed to maintain the integrity of the dataset. Furthermore, considering the discrepancies in units and value ranges among various features, this study applied the Z-score normalization method, utilizing the StandardScaler class to standardize the data. This process ensured that each feature's distribution had a mean of zero and a standard deviation of one, thereby eliminating the impact of units and laying a solid foundation for subsequent model training.

2.2. Model construction and evaluation metrics

This study utilized four established machine learning models to build an air quality forecasting framework. These models include: (1) DT,

which recursively divides the dataset into smaller subsets to find the best feature split points for minimizing mean squared error; (2) RF, an ensemble learning method that enhances prediction performance by averaging or voting the results of multiple decision trees, with randomness introduced in both data sampling and feature selection to improve generalization; (3) SVR, a kernel-based method that seeks to

maximize margin in feature space and can handle non-linear problems by mapping data into higher dimensions using kernel functions like the Radial Basis Function (RBF); and (4) MLP, a feedforward neural network model with one or more hidden layers and an output layer, where neurons apply activation functions like ReLU or Sigmoid to transform inputs non-linearly, and the model adjusts weights and biases through

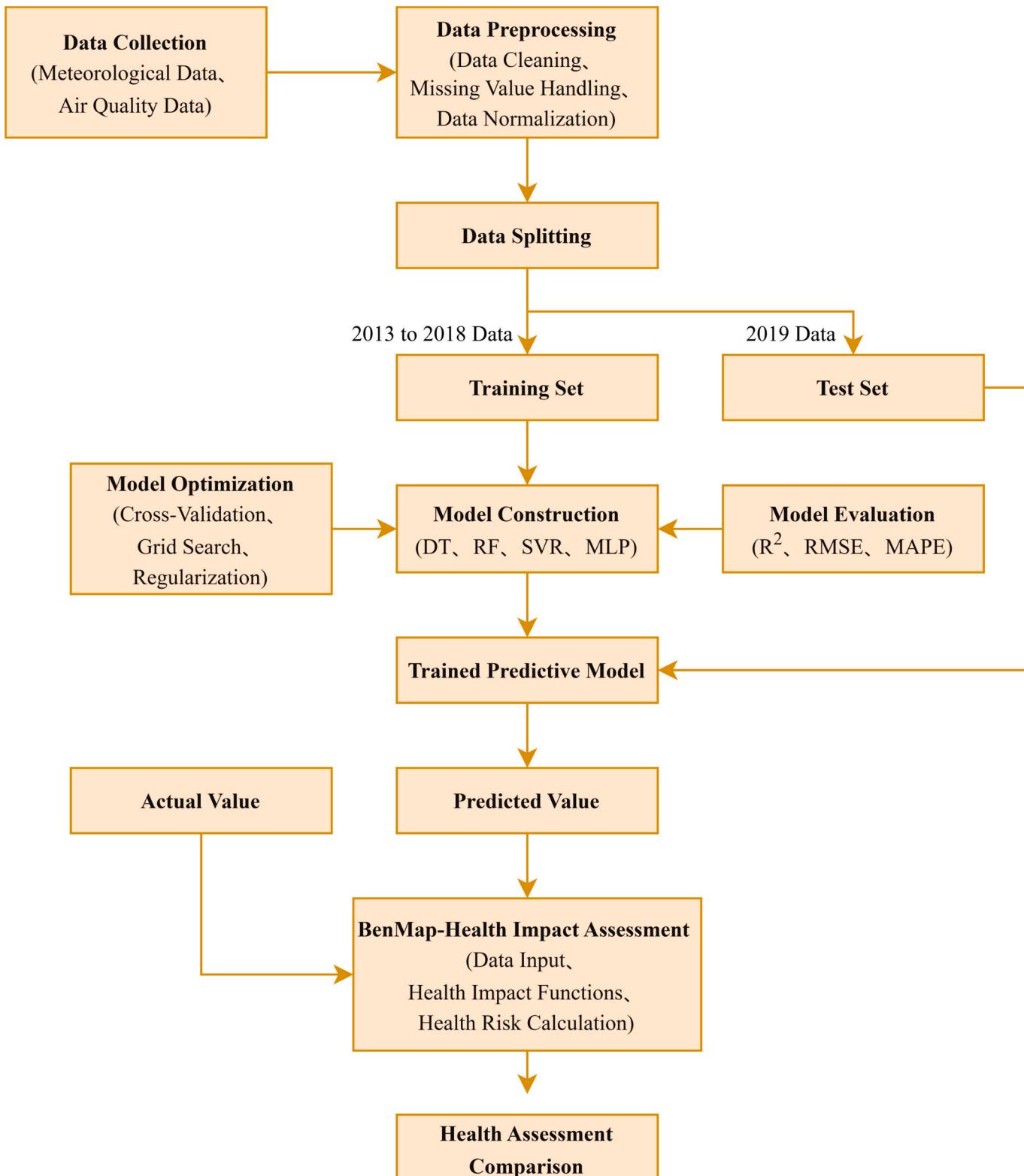


Fig. 1. Machine learning model workflow for BenMAP health impact assessment.

backpropagation and gradient descent to minimize prediction errors.

Model evaluation primarily relies on three key metrics: the Coefficient of Determination (R^2), the Root Mean Square Error (RMSE), and the Mean Absolute Percentage Error (MAPE). The R^2 metric measures the model's ability to explain the variability in the data, and it is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

Here, y_i represents the actual observed values, \hat{y}_i are the model's predicted values, and \bar{y} is the mean of the observed values. The closer the R^2 value is to 1, the better the model's predictive performance.

The RMSE metric measures the differences between the model's predicted values and the actual observed values, and it is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

The smaller the RMSE value, the higher the model's predictive accuracy.

Furthermore, the MAPE provides an intuitive percentage representation of the model's prediction error and is calculated as:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

The MAPE indicator offers a straightforward percentage measure of the average prediction error across different data points, which is useful for evaluating the model's accuracy in its predictions.

For the MLP model, in addition to the aforementioned metrics, the model's convergence is also closely monitored. By tracking the changes in the loss function, it is possible to determine whether the model has reached a state of convergence during the iterative process. The loss function typically employs Mean Squared Error (MSE) for regression problems or Cross-Entropy for classification problems, depending on the nature of the task at hand (Afzal et al., 2023; Alomari and Andó, 2024). For MLP, the Mean Squared Error (MSE) is commonly used as the loss function, which is calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (4)$$

During the model training process, it is expected that the value of the loss function will gradually decrease as the number of iterations increases, eventually stabilizing around a small value. This indicates that the model has converged. If the value of the loss function continues to drop significantly or fluctuates largely, it may be necessary to adjust the learning rate, increase the number of iterations, or employ other regularization techniques to improve the model's convergence.

2.3. Model optimization and cross-validation strategies

In the pursuit of optimizing model performance, this study employs a systematic approach combining Cross-Validation with Grid Search. Cross-Validation is a statistical technique used to evaluate a model's generalization capability, which refers to the model's performance on unseen data (Xu et al., 2024). GridSearchCV is a powerful parameter optimization tool that systematically explores the user-defined parameter grid by conducting cross-validation for each set of parameters, thereby identifying the optimal parameter configuration (Devasahayam and Albijanic, 2024). In this study, a 5-fold cross-validation strategy was employed. The dataset was divided into five equal subsets. During each iteration of GridSearchCV, the model was trained on four subsets and validated on the remaining one. This process was repeated five times,

with a different subset used for validation each time, ensuring a comprehensive assessment of model performance and reducing variance due to data partitioning. GridSearchCV aimed to minimize the loss function across the validation process.

To prevent overfitting, regularization techniques were incorporated. In SVR, the regularization parameter was adjusted to control model complexity and data fitting. For MLP, dropout layers or early stopping strategies were also used in addition to regularization parameters. The optimal parameter combination identified by GridSearchCV was then used to build the final prediction model, enhancing both its accuracy and generalization capability on unseen data. This advanced model optimization and cross-validation strategy ensured the robustness and reliability of the proposed models.

2.4. Health assessment using the BenMap model

This study utilized the BenMap model to assess the health impacts of various pollutant changes, illustrating the method's application in urban settings. As an effective spatial analysis tool, BenMap integrates GIS, environmental monitoring data, demographic information, and disease incidence rates, providing strong support for quantifying health impacts (Ghahremanloo et al., 2021; Sacks et al., 2018).

Initially, the study selected key pollutants such as PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃. Concentration distributions of these pollutants were predicted using machine learning models and combined with actual monitoring values to form the basis for assessment. The actual monitoring data for pollutant concentrations were obtained from the official websites of local government and environmental protection agencies, which publish hourly pollutant concentration data. Subsequently, baseline and target concentrations for each pollutant were established, using the 2019 monthly average concentrations as the reference, which is crucial for subsequent health risk assessments. ArcGIS software was used to georeference and vectorize the administrative map of Tianjin into a shapefile format with latitude and longitude information, which was input into the custom dataset. Population distribution and annual average mortality data were collected from authoritative statistical yearbooks and bulletins, specifically from the local statistics bureau's "Statistical Bulletin on National Economic and Social Development." This ensures the accuracy and scientific nature of the assessment process. By integrating these reliable data sources, the study ensures a robust and precise evaluation of the health impacts of various pollutants in an urban setting.

The BenMap model was used to evaluate the health impacts of air pollution by quantifying the relationship between pollutant concentrations and health effects using specific health impact functions. These functions are based on the latest scientific research and epidemiological data, ensuring the scientific and accurate assessment. Relevant data and models were imported into the BenMap software to comprehensively assess the potential health impacts of air pollution. BenMap offers multiple peer-reviewed health impact functions that can be used directly or customized. In this study, we customized functions using the BenMap editor to assess four main health outcomes: all-cause mortality, cardiovascular diseases, respiratory diseases, and lung cancer, providing precise scientific evidence for public health policy formulation. The formula for calculating the number of premature deaths (ΔM) caused by pollutants is shown in Eq. (5):

$$\Delta M = y_0 \times P \times \frac{RR - 1}{RR} \quad (5)$$

Here, y_0 represents the baseline mortality rate for a certain disease, which is the non-accidental mortality rate (total mortality rate - accidental mortality rate). P denotes the number of people exposed, and RR is the relative risk, which is the ratio of the probability of an event occurring in a group exposed to a certain risk compared to those not exposed. It can be calculated using Equation (Meng et al., 2023; So et al., 2022; Sun et al., 2023):

$$RR = \exp^{\beta(X - X_c)} \quad (6)$$

Here, β represents the coefficient between concentration and non-accidental mortality; X_c is the concentration threshold, indicating the theoretical minimum risk exposure value, which is generally considered negligible for human health impacts when the pollutant concentration is below this level; X is the average pollutants concentration derived from the simulation, in $\mu\text{g}/\text{m}^3$.

The open-source nature of the BenMap software offers great flexibility for processing raw data from different countries and regions. In this study, a custom dataset was specifically used to ensure the targeted and applicability of the method construction and data collection. Combined with the simulation study results, an in-depth assessment of the health benefits of pollutants was conducted.

3. Results

3.1. Impact of training data volume on machine learning model performance

During the preparation of BenMap pollution data, we found the dataset's scale often insufficient, affecting the accuracy of predictions. To address this, our study adopted an incremental training strategy where data was uniformly sampled from the original set and increased by 20 % each time. This method aims to evaluate and compare the performance of different models under data-constrained conditions, as illustrated in Table S1.

For CO: DT model showed significant improvement as data volume increased, culminating in an R^2 of 0.99 at 100 % data volume, showcasing nearly perfect prediction accuracy. The RF model's performance peaked at 80 % data volume with an R^2 of 0.56 and an RMSE of 0.42, while the SVR model reached its best at 80 % data volume with an R^2 of 0.68. The MLP model showed a modest increase in R^2 to 0.47 at full data volume. The DT model demonstrated the best overall performance for CO at full data volume. For NO_2 : The DT model again reached an R^2 of 0.99 at 100 % data volume, showing exceptional predictive accuracy. The RF model exhibited strong performance, especially at 100 % data volume with an R^2 of 0.87. The SVR and MLP models also showed improved performance with increased data volumes, with the MLP reaching an R^2 of 0.77 at 100 % data volume. The DT and RF models were the top performers for NO_2 , particularly at higher data volumes.

For O_3 : The DT model's performance peaked dramatically at 100 % data volume with an R^2 of 0.99. The RF model performed optimally at 80 % data volume with an R^2 of 0.75. MLP and SVR models also showed improvements with increasing data volumes, with the MLP reaching an R^2 of 0.65 at full data capacity. The DT model proved to be the most effective for O_3 at full data volume. For PM_{10} : The RF model displayed the best performance at 60 % data volume with an R^2 of 0.72. The DT model showed a significant jump to an R^2 of 0.99 at 100 % data volume, indicating a very high predictive performance with complete data. MLP and SVR models also demonstrated increasing trends, but the RF and DT models stood out, especially the DT at full data volume.

For $\text{PM}_{2.5}$: The RF model showed its best performance at 80 % data volume with an R^2 of 0.76 and an RMSE of 23.40. The DT model also reached an R^2 of 0.99 at 100 % data volume, echoing its strong performance across other pollutants. The MLP and SVR models improved consistently with increased data volume, but the RF and DT models were the most effective, with the DT model achieving the highest accuracy at full data volume. For SO_2 : The DT model dramatically improved to an R^2 of 0.99 at 100 % data volume, showcasing its effectiveness with complete data. The SVR model showed steady improvement, reaching an R^2 of 0.35 at 80 % data volume. The RF model had lower performance, and MLP showed negative trends at full data volume. For SO_2 , the DT model was notably the best performer at full data volume.

In conclusion, our expanded analysis across different training set sizes illustrates the significant impact of data volume on the predictive

performance of machine learning models. The DT model consistently shows exceptional performance at full data volume across all pollutants, indicating its robustness and effectiveness in environmental pollution prediction under optimal data conditions.

3.2. Impact of time steps on machine learning model performance

The time step for which a machine learning model can accurately predict has always been a critical focus. In this study, the performance of four machine learning models in predicting air pollutant concentrations was evaluated. Each model demonstrated specific advantages in predicting different pollutants. Fig. 2 illustrates the Heatmap of R^2 , RMSE, and MAPE scores across various pollutants and models, providing a comprehensive visualization of their predictive performance.

For DT, the highest R^2 value (0.99) for CO occurred at the 1872nd time step, while the lowest RMSE (10.44) and MAPE (10.40) were observed at the 88th time step. DT achieved its best predictive performance for NO_2 at the 2208th time step with an R^2 value of 0.99. O_3 prediction reached its peak for DT at the 2136th time step, also with an R^2 of 0.99. The prediction performance for particulate matter (PM_{10} and $\text{PM}_{2.5}$) was best at the 1896th and 2208th time steps, respectively, with R^2 values of 0.99. SO_2 prediction was most accurate at the 2064th time step, with an RMSE of 6.50 and a MAPE of 0.55. For MLP, the highest R^2 (0.87) for CO prediction was achieved at the 24th time step. NO_2 prediction was best at the 624th time step, with an R^2 of 0.42. O_3 prediction performance was optimal at the 2064th time step, with an R^2 of 0.77. PM_{10} prediction was best at the 2112th time step, with an R^2 of 0.25, while $\text{PM}_{2.5}$ prediction was optimal at the 624th time step, with an R^2 of 0.25. SO_2 prediction was most accurate at the 1488th time step, with an RMSE of 5.46 and a MAPE of 0.49.

For RF, the highest R^2 value (0.96) for CO prediction was observed at the 24th time step. NO_2 prediction was best at the 624th time step, with an R^2 of 0.95. O_3 prediction was optimal at the 2064th time step, with an R^2 of 0.98. PM_{10} and $\text{PM}_{2.5}$ predictions were best at the 912th and 624th time steps, respectively, with R^2 values of 0.93 and 0.95. SO_2 prediction was most accurate at the 24th time step, with an RMSE of 2.44 and a MAPE of 0.06. For SVR, the highest R^2 value (0.87) for CO prediction was observed at the 24th time step. NO_2 prediction was best at the 624th time step, with an R^2 of 0.49. O_3 prediction was optimal at the 72nd time step, with an R^2 of 0.45. PM_{10} prediction was best at the 912th time step, with an R^2 of 0.53, while $\text{PM}_{2.5}$ prediction was optimal at the 624th time step, with an R^2 of 0.55. SO_2 prediction was most accurate at the 24th time step, with an RMSE of 2.44 and a MAPE of 0.06.

In this section, we compared four different machine learning models to determine which model performs best for predicting specific air pollutants at specific time steps. Specifically, the DT showed the best performance in predicting CO and O_3 , with the highest R^2 values of 0.99 at the 1872nd and 2136th time steps, respectively. The RF excelled in predicting NO_2 , particulate matter (PM_{10} and $\text{PM}_{2.5}$), particularly at the 624th time step for NO_2 and $\text{PM}_{2.5}$, and at the 912th time step for PM_{10} , achieving high R^2 values. Additionally, the RF demonstrated high accuracy in predicting SO_2 at the 24th time step, with the lowest RMSE and MAPE values. These results suggest that when selecting models for specific pollutant predictions, the DT and RF provide the most reliable performance. Although the MLP and SVR models also performed well in certain scenarios, they did not exhibit as comprehensive or consistent performance as the DT and RF.

3.3. Impact of the number of meteorological factors on machine learning model performance

When exploring the impact of the number of meteorological factors on model prediction performance, we noticed that different combinations of meteorological factors directly affect prediction accuracy. Particularly when using machine learning techniques in conjunction with BenMap, incomplete meteorological data may be encountered.

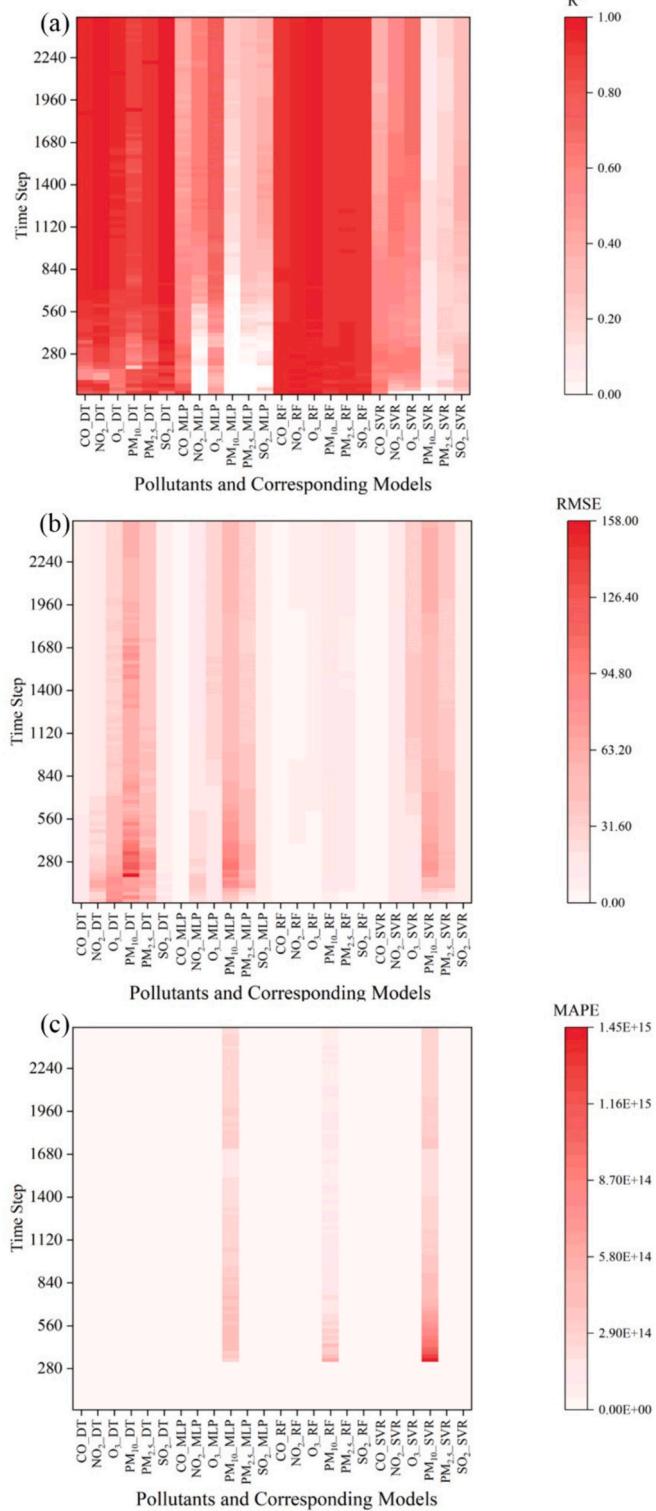


Fig. 2. Heatmap of R^2 , RMSE, and MAPE for different pollutants and models. (a-c) indicates the heatmaps of R^2 , RMSE, and MAPE, respectively.

Therefore, it is essential to identify which machine learning model can provide more accurate predictions when meteorological data is limited. As shown in Fig. 3, the performance indicators R^2 (coefficient of determination), RMSE (root mean square error), and MAPE (mean absolute percentage error) were compared across different models with varying numbers of meteorological factors, aiming to find the most suitable model for the current data situation. Fig. 4 further illustrates the

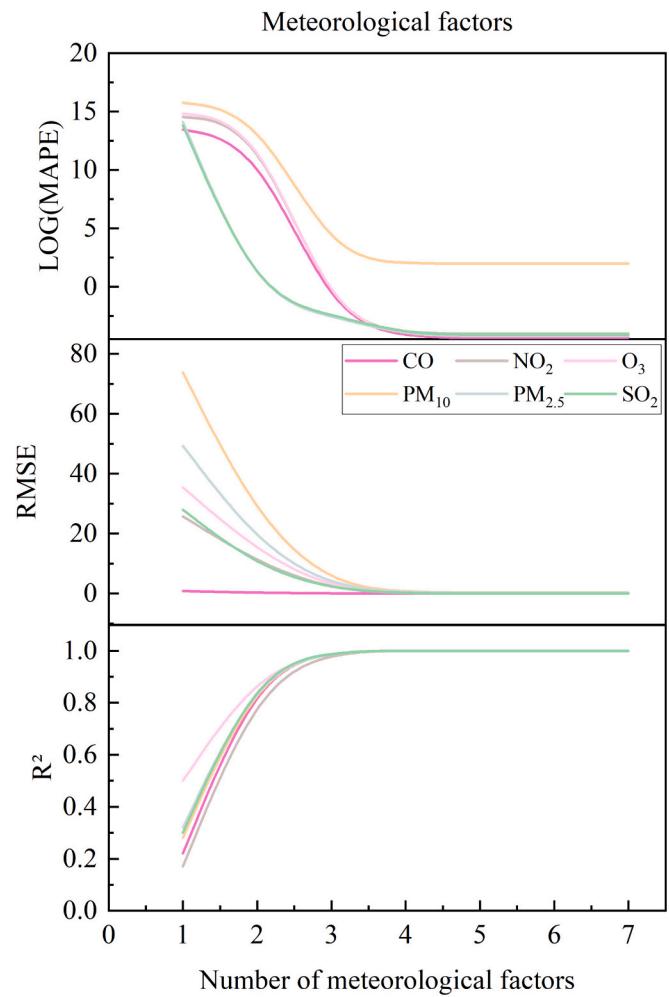


Fig. 3. R^2 , EMSE, MAPE of the best model corresponding to the number of meteorological factors.

prediction results for different pollutants with varying numbers of meteorological factors, providing a more intuitive comparison of prediction accuracy.

To further understand the individual impact of each meteorological factor on the prediction of atmospheric pollutant concentrations, this study specifically focused on the performance of prediction models with a single meteorological factor. Table S2 presents the best prediction model performance for different pollutants under various combinations of meteorological factors. From the table, it can be seen that both DT and RF models generally perform poorly when considering a single factor. For example, for the prediction of CO, both models have an R^2 value of 0.22 when only considering precipitation amount, indicating low model explanatory power. Additionally, the RMSE and MAPE values are relatively high, indicating limited prediction accuracy. Specifically, for the prediction of PM₁₀, the DT model has an R^2 value of only 0.28 when only considering precipitation amount, with a MAPE value as high as 5.76E+15, showing significant prediction error. This reflects the complexity of particulate matter concentration response to meteorological variables, especially when considering a single meteorological factor. Overall, the role of single meteorological factors in predicting atmospheric pollutants is limited.

Focusing on the performance of prediction models with two meteorological factors, the data indicates that the combination of atmospheric pressure and dew point temperature provides significant improvements in the prediction of most pollutants. For example, in the prediction of CO, this combination yields an R^2 value of 0.92, significantly higher

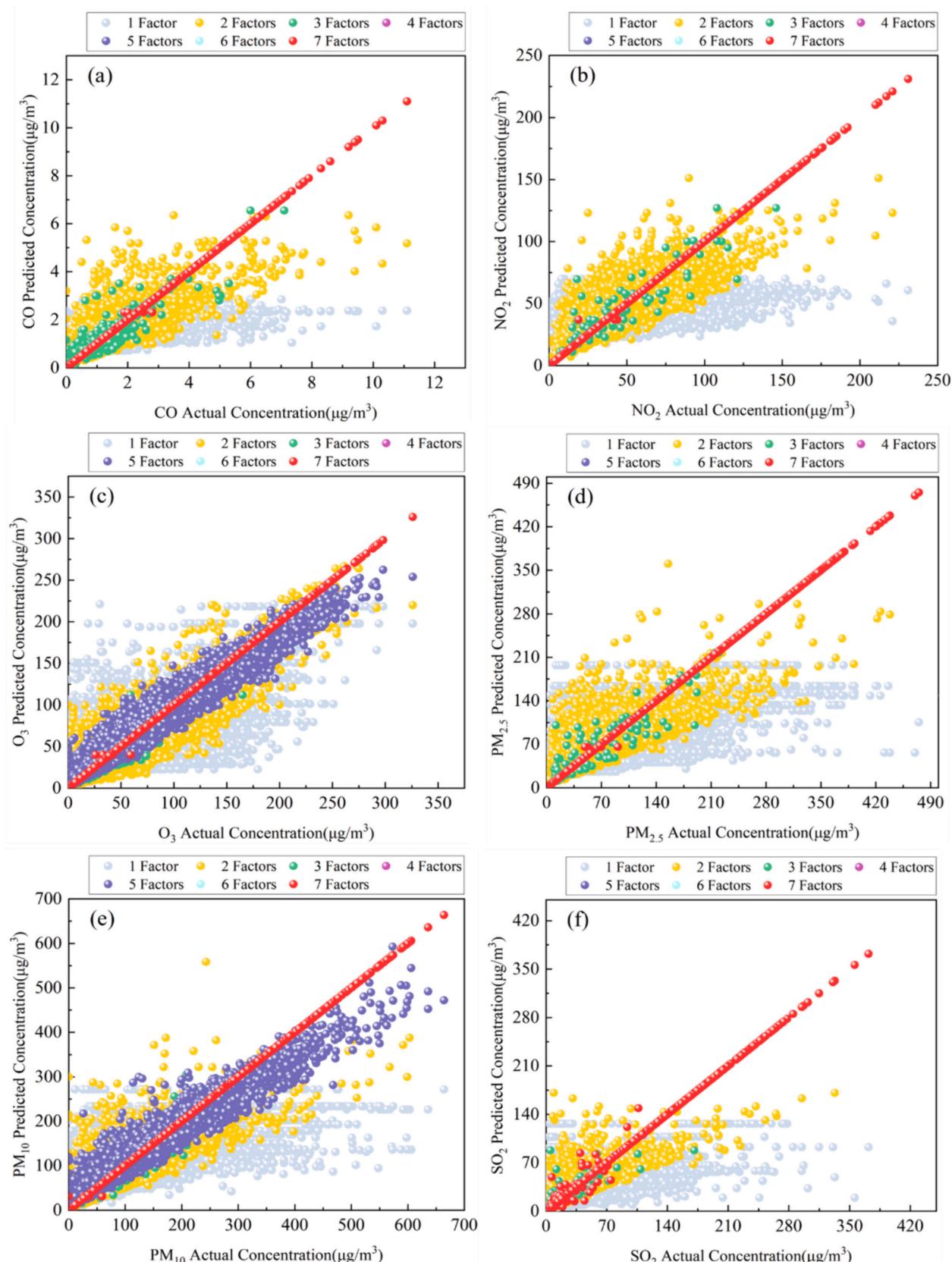


Fig. 4. Comparison of predicted and actual results for different pollutants with optimal machine learning models based on varying numbers of meteorological factors. (a-f) indicates the comparison of predicted and actual values for CO, NO_2 , O_3 , $\text{PM}_{2.5}$, PM_{10} , and SO_2 based on different numbers of meteorological factors.

than the performance with a single factor, with RMSE and MAPE values significantly reduced to 0.26 and 2.10E+12, respectively, indicating high prediction accuracy and low error. For O₃, the combination of atmospheric pressure and precipitation amount also shows high explanatory power ($R^2 = 0.92$), although the MAPE is 1.13E+14, still suggesting some prediction error. For PM₁₀, PM_{2.5}, and SO₂, using the combination of atmospheric pressure and dew point temperature not only achieves R^2 values above 0.92 but also exhibits very low MAPE values, especially for PM_{2.5} and SO₂, with MAPE values of 0.11 and 0.13, respectively, indicating very high prediction accuracy.

Further research explored the impact of three meteorological factor combinations to evaluate their contribution to prediction performance. The data shows that the combination of atmospheric pressure, relative humidity, and dew point temperature provides very high explanatory power ($R^2 = 0.99$) for the prediction of CO, NO₂, PM₁₀, PM_{2.5}, and SO₂. These combinations not only improved model explanatory power but also significantly reduced prediction errors, such as the RMSE and MAPE for CO being 0.04 and 0.01, respectively, indicating excellent prediction accuracy. For O₃, the combination of atmospheric pressure, relative humidity, and precipitation amount further improved the model's explanatory power to 0.99, with a MAPE of 0.01, highlighting the importance of composite meteorological factors in improving prediction accuracy. However, it is noteworthy that despite the generally high explanatory power, the MAPE for PM₁₀ is exceptionally high at 2.82E+2, suggesting that other potential influencing factors need to be considered when predicting particulate matter concentrations.

Table S2 also shows the optimal prediction performance of models when including combinations of four, five, and six meteorological factors. These factors include atmospheric temperature, atmospheric pressure, dew point temperature, precipitation amount, relative humidity, wind speed, and total cloud cover. The results indicate that as the number of meteorological factors increases, the prediction performance of the models significantly improves. For pollutants such as CO, PM₁₀, PM_{2.5}, SO₂, NO₂, and O₃, the explanatory power (R^2) is close to or reaches 0.99, demonstrating that the models can very accurately reflect changes in actual observations. Especially in the DT model, the predictions for CO, PM_{2.5}, and SO₂ show outstanding performance, with very low RMSE and MAPE values, indicating very high prediction accuracy. For NO₂ and O₃, when the models include atmospheric temperature, atmospheric pressure, relative humidity, and precipitation amount, they also show very high explanatory power. However, for PM₁₀, despite the very high explanatory power (0.99) with a combination of six meteorological factors, the MAPE is exceptionally high (9.31E+01), possibly indicating specific challenges in predicting PM₁₀, such as the complexity of its sources and the variability of environmental factors, which may not be fully captured by the existing models. Additionally, the RF model also shows excellent prediction performance under multi-factor combinations, especially when including wind speed and total cloud cover, demonstrating good model adaptability and prediction accuracy for O₃ and PM₁₀. These results highlight the importance of including comprehensive meteorological factors in prediction models to significantly improve prediction accuracy and reliability.

3.4. BenMap health impact assessment

The DT model exhibited exceptional performance in predicting atmospheric pollutant concentrations in Tianjin. The model achieved near-perfect explanatory power ($R^2 = 0.99$) for various pollutants, including CO, NO₂, O₃, particulate matter (PM₁₀ and PM_{2.5}), and SO₂, demonstrating high precision and reliability. Specifically, the DT model achieved a root mean square error (RMSE) of 0.01 and a mean absolute percentage error (MAPE) of 4.24E-05 for CO, indicating extremely high prediction accuracy. For SO₂, the RMSE decreased to 0.02, and the MAPE was 7.97E-05, further confirming the model's effectiveness in handling minimal error levels.

A combination of machine learning predictive models and BenMap software was employed to assess the health impacts of different pollutants, including PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃. As shown in Fig. 5, the results indicate a high degree of consistency between predicted and actual values for various pollutants and different causes of death. This high degree of consistency is attributed to the fact that the R^2 values for the pollutant prediction results were all above 0.99, indicating that the model predictions were extremely close to the actual values. For example, the predicted all-cause mortality for PM_{2.5} was 3120, which matched the actual value of 3120; the predicted and actual values for cardiovascular disease were both 1560, and for respiratory disease, both were 780. For PM₁₀, the predicted all-cause mortality was 3120, which also matched the actual value of 3120; the predicted and actual values for cardiovascular disease were both 1560, and for respiratory disease, the predicted and actual values were both 780. These results demonstrate that the machine learning model combined with BenMap can accurately assess the health risks and associated mortality from different pollutants, providing an effective tool for environmental health risk assessment.

4. Discussion on the generalizability of the research

This study addresses the issue of limited meteorological factor data and missing pollutant data when using BenMap for evaluation. In this section, data from Chengdu, a city with distinct geographical and climatic characteristics compared to Tianjin, is utilized to validate the generalizability of our approach. Chengdu, situated in the Sichuan Basin, features complex terrain and significant local climate variations influenced by factors such as basin effects and monsoon patterns. In contrast, Tianjin is located in the North China Plain, characterized by relatively flat terrain and a temperate continental climate with cold, dry winters and hot, humid summers. These pronounced geographical and climatic differences contribute to diverse challenges in air quality formation and management between the two cities. Despite these geographic and climatic disparities, Chengdu and Tianjin share common challenges in urban air quality management and pollution monitoring. Both cities confront issues related to urbanization, including traffic pollution, industrial emissions, and the management of residential waste, all of which pose potential environmental and public health impacts. Effective pollution monitoring and prediction tools are essential for both cities to support decision-making and implement environmental policies aimed at improving air quality and reducing health risks. Therefore, Chengdu was selected to test the portability of our machine learning models not only due to its shared urban air quality challenges with Tianjin but also to demonstrate the applicability and robustness of our models in different urban environments. This choice aims to validate the broad applicability of our approach and to provide new perspectives and methodologies for global urban air quality assessment.

Based on the above research, the machine learning methods from this study were applied to data from Chengdu, using data from 2013 to 2018 for training and predicting PM_{2.5}, CO, NO₂, O₃, PM₁₀, and SO₂ for 2019, incorporating atmospheric pressure, relative humidity, and dew point temperature, as shown in Fig. 6. The model's performance was validated using various evaluation metrics. The results showed a high degree of fit to the data with R^2 values of 0.90 for PM_{2.5}, 0.78 for CO, 0.73 for NO₂, 0.81 for O₃, 0.89 for PM₁₀, and 0.63 for SO₂. The RMSE values indicated small average errors between predicted and actual values, with 8.63 for PM_{2.5}, 0.11 for CO, 8.70 for NO₂, 13.13 for O₃, 13.35 for PM₁₀, and 1.70 for SO₂. Additionally, the MAPE values showed that the prediction errors for pollutant concentrations were within acceptable ranges, with 0.23 for PM_{2.5}, 0.11 for CO, 0.22 for NO₂, 0.54 for O₃, 0.21 for PM₁₀, and 0.20 for SO₂. These comprehensive results indicate that the model performs well in predicting different pollutants, further demonstrating the generalizability and robustness of the method.

Based on the analysis of the predicted and actual data of various pollutants' health impacts in Chengdu using the BenMap model, the

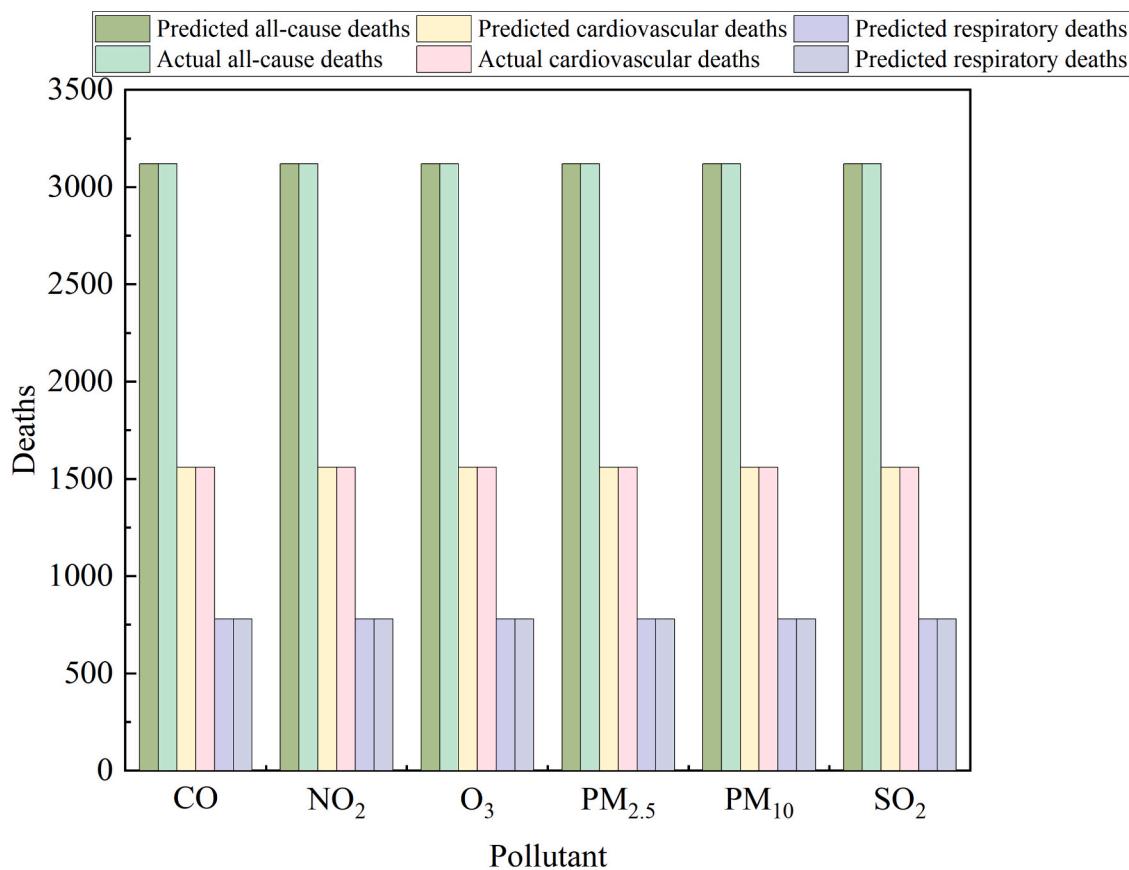


Fig. 5. BenMap predicted and actual health assessment results.

results show that the predicted values are almost completely consistent with the actual values for all-cause, cardiovascular disease, and respiratory disease impacts. Specifically, for PM_{2.5}, the predicted all-cause health impact value is 195,965, matching the actual value; the predicted and actual values for cardiovascular diseases are both 97,983; and for respiratory diseases, the predicted and actual values are both 48,991. Similarly, for SO₂, the predicted all-cause health impact value is 195,960, very close to the actual value of 195,960; for cardiovascular diseases, both predicted and actual values are 97,980; and for respiratory diseases, both predicted and actual values are 48,990. For NO₂, the predicted all-cause health impact value is 195,962, very close to the actual value of 195,962; for cardiovascular diseases, the predicted value is 97,981, close to the actual value of 97,981; and for respiratory diseases, the predicted value is 48,990, close to the actual value of 48,990. For CO, the predicted all-cause health impact value is 195,960, matching the actual value; for cardiovascular diseases, both predicted and actual values are 97,980; and for respiratory diseases, both predicted and actual values are 48,990. For O₃, the predicted all-cause health impact value is 195,962, very close to the actual value of 195,962; for cardiovascular diseases, the predicted value is 97,981, close to the actual value of 97,981; and for respiratory diseases, the predicted value is 48,990, close to the actual value of 48,990. For PM₁₀, the predicted all-cause health impact value is 195,975, very close to the actual value of 195,974; for cardiovascular diseases, the predicted value is 97,987, close to the actual value of 97,987; and for respiratory diseases, the predicted value is 48,994, close to the actual value of 48,994. These results indicate that the methodology in this study is scalable and that the combination with the BenMap model provides high accuracy and reliability in assessing the health impacts of various pollutants.

5. Conclusions

This study aims to address accuracy issues caused by limited meteorological factor data and missing pollutant data when using BenMap to assess the health impacts of air pollution. By employing different data increment strategies and various machine learning models, the study explores the effects of data volume, time steps, and the number of meteorological factors on model prediction performance. The key findings are:

- (1) Impact of Training Data Volume: Increasing the training data volume significantly improves the predictive performance of RF and DT models, especially for CO, NO₂, and PM_{2.5} predictions.
- (2) Impact of Time Steps: The optimal prediction time step varies by pollutant. The DT model shows high explanatory power, achieving the highest R² value (0.99) for CO and O₃.
- (3) Impact of Meteorological Factors: Combining multiple meteorological factors (e.g., atmospheric pressure, relative humidity, dew point temperature) significantly enhances model accuracy, with an R² of 0.99 for CO, NO₂, PM₁₀, PM_{2.5}, and SO₂ predictions.
- (4) Method Generalizability: Applying the proposed methods to Chengdu data confirms excellent performance across multiple pollutants, validating the method's broad applicability.

Overall, the machine learning and BenMap-based methods demonstrate high accuracy and reliability in predicting air pollutant concentrations and health impacts, showing robust performance under varying data and meteorological conditions. These methods are valuable for air pollution assessments in other urban contexts. Future research will explore the cumulative and interactive health impacts of multiple pollutants to deepen our understanding of compounded health risks.

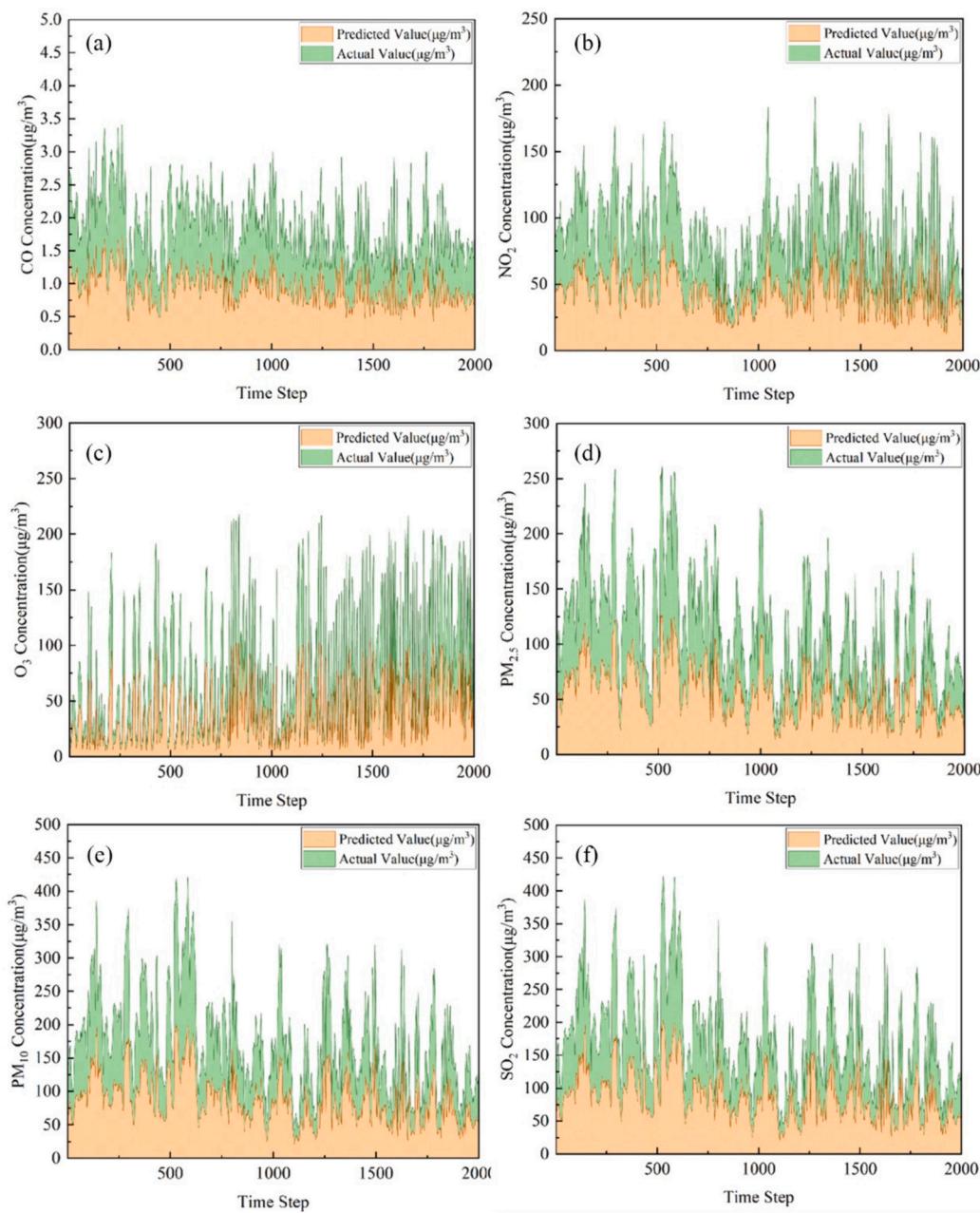


Fig. 6. Extended study: BenMap predicted and actual health assessment results. (a-f) indicates the comparison of predicted and actual values for CO, NO_2 , O_3 , $\text{PM}_{2.5}$, PM_{10} , and SO_2 .

CRediT authorship contribution statement

Juncheng Wu: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Qili Dai:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Shaojie Song:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (42205109), the Natural Science Foundation of Tianjin City (22JCYBJC01330), and TianHe Qingsuo open research fund of TSYS in 2022 & NSCC-TJ.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2024.175246>.

References

- Afzal, S., Ziapour, B.M., Shokri, A., Shakibi, H., Sobhani, B., 2023. Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms. *Energy* 282, 128446. <https://doi.org/10.1016/j.energy.2023.128446>.
- Aguiar-Gil, D., Gómez-Peláez, L.M., Álvarez-Jaramillo, T., Correa-Ochoa, M.A., Saldarriaga-Molina, J.C., 2020. Evaluating the impact of PM_{2.5} atmospheric pollution on population mortality in an urbanized valley in the American tropics. *Atmos. Environ.* 224, 117343 <https://doi.org/10.1016/j.atmosenv.2020.117343>.
- Ai, B., Zhang, J., Zhang, S., Chen, G., Tian, F., Chen, L., et al., 2024. Causal association between long-term exposure to air pollution and incident Parkinson's disease. *J. Hazard. Mater.* 469, 133944 <https://doi.org/10.1016/j.jhazmat.2024.133944>.
- Alomari, Y., Andó, M., 2024. SHAP-based insights for aerospace PHM: temporal feature importance, dependencies, robustness, and interaction analysis. *Results Eng.* 21, 101834 <https://doi.org/10.1016/j.rineng.2024.101834>.
- Chen, L., Shi, M., Gao, S., Li, S., Mao, J., Zhang, H., et al., 2017. Assessment of population exposure to PM_{2.5} for mortality in China and its public health benefit based on BenMap. *Environ. Pollut.* 221, 311–317. <https://doi.org/10.1016/j.envpol.2016.11.080>.
- Chen, Z., Dou, S., Zhao, C., Xiao, L., Lu, Z., Qiu, Y., 2024. Machine learning-assisted assessment of key meteorological and crop factors affecting historical mulch pollution in China. *J. Hazard. Mater.* 465, 133281 <https://doi.org/10.1016/j.jhazmat.2023.133281>.
- Devasahayam, S., Albijanic, B., 2024. Predicting hydrogen production from co-gasification of biomass and plastics using tree based machine learning algorithms. *Renew. Energy* 222, 119883. <https://doi.org/10.1016/j.renene.2023.119883>.
- Ducruet, C., Polo Martin, B., Sene, M.A., Lo Prete, M., Sun, L., Itoh, H., et al., 2024. Ports and their influence on local air pollution and public health: a global analysis. *Sci. Total Environ.* 915, 170099 <https://doi.org/10.1016/j.scitotenv.2024.170099>.
- Gahremanloo, M., Choi, Y., Sayeed, A., Salman, A.K., Pan, S., Amani, M., 2021. Estimating daily high-resolution PM_{2.5} concentrations over Texas: Machine Learning approach. *Atmos. Environ.* 247, 118209 <https://doi.org/10.1016/j.atmosenv.2021.118209>.
- Goudarzi, G., Hopke, P.K., Yazdani, M., 2021. Forecasting PM_{2.5} concentration using artificial neural network and its health effects in Ahvaz, Iran. *Chemosphere* 283, 131285. <https://doi.org/10.1016/j.chemosphere.2021.131285>.
- Han, S.-B., Song, S.-K., Shon, Z.-H., Kang, Y.-H., Bang, J.-H., Oh, I., 2021. Comprehensive study of a long-lasting severe haze in Seoul megacity and its impacts on fine particulate matter and health. *Chemosphere* 268, 129369. <https://doi.org/10.1016/j.chemosphere.2020.129369>.
- Kalisa, E., Clark, M.L., Ntakirutimana, T., Amani, M., Volckens, J., 2023. Exposure to indoor and outdoor air pollution in schools in Africa: current status, knowledge gaps, and a call to action. *Heliyon* 9, e18450. <https://doi.org/10.1016/j.heliyon.2023.e18450>.
- Li, Y., Wu, Q., Wang, X., Cheng, H., Sun, Y., Li, D., et al., 2024. Effects of chemical mechanism and meteorological factors on the concentration of atmospheric pollutants in the megacity Beijing, China. *Atmos. Environ.* 323, 120393 <https://doi.org/10.1016/j.atmosenv.2024.120393>.
- Meng, X., Hang, Y., Lin, X., Li, T., Wang, T., Cao, J., et al., 2023. A satellite-driven model to estimate long-term particulate sulfate levels and attributable mortality burden in China. *Environ. Int.* 171, 107740 <https://doi.org/10.1016/j.envint.2023.107740>.
- Nguyen, T.H., Nagashima, T., Doan, Q.-V., Khan, A., Niyogi, D., 2023. Source apportionment of PM_{2.5} and the impact of future PM_{2.5} changes on human health in the monsoon-influenced humid subtropical climate. *Atmos. Pollut. Res.* 14, 101777 <https://doi.org/10.1016/j.apr.2023.101777>.
- Ni, J., Zhao, Y., Li, B., Liu, J., Zhou, Y., Zhang, P., et al., 2023. Investigation of the impact mechanisms and patterns of meteorological factors on air quality and atmospheric pollutant concentrations during extreme weather events in Zhengzhou city, Henan Province. *Atmos. Pollut. Res.* 14, 101932 <https://doi.org/10.1016/j.apr.2023.101932>.
- Parthum, B., Pindilli, E., Hogan, D., 2017. Benefits of the fire mitigation ecosystem service in The Great Dismal Swamp National Wildlife Refuge, Virginia, USA. *J. Environ. Manag.* 203, 375–382. <https://doi.org/10.1016/j.jenvman.2017.08.018>.
- Sacks, J.D., Lloyd, J.M., Zhu, Y., Anderton, J., Jang, C.J., Hubbell, B., et al., 2018. The environmental Benefits Mapping and Analysis Program – Community Edition (BenMap–CE): a tool to estimate the health and economic benefits of reducing air pollution. *Environ. Model Softw.* 104, 118–129. <https://doi.org/10.1016/j.envsoft.2018.02.009>.
- Sadiq, M.A., Sarkar, S.K., Raisa, S.S., 2023. Meteorological drought assessment in northern Bangladesh: a machine learning-based approach considering remote sensing indices. *Ecol. Indic.* 157, 111233 <https://doi.org/10.1016/j.ecolind.2023.111233>.
- So, R., Andersen, Z.J., Chen, J., Stafoggia, M., de Hoogh, K., Katsouyanni, K., et al., 2022. Long-term exposure to air pollution and mortality in a Danish nationwide administrative cohort study: beyond mortality from cardiopulmonary disease and lung cancer. *Environ. Int.* 164, 107241 <https://doi.org/10.1016/j.envint.2022.107241>.
- de Souza Fernandes Duarte, E., Lucio, P.S., Costa, M.J., Salgueiro, V., Salgado, R., Potes, M., et al., 2024. Pollutant-meteorological factors and cardio-respiratory mortality in Portugal: seasonal variability and associations. *Environ. Res.* 240, 117491 <https://doi.org/10.1016/j.envres.2023.117491>.
- Su, D., Chen, L., Wang, J., Zhang, H., Gao, S., Sun, Y., et al., 2024. Long- and short-term health benefits attributable to PM_{2.5} constituents reductions from 2013 to 2021: a spatiotemporal analysis in China. *Sci. Total Environ.* (907), 168184 <https://doi.org/10.1016/j.scitotenv.2023.168184>.
- Sun, H.Z., Zhao, J., Liu, X., Qiu, M., Shen, H., Guillas, S., et al., 2023. Antagonism between ambient ozone increase and urbanization-oriented population migration on Chinese cardiopulmonary mortality. *The Innovation* 4, 100517. <https://doi.org/10.1016/j.xinn.2023.100517>.
- Verma, P., Verma, R., Mallet, M., Sisodiya, S., Zare, A., Dwivedi, G., et al., 2024. Assessment of human and meteorological influences on PM₁₀ concentrations: insights from machine learning algorithms. *Atmos. Pollut. Res.* 15, 102123 <https://doi.org/10.1016/j.apr.2024.102123>.
- Voorhees, A.S., Wang, J., Wang, C., Zhao, B., Wang, S., Kan, H., 2014. Public health benefits of reducing air pollution in Shanghai: a proof-of-concept methodology with application to BenMap. *Sci. Total Environ.* 485–486, 396–405. <https://doi.org/10.1016/j.scitotenv.2014.03.113>.
- Wang, H., Sun, F., Liu, F., Wang, T., Liu, W., Feng, Y., 2023. Reconstruction of the pan evaporation based on meteorological factors with machine learning method over China. *Agric. Water Manag.* 287, 108416 <https://doi.org/10.1016/j.agwat.2023.108416>.
- Xu, Y., Kong, X., Cai, Z., 2024. Cross-validation strategy for performance evaluation of machine learning algorithms in underwater acoustic target recognition. *Ocean Eng.* 299, 117236 <https://doi.org/10.1016/j.oceaneng.2024.117236>.
- Zhang, Z., Xu, B., Xu, W., Wang, F., Gao, J., Li, Y., et al., 2022. Machine learning combined with the PMF model reveal the synergistic effects of sources and meteorological factors on PM_{2.5} pollution. *Environ. Res.* 212, 113322 <https://doi.org/10.1016/j.enres.2022.113322>.
- Zhong, J., Hodgson, J.R., James Bloss, W., Shi, Z., 2023. Impacts of net zero policies on air quality in a metropolitan area of the United Kingdom: towards world health organization air quality guidelines. *Environ. Res.* 236, 116704 <https://doi.org/10.1016/j.enres.2023.116704>.