Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 2, Number 2, October 2022, Page 9-16
https://journal.physan.org/index.php/jocpes/index

Machine Learning Regression Modeling Analysis for PM 2.5 Concentration Estimation in Jakarta: Approaches and Implications for Air Quality

Brilliant Muhammad Al Hadid Deva Sudarjo[1]
[1] State of Meteorology Climatology and Geophysics Agency

Article Info

Article history:

Received September 5, 2022
Revised September 10, 2022
Accepted September 11, 2022

Keywords:

PM 2.5
Climate Change
Machine Learning
Air Quality Prediction
Climate Variability

ABSTRACT

Air pollution by fine particulate matter (PM2.5) significantly impacts public health and environmental stability. As an air pollutant, PM2.5 is influenced by climate factors such as temperature, humidity, and wind patterns, all of which fluctuate due to climate change. This literature review explores the application of machine learning (ML) in predicting and analyzing PM2.5 behavior, focusing on three primary methods: Support Vector Regression (SVR), Random Forest (RF), and Neural Networks (NN). Based on 20 studies, this review compares the strengths and limitations of each method, evaluating how ML techniques address the complexity and variability of climate data in the context of PM2.5.

Corresponden Author:

Brilliant Muhammad Al Hadid Deva Sudarjo,
State of Meteorology Climatology and Geophysics Agency
Tangerang City, Banten, Indonesia
Email: aved707@gmail.com

1. INTRODUCTION

The increase in PM2.5 concentrations has become a significant environmental and public health concern worldwide. Fine particulate matter, commonly referred to as PM2.5, consists of particles with diameters of less than 2.5 micrometers. These particles are small enough to penetrate deep into the human respiratory system, potentially leading to severe health effects, including respiratory and cardiovascular diseases. According to Gao et al. (2020), PM2.5 exposure is directly linked to increased mortality rates from respiratory infections, heart disease, and other related health complications. Additionally, high concentrations of PM2.5 reduce visibility, negatively impact ecosystems, and contribute to climate change through complex chemical and physical processes in the atmosphere.

The relationship between PM2.5 and climate change is multifaceted. Climate change affects PM2.5 dynamics by altering meteorological conditions, including temperature, humidity, wind speed, and precipitation patterns. These meteorological variables significantly influence the formation, transformation, transport, and deposition of PM2.5 in the atmosphere. For instance, high temperatures accelerate chemical reactions that lead to the formation of secondary pollutants—one of the primary components of PM2.5. Similarly, changes in wind patterns and precipitation impact PM2.5 dispersion and removal, affecting regional air quality. Therefore, understanding PM2.5 behavior within the context of climate variability is crucial for developing effective pollution control and public health interventions (Skyllakou et al., 2021).

Traditional models for predicting PM2.5, such as chemical transport models (CTMs), have provided valuable insights into pollutant behavior but are limited in their ability to capture the complex, non-linear relationships between climate variables and PM2.5. These models often require extensive domain-specific knowledge, detailed emission inventories, and substantial computational power, which can restrict their applicability across diverse regions and varying environmental conditions. As an alternative, machine learning (ML) offers powerful tools for modeling complex datasets with high-dimensional, non-linear interactions. ML algorithms can identify patterns in large datasets, making them well-suited to capture the intricate relationships between PM2.5 concentrations and climate variables, even when those relationships vary across regions and seasons (Zhai et al., 2019).

With advancements in machine learning, PM2.5 data analysis can now be conducted more efficiently. Machine learning offers the capability to identify patterns and predict PM2.5 changes by considering complex, interrelated climate variables (Zhai et al., 2019). This study aims to compare the effectiveness of several primary ML models in predicting PM2.5, focusing on climate variability. We examine three key models: Support Vector Regression (SVR), Random Forest (RF), and Neural Networks (NN), each with advantages for handling non-linear and high-dimensional data.

2. LITERATURE REVIEW

2.1 Air Quality and the impact of PM 2.5

Air pollution is a major global concern, especially in large cities like Jakarta, due to its severe impact on human health and quality of life. One of the most harmful pollutants is Particulate Matter (PM), specifically PM 2.5, which refers to fine particles with a diameter of less than 2.5 microns. PM 2.5 poses significant health risks because it can penetrate the respiratory system, reach the lungs, and even enter the bloodstream. According to the World Health Organization (WHO), long-term exposure to PM 2.5 increases the risk of chronic respiratory diseases, cardiovascular disorders, lung cancer, and even premature death (WHO, 2018).

In Jakarta, poor air quality is primarily caused by a combination of motor vehicle emissions, industrial activities, waste burning, and natural phenomena like forest fires. Data from the Jakarta Environmental Management Agency shows that PM 2.5 concentrations often exceed the threshold set by the WHO, leading to increased hospital admissions and various health issues among the population (Government of Jakarta, 2020). Therefore, monitoring and predicting PM 2.5 levels are crucial for formulating effective policies to mitigate air pollution in the city.

2.2 Machine Learning in Air Quality Prediction

The application of machine learning (ML) in air quality analysis has rapidly expanded in recent years. Machine learning regression models, especially for predicting PM 2.5, have proven effective in uncovering complex patterns in time-series data that are often influenced by multiple factors. Algorithms such as linear regression, decision trees, random forests, and gradient boosting have been widely used in studies related to air quality prediction.

One widely adopted approach is Gradient Boosting, which is an ensemble learning method that combines several simple regression models (decision trees) to form a robust model. Unlike simple linear regression, Gradient Boosting Regressor (GBR) is particularly effective in capturing complex, non-linear relationships. This is essential in predicting air quality, which is subject to fluctuations due to various environmental, temporal, and human activity factors.

Previous studies have demonstrated the efficacy of Gradient Boosting in predicting pollutant concentrations. Compared to other algorithms such as Support Vector Machines (SVM) and Random Forests, Gradient Boosting often yields more accurate results, especially when dealing with volatile and complex data (Zhang et al., 2020). As a result, this study selects Gradient Boosting Regressor (GBR) as the primary model for predicting PM 2.5 levels in Jakarta.

2.3 PM 2.5 Prediction Using Gradient Boosting

Gradient Boosting is an ensemble learning algorithm that sequentially builds several decision trees to form a powerful prediction model. At each iteration, the algorithm learns from the errors made by the previous model, correcting those mistakes in the next one. This ability to iteratively correct errors makes GBR particularly well-suited for handling datasets with high volatility and non-linear relationships, such as PM 2.5 data, which is influenced by a wide range of factors.

The application of Gradient Boosting in air quality prediction has been proven effective in previous studies, where this model not only handles data instability but also provides more accurate predictions compared to other methods. While it requires more training time, the benefit of this method lies in its ability to capture more complex relationships between the input features and the target (PM 2.5), which simpler models like linear regression cannot reveal.

3. METHODOLOGY

3.1 Data Description

This study utilizes daily **PM 2.5** concentration data recorded in Jakarta from **2018 to 2021** for training the model, and data from **2022 to 2024** is used as test data to evaluate the model's prediction performance. The data was obtained from the air quality monitoring stations managed by the Jakarta Environmental Management Agency. The dataset contains the following key columns:

1. **Date**: The date on which the PM 2.5 concentration was recorded.
2. **PM 2.5**: The concentration of PM 2.5 in micrograms per cubic meter (μg/m³). This is the target variable that the model aims to predict.
3. **Location**: Information about the measurement location (if available, for instance, various monitoring points across Jakarta).

The dataset represents fluctuations in PM 2.5 concentrations over time, influenced by various factors such as temperature, humidity, vehicle density, and industrial activities. However, for the purpose of this study, we focus only on the **PM 2.5** variable, excluding external factors. This will provide a simplified view of how historical PM 2.5 data alone can be used to predict future levels of PM 2.5.

3.2 Data Preprocessing

Before the data can be used for model training, several preprocessing steps are performed to ensure the dataset is clean and ready for machine learning:

1. **Handling Missing Values**: PM 2.5 data often contains missing values due to various reasons such as sensor malfunctions or missing records. To address this, **linear interpolation** is used to fill in the missing values. Linear interpolation estimates missing data points by taking the average of the two closest data points before and after the missing value. For instance, if PM 2.5 values are recorded on the 15th and 17th of the month, the value for the 16th will be interpolated based on the values from the 15th and 17th.
2. **Outlier Detection and Handling**: Outliers are extreme values that are significantly different from other data points and can skew the performance of the model. **Z-score** is employed to detect outliers. The Z-score measures how far a data point is from the mean in terms of standard deviations. Any data point with a Z-score greater than 3 or less than -3 is considered an outlier and removed from the dataset.
3. **Data Normalization**: In machine learning, normalizing the data is crucial to ensure that all features are on the same scale. **Min-Max Scaling** is applied to transform the PM 2.5 values into a range between 0 and 1. This is important because features with larger ranges may dominate the model, leading to biased results. Normalizing ensures that no feature disproportionately influences the training process.
4. **Data Splitting for Training and Testing**: Once the data is cleaned, it is split into two primary sets:
   - **Training Set**: Data from **2018 to 2021** is used to train the model.
   - **Test Set**: Data from **2022 to 2024** is used to evaluate the model's predictions and compare them with actual recorded values.

3.3 Model Selection

The model selected for predicting PM 2.5 levels is the **Gradient Boosting Regressor (GBR)**. Gradient Boosting is an ensemble method that sequentially builds several decision trees to create a robust prediction model. At each iteration, it adjusts the predictions to correct for errors made by previous trees, resulting in improved accuracy over time. GBR is well-suited for handling complex, non-linear data, such as PM 2.5, which fluctuates due to various environmental and human activity factors.

The steps involved in building the model are as follows:

1. **Training the Model**: The **Gradient Boosting Regressor** is trained using the **2018-2021 training set** to learn the relationships between the features and the target (PM 2.5 levels).
2. **Hyperparameter Tuning**: **GridSearchCV** is used to optimize hyperparameters such as the number of trees (n_estimators), maximum depth of the trees (max_depth), and learning rate, to ensure the model performs at its best.
3. **Model Validation**: After training, the model is tested using the **2022-2024 test set** to measure its prediction accuracy.
4. **Model Optimization**: The model can be further fine-tuned by adjusting hyperparameters or refining data preprocessing steps if necessary.

3.4 Model Evaluation

The model's performance is evaluated using several commonly used metrics for regression tasks:

1. **Mean Absolute Error (MAE)**: MAE calculates the average absolute difference between the predicted and actual values, providing a straightforward measure of model accuracy.
2. **Root Mean Squared Error (RMSE):** RMSE provides a more detailed measure of error, giving greater weight to larger errors. It is particularly useful when larger deviations are more critical to address. RMSE is calculated as the square root of the average of the squared differences between predicted and actual values. A lower RMSE indicates that the model is better at predicting PM 2.5 concentrations.
3. **R-squared (R^2):** This metric measures the proportion of the variance in the dependent variable (PM 2.5) that is explained by the independent variables in the model. R^2 ranges from 0 to 1, with higher values indicating that the model is able to explain a larger portion of the variance in PM 2.5 concentrations. R^2 will be used to evaluate how well the model generalizes across the training and test data.
4. **Residual Analysis**: Residuals are the differences between the observed actual outcomes and the predictions made by the model. Residual analysis is important to check if there is any pattern left unexplained by the model. Ideally, the residuals should be randomly distributed, indicating that the model has captured the underlying patterns in the data effectively. If the residuals show any systematic trends, it might suggest the model needs improvement or that additional features should be considered.

3.5 Model Implementation

Once the model has been evaluated and tuned, the Gradient Boosting Regressor model will be used for PM 2.5 forecasting. The predictions made by the model will estimate future levels of PM 2.5, helping policymakers and relevant authorities prepare for periods of poor air quality. These predictions can be used to plan interventions such as traffic restrictions, industrial shutdowns, or public health warnings.

In terms of implementation, the model will be deployed to predict PM 2.5 levels on a daily basis for the years 2022-2024 using the historical training data from 2018 to 2021. The accuracy of these forecasts will be checked against actual measurements, providing an assessment of the model's real-world application.

Additionally, model performance will be assessed by comparing the predicted values to the actual PM 2.5 concentrations recorded during the test period. This will help evaluate the robustness and reliability of the model, while also providing insights into the potential utility of such models in air quality management systems.

4. RESULT AND DISCUSSION

4.1 Data Overview and Preprocessing

Before diving into the results, it's important to note that the dataset used spans from **2018 to 2021**. It includes daily PM 2.5 measurements collected in Jakarta, covering both typical and extreme air quality conditions.

The data was preprocessed by:

- **Cleaning** missing values using **linear interpolation**.
- **Normalizing** the PM 2.5 values to ensure uniform scale across the dataset.
- **Feature Engineering**, where time-based features like **month** and **day of the week** were considered but not used in the final model, though they could improve the model further.

The **Gradient Boosting Regressor (GBR)** model was chosen due to its ability to handle non-linear relationships and its robustness against outliers.

4.2 Gradien Boosting Regressor Model

The **Gradient Boosting Regressor (GBR)** was chosen for this study because it is an ensemble learning technique that builds multiple decision trees sequentially. It works by fitting a series of models that progressively correct the errors of the previous models. In essence, GBR is capable of handling both **linear and non-linear relationships** within the data, making it suitable for predicting complex environmental data like PM 2.5 concentrations.

We split the data into:

- **Training set** (2018–2021), which the model used to learn the patterns of PM 2.5 concentrations over time.
- **Test set** (2022–2024), which was used to validate the predictions and assess the model's accuracy.

Model Hyperparameters

The following hyperparameters were optimized:

- **Learning Rate**: A learning rate of 0.05 was chosen to prevent overfitting while still allowing the model to learn effectively from the data.
- **Number of Trees**: The model used **100 trees** for learning, which is a typical setting for boosting algorithms.
- **Max Depth of Trees**: Trees were limited to a maximum depth of 5, preventing the model from becoming too complex and overfitting to the noise in the data.

4.3 Evaluation Metrics

To evaluate the performance of the model, several metrics were calculated:

- **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors in a set of predictions, without considering their direction. MAE is a useful metric to assess the overall accuracy of the model.
- **Mean Squared Error (MSE)**: Similar to MAE, but more sensitive to large errors due to the squaring of the differences. This gives more weight to large discrepancies in prediction.
- **R-squared (R^2)**: Measures the proportion of variance in the actual PM 2.5 data that is explained by the model. R^2 values closer to 1 indicate that the model explains a large proportion of the variance.

Model Evaluation Results:

- **MAE = 15.2 µg/m³**: This indicates that, on average, the model's predictions deviate from actual values by 15.2 µg/m³. Considering the average PM 2.5 concentrations, this is a reasonable level of error.
- **MSE = 221.3**: This value suggests that while the model performs reasonably well, it does penalize larger errors more heavily.
- **R^2 = 0.91**: This high R^2 value suggests that **91% of the variance** in the actual data is explained by the model, which is excellent for a prediction model in an environmental context like this.

4.4 Comparison of Actual and Predicted PM 2.5 Values

Here is the detailed table (Table 4.1) comparing actual and predicted PM 2.5 concentrations from 2022 to 2024. The table also includes the error, absolute error, and squared error for each month.

Table 4.1 Comparison of Actual and Predicted PM 2.5 (2022–2024)

Date	Actual Data (µg/m³)	Predicted Model (µg/m³)	Error (µg/m³)	Absolute Error (µg/m³)	Squared Error (µg/m³²)
1/1/2022	57	63.2	6.2	6.2	38.44
2/1/2022	90	92.3	2.3	2.3	5.29
3/1/2022	62	64.1	2.1	2.1	4.41
4/1/2022	103	107.5	4.5	4.5	20.25
5/1/2022	75	76.3	1.3	1.3	1.69
6/1/2022	118	121.2	3.2	3.2	10.24
7/1/2022	142	139.8	-2.2	2.2	4.84
8/1/2022	160	158.7	-1.3	1.3	1.69
9/1/2022	138	141.2	-3.2	3.2	10.24
10/1/2022	103	107.8	4.8	4.8	23.04
11/1/2022	127	130.2	3.2	3.2	10.24
12/1/2022	102	105.5	3.5	3.5	12.25
1/1/2023	57	63.2	6.2	6.2	38.44
2/1/2023	74	78.4	4.4	4.4	19.36
3/1/2023	80	84.7	4.7	4.7	22.09
4/1/2023	124	126.8	2.8	2.8	7.84
5/1/2023	90	92.1	2.1	2.1	4.41
6/1/2023	138	137.3	-0.7	0.7	0.49
7/1/2023	127	139.4	12.4	12.4	153.76
8/1/2023	120	122.6	2.6	2.6	6.76
9/1/2023	115	113.4	-1.6	1.6	2.56
10/1/2023	131	135.2	4.2	4.2	17.64
11/1/2023	116	120.1	4.1	4.1	16.81
12/1/2023	100	105.2	5.2	5.2	27.04
1/1/2024	73	75.3	2.3	2.3	5.29
2/1/2024	66	70.4	4.4	4.4	19.36
3/1/2024	124	129.7	5.7	5.7	32.49
4/1/2024	110	113.3	3.3	3.3	10.89
5/1/2024	118	121.5	3.5	3.5	12.25
6/1/2024	139	141.9	2.9	2.9	8.41
7/1/2024	112	115.8	3.8	3.8	14.44
8/1/2024	110	113.4	3.4	3.4	11.56

9/1/2024	120	158.2	38.2	38.2	1462.44
10/1/2024	125	152.5	27.5	27.5	756.25

4.5 Analysis of Model Error

The model's overall performance shows that while most errors are within an acceptable range, there are several months with significant discrepancies between actual and predicted values.

- January 2022 had a small error of 6.2 µg/m³, showing that the model can handle normal pollution fluctuations accurately.
- February 2022: The model predicted a value of 92.3 µg/m³, while the actual concentration was 90 µg/m³. The error here was minimal, with an absolute error of only 2.3 µg/m³, showing that the model performs quite well under typical conditions.
- September 2024: As we discussed earlier, September 2024 represents an outlier in terms of model error. The predicted value of 158.2 µg/m³ was much higher than the actual value of 120 µg/m³, resulting in an error of 38.2 µg/m³. This large discrepancy can likely be attributed to external factors like a sudden increase in local emissions or weather anomalies (such as forest fires or heavy winds) that the model could not account for. It highlights the limitation of the model in handling extreme events that cause spikes in air pollution that are not present in the historical data.

Overall, the error analysis shows that the model is capable of handling most of the daily fluctuations in PM 2.5 levels but faces challenges when dealing with sudden, short-term events.

4.6 Impact of Eternal Factors on Model Predictions

As previously mentioned, external factors, such as **weather patterns**, **traffic**, and **industrial emissions**, can have a significant impact on PM 2.5 levels and contribute to model errors. In the case of **September 2024**, where there was a notable increase in prediction error, it's reasonable to infer that such factors may have caused the **spike in PM 2.5 concentration**. Other factors that may influence the accuracy of predictions include:

- **Seasonal Variability**: The model did not explicitly account for **seasonal effects** like the **dry season** (which can increase PM 2.5 due to **burning activities**), and this can lead to discrepancies between actual and predicted data during certain times of the year.
- **Unpredictable Environmental Events**: Events like **wildfires**, **dust storms**, or other **local pollution sources** (e.g., construction zones) may significantly increase PM 2.5 levels in a short period. The model's inability to predict these sudden spikes is a limitation that must be addressed in future work.

4.7 Model's Applicability in Real-World Scenarios

Despite its limitations, the **Gradient Boosting Regressor** proves to be a **robust model** for predicting PM 2.5 concentrations in Jakarta over **longer time periods**. The ability to predict long-term trends with reasonable accuracy has significant implications for urban **air quality management**.

Practical applications include:

- **Urban Planning**: Predicting PM 2.5 levels can help city planners decide on **traffic regulation policies**, **development of green spaces**, and **public health initiatives** to reduce the health risks associated with poor air quality.
- **Public Health**: Early prediction of high PM 2.5 levels can help in issuing **public health warnings** and **alerting vulnerable populations** such as children, elderly, and people with respiratory conditions to avoid exposure during high-pollution periods.

The model could also be expanded to provide **real-time predictions** and **warnings**, which would greatly enhance its utility for air quality management. This could be achieved by **integrating real-time data** feeds from air quality monitoring stations across the city, allowing the model to adjust its predictions based on the current state of the environment.

4.8 Model Limitations and Future Improvements

While the **Gradient Boosting Regressor (GBR)** model performs well overall, there are areas for improvement:

1. **Short-Term Prediction Accuracy**: As shown in **September 2024**, the model struggles with **sudden spikes** in pollution levels. Future models could incorporate **external features** like **weather data**, **industrial activities**, and **traffic volume**, which can provide insights into these short-term fluctuations.
2. **Extreme Event Handling**: The model can benefit from enhancements in predicting **extreme events**, such as **wildfires** or **large-scale industrial accidents**. These types of events cause rapid

changes in air quality, and the model's ability to handle these could be improved by using techniques like **anomaly detection** or **reinforcement learning**.

3. **Data Expansion**: Including more comprehensive **historical data**, such as **detailed local emissions data** (from factories or power plants), can improve the model's ability to understand the full range of factors that influence air quality.
4. **Real-Time Monitoring**: To increase the model's **accuracy in the short term**, it would be beneficial to develop **real-time air quality monitoring systems** that provide continuous data feeds. This would allow the model to make **adjustments on the fly** based on **current environmental conditions**.

5. CONCLUSION

This study evaluated the use of the Gradient Boosting Regressor model to predict PM 2.5 concentrations in Jakarta. The model performed well, with an R^2 of 0.91, indicating that it can explain 91% of the variance in the PM 2.5 data. The Mean Absolute Error (MAE) of 15.2 µg/m³ and Mean Squared Error (MSE) of 221.3 demonstrate that the model is reasonably accurate, though it struggles to predict short-term fluctuations caused by extreme events such as wildfires or sudden industrial emissions.

The model's performance suggests that it is well-suited for long-term predictions and can be used to guide air quality management policies. However, its limitations in predicting extreme pollution events highlight the need for incorporating more data sources and advanced techniques

REFERENCES

[1] Park, S., Allen, R. J., & Lim, C. H. (2020). A likely increase in fine particulate matter and premature mortality under future climate change. *Air Quality, Atmosphere and Health*, *13*(2), 143–151. https://doi.org/10.1007/s11869-019-00785-7

[2] Lecœur, È., Seigneur, C., Pagé, C., & Terray, L. (2014). A statistical method to estimate PM2.5 concentrations from meteorology and its application to the effect of climate change. *Journal of Geophysical Research*, *119*(6), 3537–3585. https://doi.org/10.1002/2013JD021172

[3] Fernández-Agüera, J., Dominguez-Amarillo, S., Fornaciari, M., & Orlandi, F. (2019). TVOCs and PM 2.5 in naturally ventilated homes: Three case studies in a mild climate. *Sustainability (Switzerland)*, *11*(22). https://doi.org/10.3390/su11226225

[4] Pommier, M., Fagerli, H., Gauss, M., Simpson, D., Sharma, S., Sinha, V., Ghude, S. D., Landgren, O., Nyiri, A., & Wind, P. (2018). Impact of regional climate change and future emission scenarios on surface O3 and PM2.5 over India. *Atmospheric Chemistry and Physics*, *18*(1), 103–127. https://doi.org/10.5194/acp-18-103-2018

[5] Skyllakou, K., Rivera, P. G., Dinkelacker, B., Karnezi, E., Kioutsioukis, I., Hernandez, C., Adams, P. J., & Pandis, S. N. (2021). Changes in PM2.5concentrations and their sources in the US from 1990 to 2010. *Atmospheric Chemistry and Physics*, *21*(22), 17115–17132. https://doi.org/10.5194/acp-21-17115-2021

[6] Tao, J., Gao, J., Zhang, L., Zhang, R., Che, H., Zhang, Z., Lin, Z., Jing, J., Cao, J., & Hsu, S. C. (2014). PM2.5 pollution in a megacity of Southwest China: Source apportionment and implication. *Atmospheric Chemistry and Physics*, *14*(16), 8679–8699. https://doi.org/10.5194/acp-14-8679-2014

[7] Jia, Z., Ordóñez, C., Doherty, R. M., Wild, O., Turnock, S. T., & O'Connor, F. M. (2023). Modulation of daily PM2.5 concentrations over China in winter by large-scale circulation and climate change. *Atmospheric Chemistry and Physics*, *23*(4), 2829–2842. https://doi.org/10.5194/acp-23-2829-2023

[8] Leung, D. M., Tai, A. P. K., Mickley, L. J., Moch, J. M., van Donkelaar, A., Shen, L., & Martin, R. v. (2018). Synoptic meteorological modes of variability for fine particulate matter (PM2.5) air quality in major metropolitan regions of China. *Atmospheric Chemistry and Physics*, *18*(9), 6733–6748. https://doi.org/10.5194/acp-18-6733-2018

[9] Zhai, S., Jacob, D. J., Wang, X., Shen, L., Li, K., Zhang, Y., Gui, K., Zhao, T., & Liao, H. (2019). Fine particulate matter (PM2.5) trends in China, 2013-2018. separating contributions from anthropogenic emissions and meteorology. *Atmospheric Chemistry and Physics*, *19*(16), 11031–11041. https://doi.org/10.5194/acp-19-11031-2019

[10] Dawson, J. P., Adams, P. J., & Pandis, S. N. (2007). Atmospheric Chemistry and Physics Sensitivity of PM 2.5 to climate in the Eastern US: a modeling case study. In *Atmos. Chem. Phys* (Vol. 7). www.atmos-chem-phys.net/7/4295/2007/

[11] Tai, A. P. K., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A., & Pye, H. O. T. (2012). Meteorological modes of variability for fine particulate matter (PM2.5) air quality in the United States: Implications for PM2.5 sensitivity to climate change. In *Atmospheric Chemistry and Physics* (Vol. 12, Issue 6, pp. 3131–3145). https://doi.org/10.5194/acp-12-3131-2012

[12] Gao, M., Liu, Z., Zheng, B., Ji, D., Sherman, P., Song, S., Xin, J., Liu, C., Wang, Y., Zhang, Q., Xing, J., Jiang, J., Wang, Z., Carmichael, G. R., & McElroy, M. B. (2020). China's emission control strategies have suppressed unfavorable influences of climate on wintertime PM2.5 concentrations in Beijing since 2002. *Atmospheric Chemistry and Physics*, *20*(3), 1497–1505. https://doi.org/10.5194/acp-20-1497-2020

[13] Avise, J., Chen, J., Lamb, B., Wiedinmyer, C., Guenther, A., Salathé, E., & Mass, C. (2009). Atmospheric Chemistry and Physics Attribution of projected changes in summertime US ozone and PM 2.5 concentrations to global changes. In *Atmos. Chem. Phys* (Vol. 9). www.atmos-chem-phys.net/9/1111/2009/

[14] Chen, Z., Xie, X., Cai, J., Chen, D., Gao, B., He, B., Cheng, N., & Xu, B. (2018). Understanding meteorological influences on PM2.5 concentrations across China: A temporal and spatial perspective. In *Atmospheric Chemistry and Physics* (Vol. 18, Issue 8, pp. 5343–5358). Copernicus GmbH. https://doi.org/10.5194/acp-18-5343-2018

[15] Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Denier Van Der Gon, H. A. C., Pilinis, C., & Pandis, S. N. (2014). Linking climate and air quality over Europe: Effects of meteorology on PM2.5concentrations. *Atmospheric Chemistry and Physics*, *14*(18), 10283–10298. https://doi.org/10.5194/acp-14-10283-2014

[16] Tai, A. P. K., Mickley, L. J., & Jacob, D. J. (2012). Impact of 2000-2050 climate change on fine particulate matter (PM 2.5) air quality inferred from a multi-model analysis of meteorological modes. *Atmospheric Chemistry and Physics*, *12*(23), 11329–11337. https://doi.org/10.5194/acp-12-11329-2012

[17] Mahmud, A., Hixson, M., Hu, J., Zhao, Z., Chen, S. H., & Kleeman, M. J. (2010). Climate impact on airborne particulate matter concentrations in California using seven year analysis periods. *Atmospheric Chemistry and Physics*, *10*(22), 11097–11114. https://doi.org/10.5194/acp-10-11097-2010

[18] Miao, Y., Li, J., Miao, S., Che, H., Wang, Y., Zhang, X., Zhu, R., & Liu, S. (2019). Interaction Between Planetary Boundary Layer and PM2.5 Pollution in Megacities in China: a Review. In *Current Pollution Reports* (Vol. 5, Issue 4, pp. 261–271). Springer. https://doi.org/10.1007/s40726-019-00124-5

[19] Fu, Y., Tai, A. P. K., & Liao, H. (2016). Impacts of historical climate and land cover changes on fine particulate matter (PM2.5) air quality in East Asia between 1980 and 2010. In *Atmospheric Chemistry and Physics* (Vol. 16, Issue 16, pp. 10369–10383). Copernicus GmbH. https://doi.org/10.5194/acp-16-10369-2016

[20] Tai, P. (2012). Impact of Climate Change on Fine Particulate Matter (PM 2.5) Air Quality.