AIR QUALITY PREDICTION IN MOUNTAINOUS AREAS OF CHINA: BASED ON INTERPRETABLE MACHINE LEARNING MODEL

GUAN, T. Z.[1,2] – WANG, W. B.[1,2,3*] – CHEN, D. N.[1,2] – JIA, H. F.[1,2] – ZHANG, X. Z.[1,2]

[1]*Yunnan Academy of Ecological and Environmental Sciences, Kunming 650034, China*

[2]*Yunnan Key Laboratory for Pollution Processes and Control of Plateau Lake-Watersheds, Kunming 650034, China*

[3]*College of Soil and Water Conservation, Southwest Forestry University, Kunming 650224, China*

**Corresponding author*
e-mail: 774094115@qq.com

(Received 29th Oct 2025; accepted 11th Feb 2026)

Abstract. The spatial heterogeneity characteristic of mountainous regions poses substantial challenges to the accurate prediction of air quality. While machine learning has been widely applied in air quality prediction, there is a notable lack of research focusing on interpretable models in highland mountainous regions. This study proposes an interpretable machine learning-based approach to predict key air pollutants, focusing on air quality indicators—Fine Particulate Matter ($PM_{2.5}$) and Ozone (O_3)—in Yunnan Province, China. The main findings are as follows: (1) Air quality indicators in mountainous areas generally exhibit strong temporal autocorrelation. (2) The machine learning model exhibited reliable short- to medium-term forecasting ability, as evidenced by R^2 values above 0.5 and MAPE values below 32% for 5-day-ahead predictions of $PM_{2.5}$ and O_3. (3) Air quality in most areas shows higher sensitivity to meteorological factors, reflecting complex coupling relationships. In particular, low humidity during March and April emerged as a major factor contributing to air pollution. These findings provide valuable insights and practical tools for environmental management authorities, supporting the development and implementation of more effective environmental monitoring and early warning systems.
Keywords: *$PM_{2.5}$, O_3, meteorological factors, partial dependence plot (PDP), mountainous regions*

Introduction

The accurate prediction of air quality indicators is essential for informed decision-making in environmental management. Prediction models are generally classified as mechanistic and non-mechanistic approaches. Mechanistic simulation models such as the Community Multiscale Air Quality (CMAQ) model and the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) are commonly used (Gao and Zhou, 2024; Ryan, 2016). However, the implementation of mechanistic models for comprehensive air quality simulations remains challenging due to their complex structures and the large number of parameters involved. These models often require significant time and computational resources, which can hinder their practical application.

Machine learning models have been widely adopted for air quality simulation and prediction in recent years. For example: Ravindiran et al. (2025) employed various machine learning models—such as Extreme Gradient Boosting (XGBoost), Least Angle Regression with Cross-Validation (LarsCV), Bayesian Ridge, Adaptive Boosting (AdaBoost), and ensemble stacking techniques—for AQI prediction in India. Liao et al. (2023) integrated mechanistic modeling and machine learning by proposing a novel air quality prediction (AQP) method based on a Dynamic Multi-granularity Spatio-temporal

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

Graph Neural Network (DM_STGNN), which was applied to the Yangtze River Delta urban agglomeration. Xu et al. (2025) used Baidu search data from 290 cities across China (2011–2022) to investigate the spatiotemporal patterns of public concern over air pollution and their mismatch with observed pollution levels. Zeng et al. (2024b) utilized a data-driven two-step approach to assess the short- and long-term impacts of COVID-19 lockdowns on air pollution in China. Wei et al. (2025) developed a novel hybrid model combining the whale optimization algorithm (WOA), CNN, LSTM, and an attention mechanism (AM) to predict daily $PM_{2.5}$ concentrations in Beijing. Furthermore, Wang et al. constructed two models—a nonlinear autoregressive model with exogenous inputs (NARX) and a temporal convolutional network (TCN)—to forecast O_3 concentrations in the plateau region of Kunming (Wang et al., 2024). The extreme heterogeneity of highland mountainous environments has hindered the development of interpretable machine learning frameworks for air quality studies. Consequently, research integrating simulation with mechanistic insight in such regions remains notably limited.

Mountains, hills, and rugged highlands are collectively referred to as mountainous regions, which account for approximately two-thirds of China's total land area. Predicting air quality in these regions presents significant challenges due to the complex and spatially heterogeneous characteristics of meteorological, hydrological, topographic, and geomorphological conditions. This study focuses on Yunnan Province, China, and employs interpretable machine learning methods to develop predictive models for air environmental quality ($PM_{2.5}$ and O_3). The main contributions of this study are twofold: (1) evaluating the applicability of interpretable machine learning models in predicting comprehensive air quality indicators in mountainous areas, and (2) conducting an in-depth analysis of the mechanisms and spatial variability by which different meteorological factors influence future air quality in such regions using the PDP method. Our approach provides new insights into air quality forecasting and early warning in highland mountainous areas. Furthermore, it offers practical guidance for local environmental management authorities in pollution prevention and in the development of effective forecasting and early warning systems.

Materials and methods

Study area

Yunnan Province, located in southwest China, was selected as the case study area for several compelling reasons: (1) Typical topographical features. Yunnan is characterized by a highly diverse terrain that descends from the northwest to the southeast, classifying it as a mountainous plateau. Approximately 88.6% of the province's land area is mountainous (https://www.yn.gov.cn/). The elevation ranges from 77 to 5963 m, and the slope varies between 0 and 85.25 degrees. (2) Typical spatial heterogeneity of meteorology**.** Yunnan experiences a subtropical plateau monsoon climate with pronounced three-dimensional climatic characteristics. The region is susceptible to compound extreme weather events, such as concurrent heatwaves and droughts, particularly in its southwestern parts (Zeng et al., 2024a). Climatic variations in eastern China, including potential influences on Yunnan, are also modulated by large-scale atmospheric patterns and oceanic anomalies (Zhang et al., 2024b). Over the past five years, meteorological station data show an annual average temperature of 16.4 °C, with maximum and minimum values of 23.9 °C and 5.8 °C, respectively. Monthly temperatures range between 9.3 °C and 21.6 °C. The province receives an average annual

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

rainfall of 1148.4 mm, with recorded extremes ranging from 563.9 mm to 2323.5 mm. Monthly rainfall varies between 14.07 mm and 223.9 mm. The geographic location of the study area is illustrated schematically in *Figure 1*.

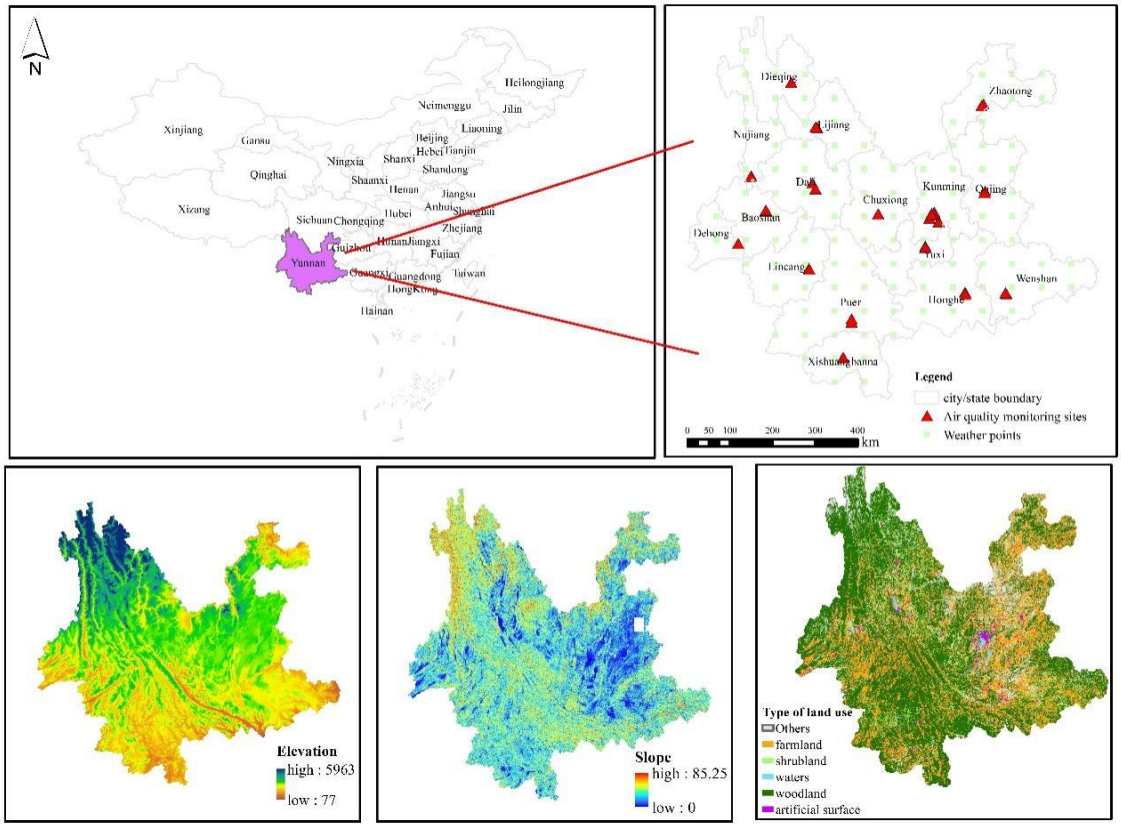

Figure 1. *Schematic map of the study area. The base map illustrates the topography of Yunnan Province, China. The red triangles represent the 34 air quality monitoring stations used in this study (see Table A1 for details)*

Data sources

Air quality data for O_3 and $PM_{2.5}$ were collected from 34 air monitoring stations (https://quotsoft.net/air/). All data, except for O_3, were averaged on a daily basis; O_3 concentrations were averaged over 8-h intervals. The dataset spans the period from 2021 to 2023, providing comprehensive day-by-day records for each pollutant. Meteorological data were sourced from the National Oceanic and Atmospheric Administration (NOAA) and include daily observations of rainfall, temperature, humidity, cloud cover, and wind speed for the same period (2021–2023). These meteorological data were preprocessed by the Wheat A team (https://wheata.cn/). To adequately reflect the spatial heterogeneity of meteorological conditions, data were extracted at a spatial resolution of 0.1° × 0.1°. Given the spatial heterogeneity of meteorological conditions in highland mountainous areas, we represented these conditions using the average values of all meteorological variables within a 50 km radius of each air quality monitoring site. The air monitoring stations are widely distributed across all municipal regions of Yunnan Province, ensuring representative coverage of the region's diverse geographical and climatic conditions.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

Method construction

(1) Impact factor assumptions

It is well established that air quality indicators also exhibit strong temporal patterns, which reflect both historical pollution levels and the emission dynamics of major point sources. These time-dependent characteristics are critical for enhancing prediction accuracy and must be integrated into any modeling framework. Additionally, the key factors influencing $PM_{2.5}$ include O_3, NO_2, CO, temperature, rainfall, wind speed, humidity, and cloud cover (Chen et al., 2020). Due to the physicochemical interactions between $PM_{2.5}$ and O_3, NO_2, and CO, the concentrations of these air pollutants from the previous day were included in the model. Furthermore, to account for the lagged effects of meteorological conditions (temperature, rainfall, wind speed, humidity, and cloud cover), data from the preceding five days were considered. A schematic overview of the technical methodology is presented in *Figure 2*.

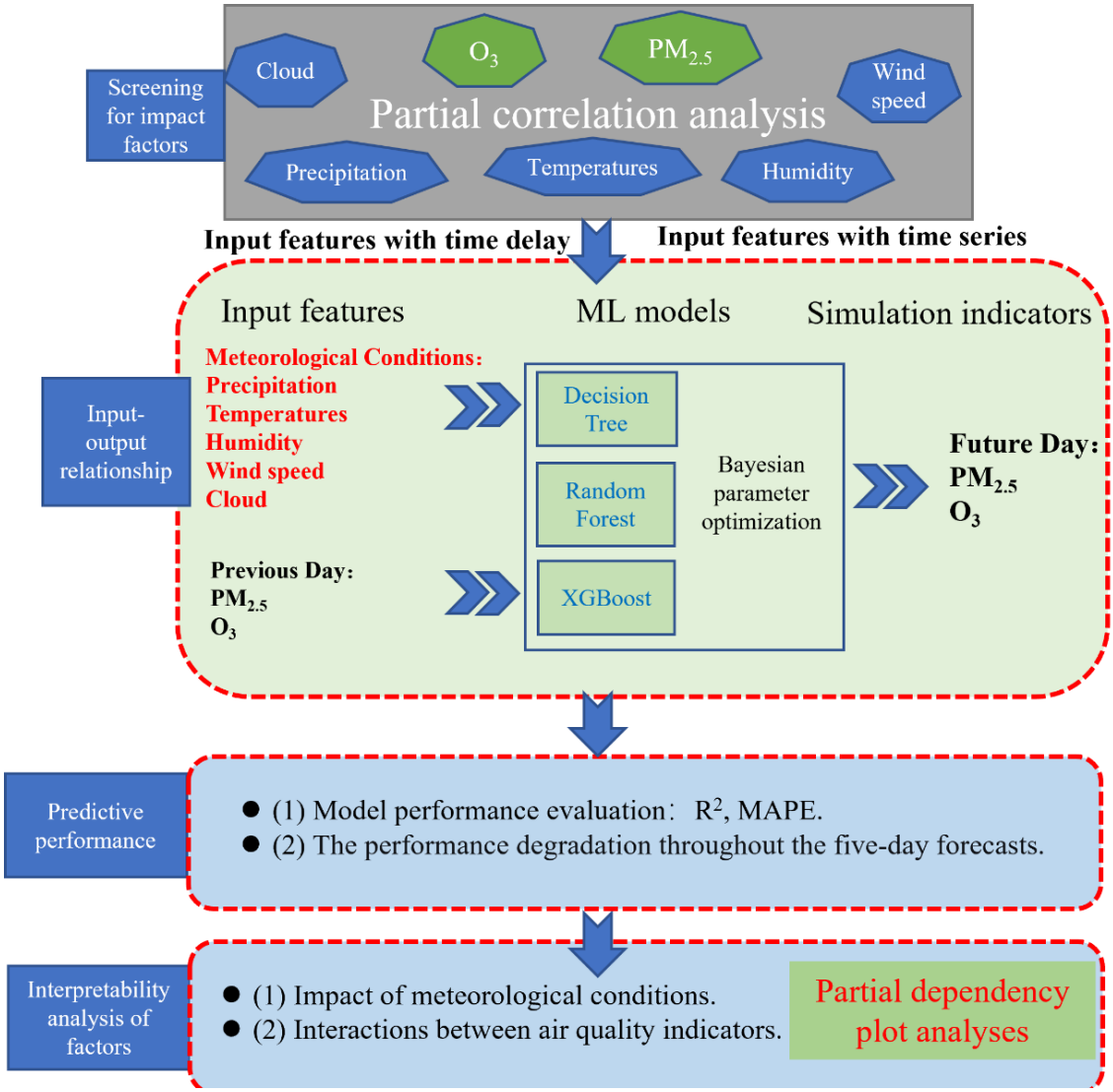

Figure 2. *Schematic diagram of the technical route of the study*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

(2) Screening for impact factors

Given the large number of variables influencing air quality, partial correlation analysis was employed to identify the most significant factors. This method enables the evaluation of the correlation between two variables while controlling for the influence of a third variable, thereby isolating the direct relationship between the remaining two. Using this approach, we identified key meteorological variables and their respective lag periods that significantly impact air quality indicators.

(3) Machine learning models

However, given the interactions among factors and the nonlinear relationships between driving factors and target variables ($PM_{2.5}$ and O_3), this study intends to employ machine learning to construct a nonlinear mapping model between them. This approach will be used to assess the prediction accuracy for the next five days and to quantify the nonlinear processes linking driving factors and target pollutants. To predict each air quality indicator—$PM_{2.5}$ and O_3—we applied three machine learning models: Extreme Gradient Boosting (XGBoost), Random Forest, and Decision Tree. The selection of these models was based on their complementary strengths: strong predictive performance (XGBoost), robustness and feature importance assessment (Random Forest), and high interpretability (Decision Tree). XGBoost is a scalable tree boosting algorithm that is widely adopted for its high efficiency and fast training capabilities (Bentejac et al., 2021). Random Forest, an ensemble learning algorithm based on the Bagging (Bootstrap Aggregating) strategy, is effective in addressing nonlinear problems, handling large datasets with numerous features, and has been widely used in fields such as genomic data analysis for uncovering complex relationships (Kolisnik et al., 2024). The Decision Tree method is a widely used rule-based approach in machine learning and stochastic decision-making, known for its interpretability and simplicity (Blockeel et al., 2023). Bayesian optimization has been extensively applied across various research domains to optimize model parameters by maximizing predictive performance (Kulshrestha et al., 2020; Mirzaee and Kamrava 2023; Tao et al., 2022; Zhang et al., 2023, 2021; Zhou et al., 2021). In our study, we developed a customized program to perform Bayesian optimization, using model fitting accuracy as the objective function to determine the optimal parameter configuration. To ensure robustness and mitigate the impact of a single random split, we employed a 10-fold cross-validation strategy during the model tuning phase. The final reported performance metrics are based on the held-out validation set (15% of data).

(4) Model interpretable analysis

To investigate the influence mechanisms of individual factors on air quality, we employed partial dependence plots (PDPs) based on the best-performing machine learning model. PDP is a global interpretability technique that illustrates the marginal effect of a single input variable on the predicted outcome, while averaging out the influence of all other variables (Greenwell, 2017; Szepannek and Luebke, 2023). For feature x_s the partial dependency function is:

$$f_s(x_s) = \frac{1}{n}\sum_{i=1}^{n} f(x_s, x_c^{(i)}) \quad \text{(Eq.1)}$$

where: x_s denotes the target feature; x_c denotes other features; n denotes the sample size; f denotes the prediction model we constructed.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

(5) Model performance evaluation

We assessed the training and validation performance of the machine learning model using two key evaluation metrics: the coefficient of determination (R^2) and the mean absolute percentage error (MAPE). R^2 quantifies how well the model captures the variability in the observed data, while MAPE reflects the relative prediction error of the model. The mathematical definitions of these metrics are given as follows:

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

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

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{(Eq.4)}$$

where y_i is the true value, $\hat{y}_i$ indicates the predicted value, and n signifies the number of samples.

Results and discussion

Spatial and temporal distribution of air quality

Spatially, the concentrations of $PM_{2.5}$ and O_3 ranged from 8.51-27.17 μg/m³ and 38.95-80.27 μg/m³, respectively (see *Fig. 3a* and *b*). Elevated levels of $PM_{2.5}$ and O_3 were observed primarily during the March–April period each year. During these months, the average concentrations of $PM_{2.5}$ and O_3 were 2.04 and 1.68 times higher, respectively, compared to the period outside March–April (see *Fig. 3c* and *d*). In summary, the selected air quality monitoring stations exhibited significant spatial and temporal variability. This heterogeneity improves the generalizability of the developed prediction models, ensuring their applicability under diverse environmental conditions and across various locations.

Screening of impact factors for environmental quality

Through partial correlation analysis, it was found that $PM_{2.5}$ and O_3 exhibit strong autocorrelation within a 5-day time lag, with correlation coefficients gradually decaying as the lag increases. Within a 5-day lag period, the partial correlation coefficients of $PM_{2.5}$ ranged from 0.27 to 0.59, while those of O_3 ranged from 0.25 to 0.63. Notably, a strong correlation coefficient of 0.472 was also observed between $PM_{2.5}$ and O_3 over the same 5-day lag period. Furthermore, both also show certain correlations with other pollution factors (CO, NO_2) and meteorological condition variables (temperature, precipitation, relative humidity, wind speed, cloud cover) within the same time window. Therefore, this study contends that historical pollutant concentrations along with meteorological conditions can be used to predict target pollutant concentrations over the next five days. Details are shown in *Figure 4*.

Predictive performance of air quality

As shown in *Figures 5* and *6*, the predictive performance (R^2) of the two machine learning models—Random Forest and XGBoost—was satisfactory, with no substantial

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

differences observed among them. Based on the validation set, the mean R^2 values and standard deviations for the first-day-ahead prediction of O_3 and $PM_{2.5}$ were 0.74 ± 0.08 and 0.74 ± 0.08, respectively. Corresponding MAPE values were $14.95 \pm 2.11\%$ and $22.86 \pm 4.54\%$, respectively. Of particular note is that the 5-day-ahead forecasts for $PM_{2.5}$ and O_3 consistently achieved $R^2 > 0.5$ and MAPE < 32%. These results suggest that the developed machine learning model is competent for short- to medium-term forecasting tasks.

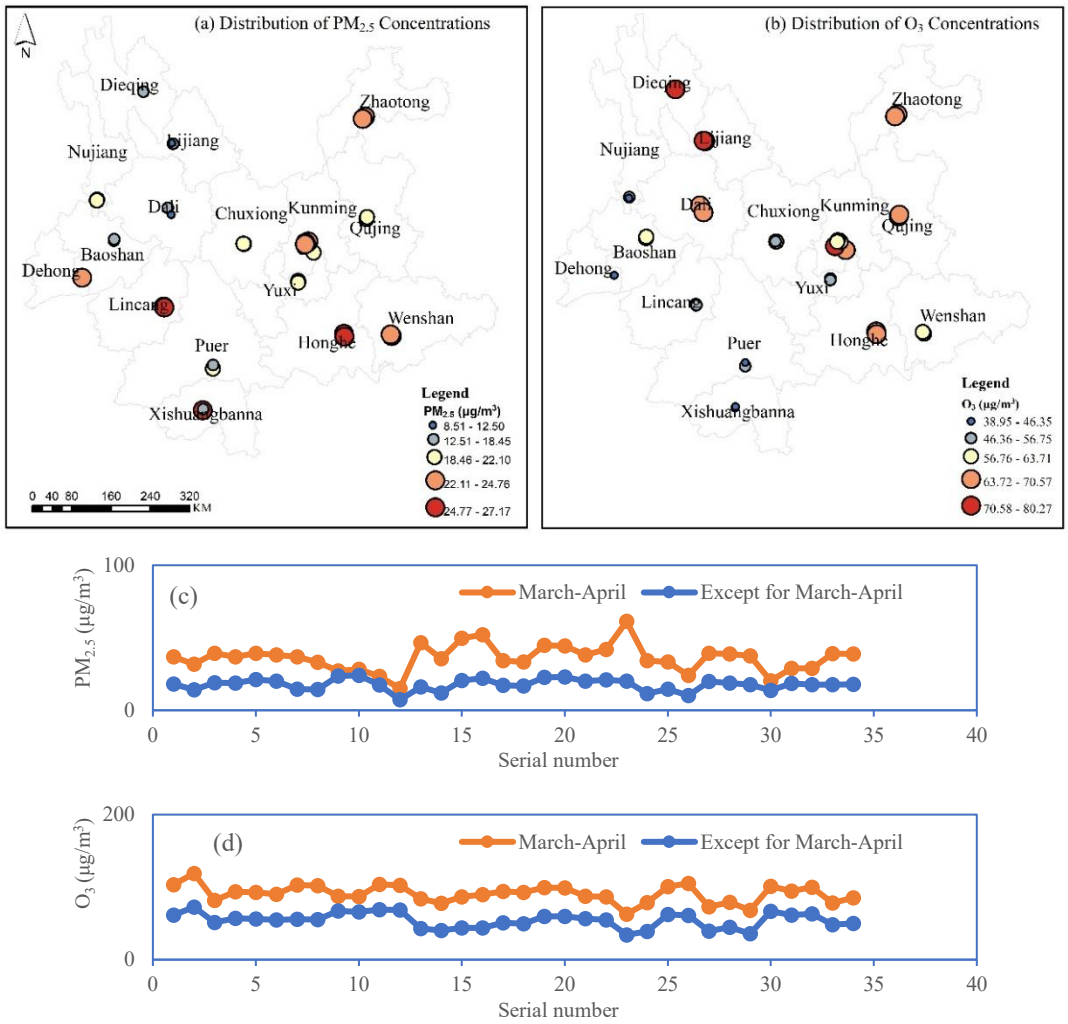

Figure 3. *Schematic diagram of pollutant concentrations ($PM_{2.5}$ and O_3) statistics. In (c) and (d): the series numbers represent the different stations, with a total of 34 for air quality*

When compared with existing studies, our results show competitive predictive accuracy. In the domain of air quality, Yang et al. (2020) proposed a hybrid model combining BPANN and SVR algorithms to predict PM2.5 concentrations in Yunnan, obtaining an R^2 of 0.8101 and a MAPE of 19.73%. Similarly, Zhao et al. (2019) employed the NARX model in Kunming, Yunnan Province, achieving MAPE values of 15.64% for PM2.5 and 36.33% for O_3, significantly outperforming the CMAQ model (MAPE: 90% for $PM_{2.5}$ and 13.67% for O_3) and the NAQPMS numerical model (MAPE: 74% and 41%, respectively) (Zhao et al., 2019). Our models achieved

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

comparable or superior MAPE for O_3 compared to Zhao et al. (14.95% vs. 36.33%), while for $PM_{2.5}$, our MAPE (22.86%) was higher than that reported by Zhao et al. (15.64%) but significantly lower than their CMAQ results (90%). This suggests that while our interpretable machine learning approach offers substantial advantages in understanding O_3 dynamics in complex terrain, there is room for improvement in $PM_{2.5}$ prediction, potentially by incorporating more localized emission sources or refining the spatial representation of meteorological fields.

Figure 4. *Partial correlation coefficients between air environmental quality indicators O_3, $PM_{2.5}$, and influencing factors. In the map, the mean partial correlation coefficient was calculated by averaging the values from all 34 stations, with a range spanning the minimum to maximum values observed across these stations. "PRE" represents rainfall, "TMP" represents temperature, 'HUM' represents relative humidity, 'WIND' stands for wind speed, 'CLOUD' represents the amount of cloud*

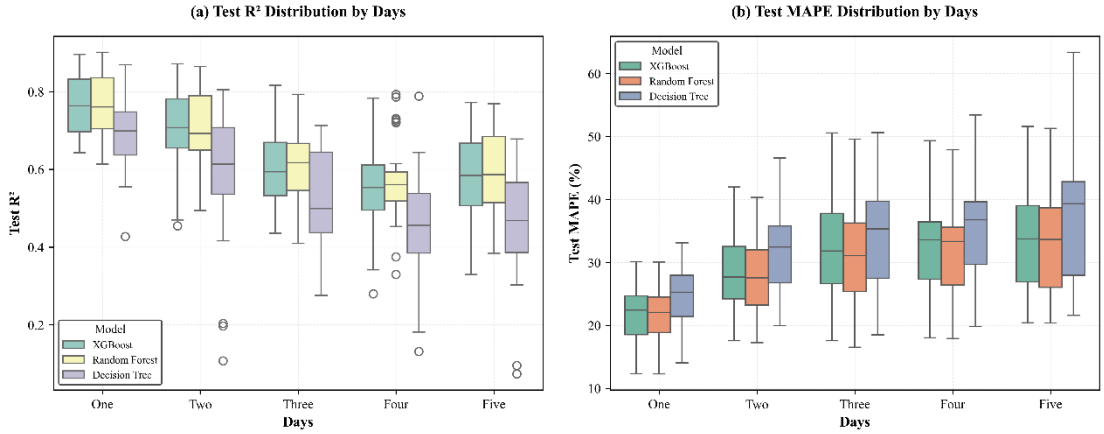

Figure 5. *Schematic diagram of the predicted performance of $PM_{2.5}$ for the validation period*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

Interpretability analysis of impact factors

Previous concentrations of air pollutants ($PM_{2.5}$ and O_3) exert a strong influence on current air quality conditions, indicating a high degree of temporal autocorrelation. Meteorological conditions have a more pronounced and complex effect on air quality, and the influencing mechanisms are more intricate. Notably, air quality indicators also exhibit complex interdependencies among themselves.

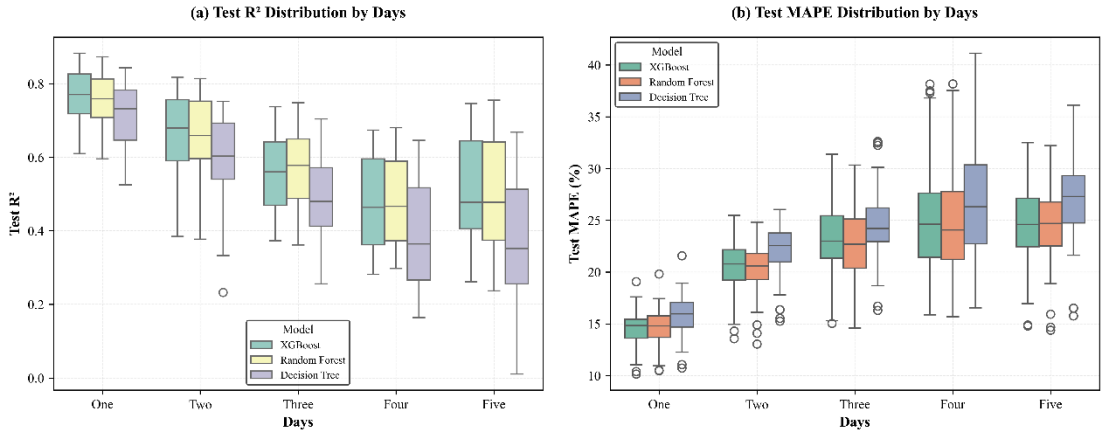

Figure 6. *Schematic diagram of the predicted performance of O_3 for the validation period*

(1) Impact of meteorological conditions

At several air quality monitoring sites, preceding meteorological variables—such as relative humidity, rainfall, air temperature, and wind speed—exerted significant negative effects on current $PM_{2.5}$ concentrations. For O_3, previous relative humidity generally exhibited a negative correlation, while temperature and wind speed tended to have a positive effect at most stations. Detailed patterns can be found in *Figures A1–S9*.

Relative humidity and precipitation. The dominant components of $PM_{2.5}$ in Yunnan are water-soluble inorganic ions, including $SO_4^{2-}, NO_3^-, and\ NH_4^+$ (Deng et al., 2020; Zhu et al., 2024). The observed negative correlation between prior relative humidity and current O_3 and $PM_{2.5}$ concentrations is primarily attributed to two mechanisms: wet deposition scavenging and photochemical inhibition. Wet deposition can directly reduce pollutant concentrations by removing $PM_{2.5}$ particles—particularly water-soluble components like sulfates and nitrates—as well as precursors of O_3, such as NO_x. During March-April, the average relative humidity in Yunnan is approximately 59.00%. Under these conditions, low humidity may not facilitate effective dry or wet deposition of $PM_{2.5}$. Instead, it may enhance particle hygroscopic growth and accumulation, potentially worsening air quality (Chen et al., 2020). As evidenced by case studies conducted in Xi'an and Shenyang, China (Hong et al., 2024; Yang et al., 2025), higher relative humidity is generally associated with lower ozone concentrations. For PM2.5, very high humidity promotes wet removal, leading to lower concentrations (e.g., in the Pearl River Delta; Weng et al., 2025).

Temperature. Elevated temperatures on preceding days create a more reactive atmospheric environment conducive to O_3 formation on the following day. Higher temperatures significantly accelerate key photochemical reactions involved in O_3 generation, such as the photolysis of NO_2 and the oxidative chain reactions between volatile organic compounds (VOCs) and hydroxyl (OH) radicals. Additionally, high

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

temperatures enhance the biogenic emission of VOCs from vegetation in Yunnan (Wu et al., 2020), thereby increasing the availability of O_3 precursors. Moreover, while elevated temperatures facilitate O_3 formation, they also consume atmospheric oxidants such as OH radicals, potentially suppressing the oxidation of SO_2 and NO_x into secondary $PM_{2.5}$ components like sulfate and nitrate. Since VOCs and NO_x are the primary precursors for photochemical O_3 pollution, higher emissions of these compounds—especially under high-temperature conditions—increase both the likelihood and severity of O_3 episodes. VOC emissions, largely from natural sources, are typically concentrated in low-latitude regions and are most active from April to September. This temperature-driven mechanism is consistent with observations in other regions (e.g., Zhang et al., 2022; Ni et al., 2024). However, the partial dependence plots in our study suggest that in Yunnan's mountainous terrain, the positive effect of temperature on O_3 may plateau or even slightly decrease beyond a certain threshold (~25°C), possibly due to enhanced convective mixing or changes in dominant VOC speciation, a nuance not commonly reported in studies focused on plains or urban agglomerations.

Wind. Higher wind speeds can promote the transport of O_3 or its precursors (e.g., VOCs, NO_x) from upwind regions. At the same time, increased wind speeds enhance the horizontal dispersion of $PM_{2.5}$, thereby reducing the accumulation of locally emitted pollutants such as vehicle exhaust and industrial particulate emissions.

(2) Interactions between air quality indicators

At certain monitoring sites, prior concentrations of CO and NO_2 were found to have a positive influence on current $PM_{2.5}$ levels (see *Figs. A10* and *A11*). However, their impact on current O_3 concentrations was relatively limited. In contrast, the interaction between $PM_{2.5}$ and O_3 was more pronounced. Elevated O_3 levels on the previous day can enhance the oxidation of VOCs, resulting in the formation of low-volatility products such as carbonyl compounds and organic acids. These compounds subsequently condense to form secondary organic aerosols (SOAs), thereby contributing directly to current $PM_{2.5}$ concentrations. Liao et al. (2019) illustrated the spatial distribution of aerosols in Yunnan Province using aerosol optical thickness data. Additionally, air masses previously enriched in $PM_{2.5}$ may carry O_3 precursors, such as NO_x and VOCs. These precursors can undergo regional transport and participate in local photochemical reactions on the following day, leading to elevated O_3 concentrations. This observed $PM_{2.5}$ and O_3 coupling in Yunnan's pre-monsoon season (March-April) is particularly strong. While Chen et al. (2023) reported widespread $PM_{2.5}$ and O_3 correlations across China, they noted positive correlations were dominant along the southern coast and in summer. Our study extends this understanding by demonstrating a significant positive coupling in the southwestern plateau during spring, potentially driven by shared meteorological drivers (low humidity, high temperature) and active photochemistry, rather than just co-transport. This insight is crucial for developing coordinated control strategies in similar mountainous environments**.** A schematic representation of the inferred interactions among meteorological variables, $PM_{2.5}$, and O_3 is provided in *Figure 7*.

During the March–April period, increased temperatures, low relative humidity, and prolonged sunshine duration contribute to the intensification of O_3 pollution (Zhang et al., 2022). Ni et al. (2024) identified temperature, relative humidity, and wind speed as the dominant meteorological drivers of O_3 variability, with significant spatial heterogeneity across regions. In Southeast Asia, temperature and humidity were shown

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

to be the two most critical factors (Ni et al., 2024). In Hangzhou, China, Zhang et al. (2024a) reported that when relative humidity drops below 70%, the positive contribution rate of temperature to O_3 pollution increases by 89%. Furthermore, Chen et al. (2023) revealed that since 2017, over 90% of monitoring sites have shown significant $PM_{2.5}$–O_3 correlations across all seasons, with positive correlations predominantly observed along the southern coast and during summer months.

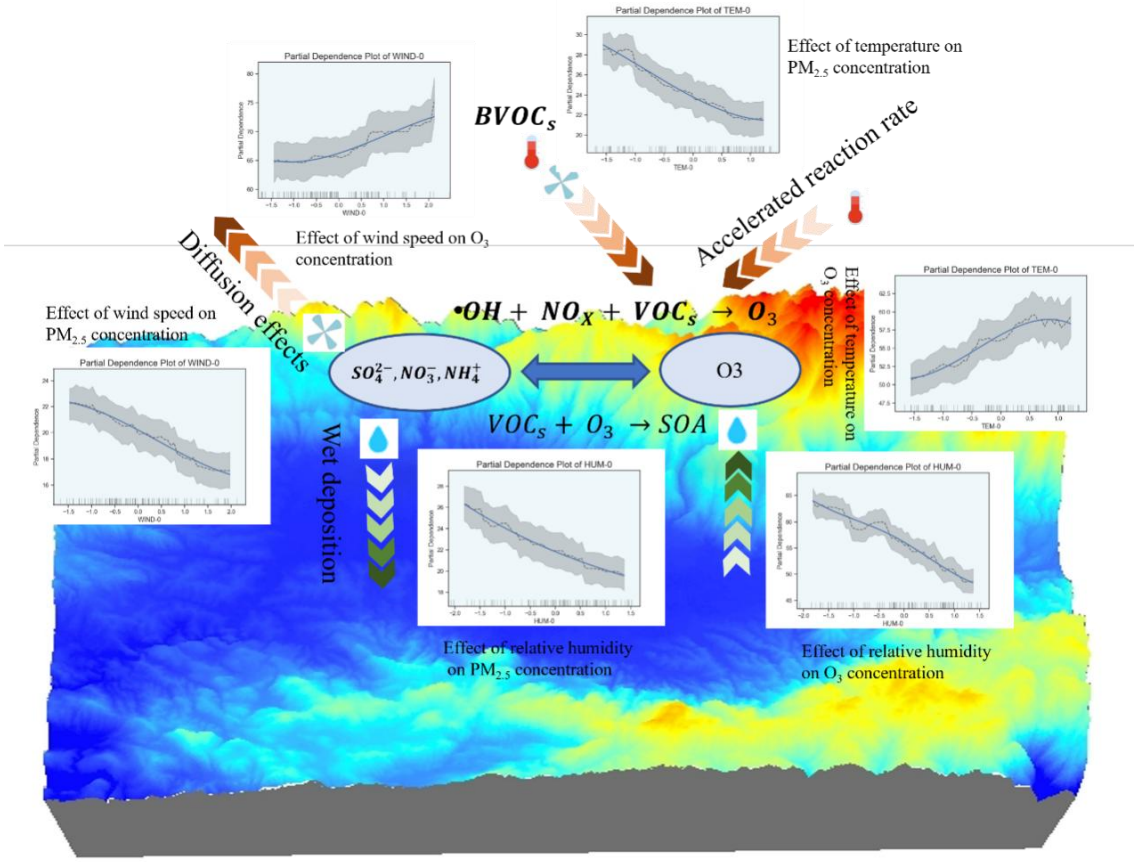

Figure 7. *Schematic diagram of the inferred relationship between meteorological factors, $PM_{2.5}$ and O_3 in Yunnan*

Further research

This study primarily focuses on the influence of meteorological conditions on future air quality in an effort to enhance prediction accuracy. During model development, we reduced the dimensionality of influencing factors to promote the practical applicability of the proposed method. The effects of solar radiation intensity and atmospheric boundary layer dynamics (e.g., mixed layer thickness) were not considered. Additionally, the potential impacts of exogenous pollution sources—particularly biomass burning in Southeast Asia—were not incorporated. These omissions may affect the accuracy of medium- to long-term air quality projections. Future studies should give greater attention to the transboundary effects of biomass burning and incorporate relevant parameters to improve forecast robustness.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

Conclusions

(1) The proposed interpretable framework, integrating time series features with meteorological variables, enables reliable 5-day-ahead air quality prediction, demonstrating its applicability in spatially complex mountainous regions. The two machine learning models (Random Forest, and XGBoost) effectively captured the nonlinear relationships between influencing factors and air quality indicators.

(2) The strong interdependence between $PM_{2.5}$ and O_3 emphasizes the importance of integrated air pollution control. Among meteorological factors, relative humidity plays a particularly important role from March to April. Low humidity conditions are not conducive to pollutant removal and tend to exacerbate $PM_{2.5}$ and O_3 pollution levels.

Acknowledgments. This study was supported by Innovative Team for Research on New Methods of Watershed Pollution Process Modelling of Yunnan Academy of Ecological and Environmental Sciences (No.2023-07CXTD) and Yunnan Fundamental Research Projects (No. 202501AT070436).

Conflict of interests. 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.

REFERENCES

[1] Bentejac, C., Csorgo, A., Martinez-Munoz, G. (2021): A comparative analysis of gradient boosting algorithms. – Artificial Intelligence Review 54(3): 1937-1967. http://dx.doi.org/10.1007/s10462-020-09896-5.

[2] Blockeel, H., Devos, L., Frenay, B., Nanfack, G., Nijssen, S. (2023): Decision trees: from efficient prediction to responsible AI. – Frontiers in Artificial Intelligence 6. http://dx.doi.org/10.3389/frai.2023.1124553.

[3] Chen, C., Gao, B., Xu, M., Liu, S., Zhu, D., Yang, J., Chen, Z. (2023): The spatiotemporal variation of PM2.5-O3 association and its influencing factors across China: dynamic *Simil-Hu* lines. – Science of the Total Environment 880. http://dx.doi.org/10.1016/j.scitotenv.2023.163346.

[4] Chen, Z., Chen, D., Zhao, C., Kwan, M.-p., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., Xu, B. (2020): Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: a review of methodology and mechanism. – Environment International 139. http://dx.doi.org/10.1016/j.envint.2020.105558.

[5] Deng, L., Han, X., Shi, Z., Yan, K., Li, J., Shi, J. (2020): Characteristics and source analysis of water soluble ions in atmospheric PM_(2.5) in low latitude plateau cities of Yunnan Province. – Environmental Chemistry 39(12): 3306-3317.

[6] Gao, Z., Zhou, X. (2024): A review of the CAMx, CMAQ, WRF-Chem and NAQPMS models: application, evaluation and uncertainty factors. – Environmental Pollution 343: 123183-Article No.: 123183. http://dx.doi.org/10.1016/j.envpol.2023.123183.

[7] Greenwell, B. M. (2017): pdp: An R Package for Constructing Partial Dependence Plots. – R Journal 9(1): 421-436.

[8] Hong, Y., Ma, Y., Su, C., Wang, Y., Ren, W., Wang, J., Wang, D., Xu, X. (2024): Analysis of the relationship between O_3 and $PM_{2.5}$ and the main controlling factors of pollution in Shenyang. – Environmental Science Research 37(3): 455-468. http://dx.doi.org/10.13198/j.issn.1001-6929.2024.01.04.

[9] Kolisnik, T., Keshavarz-Rahaghi, F., Purcell, R. V., Smith, A. N. H., Silander, O. K. (2024): pyRforest: a comprehensive R package for genomic data analysis featuring scikit-learn Random Forests in R. – Briefings in Functional Genomics 24. http://dx.doi.org/10.1093/bfgp/elae038.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

[10] Kulshrestha, A., Krishnaswamy, V., Sharma, M. (2020): Bayesian BILSTM approach for tourism demand forecasting. – Annals of Tourism Research 83. http://dx.doi.org/10.1016/j.annals.2020.102925.
[11] Liao, H., Yuan, L., Wu, M., Chen, H. (2023): Air quality prediction by integrating mechanism model and machine learning model. – Science of the Total Environment 899. http://dx.doi.org/10.1016/j.scitotenv.2023.165646.
[12] Liao, Y., Yang, F., Luo, Y., Zhao, T., Shang, Y., Zheng, X. (2019): Regional climatology of aerosol distribution over Yunnan-Guizhou Plateau. – Ecology and Environmental Sciences 28(2): 316-323.
[13] Mirzaee, H., Kamrava, S. (2023): Estimation of internal states in a Li-ion battery using BiLSTM with Bayesian hyperparameter optimization. – Journal of Energy Storage 74. http://dx.doi.org/10.1016/j.est.2023.109522.
[14] Ni, J., Jin, J., Wang, Y., Li, B. (2024): Surface ozone in global cities: a synthesis of basic features, exposure risk, and leading meteorological driving factors. – Geography and Sustainability 5(1): 64-76.
[15] Ravindiran, G., Karthick, K., Maria, A., Rajamanickam, S., Datta, D., Das, B., Shyamala, G., Hayder, G., Maria, A. (2025): Ensemble stacking of machine learning models for air quality prediction for Hyderabad city in India. – Iscience 28(2). http://dx.doi.org/10.1016/j.isci.2025.111894.
[16] Ryan, W. F. (2016): The air quality forecast rote: recent changes and future challenges. – Journal of the Air & Waste Management Association (1995) 66(6): 576-96. http://dx.doi.org/10.1080/10962247.2016.1151469.
[17] Szepannek, G., Luebke, K. (2023): How much do we see? On the explainability of partial dependence plots for credit risk scoring. – Argumenta Oeconomica 50(1): 137-+. http://dx.doi.org/10.15611/aoe.2023.1.07.
[18] Tao, Y., Sun, H., Cai, Y. (2022): Predictions of deep excavation responses considering model uncertainty: integrating BiLSTM neural networks with Bayesian updating. – International Journal of Geomechanics 22(1). http://dx.doi.org/10.1061/(asce)gm.1943-5622.0002245.
[19] Wang, Q., Liu, H., Li, Y., Li, W., Sun, D., Zhao, H., Tie, C., Gu, J., Zhao, Q. (2024): Predicting plateau atmospheric ozone concentrations by a machine learning approach: a case study of a typical city on the southwestern plateau of China. – Environmental Pollution 363. http://dx.doi.org/10.1016/j.envpol.2024.125071.
[20] Wei, Q., Zhang, H., Yang, J., Niu, B., Xu, Z. (2025): $PM_{2.5}$ concentration prediction using a whale optimization algorithm based hybrid deep learning model in Beijing, China. – Environmental Pollution 371. http://dx.doi.org/10.1016/j.envpol.2025.125953.
[21] Weng, L., Wang, P., Xiao, R., Bai, J., Zhong, J. (2025): Spatiotemporal distribution characteristics and influencing factors of $PM_{2.5}$ and O_3 in the Pearl River Delta urban agglomeration from 2000 to 2022. – Ecology and Environmental Sciences 34(2): 268-278. http://dx.doi.org/10.16258/j.cnki.1674-5906.2025.02.009.
[22] Wu, K., Yang, X., Chen, D., Gu, S., Lu, Y., Jiang, Q., Wang, K., Ou, Y., Qian, Y., Shao, P., Lu, S. (2020): Estimation of biogenic VOC emissions and their corresponding impact on ozone and secondary organic aerosol formation in China. – Atmospheric Research 231. http://dx.doi.org/10.1016/j.atmosres.2019.104656.
[23] Xu, G., Liu, H., Jia, C., Zhou, T., Shang, J., Zhang, X., Wang, Y., Wu, M. (2025): Spatiotemporal patterns and drivers of public concern about air pollution in China: leveraging online big data and interpretable machine learning. – Environmental Impact Assessment Review 114. http://dx.doi.org/10.1016/j.eiar.2025.107897.
[24] Yang, K., Teng, M., Luo, Y., Zhou, X., Zhang, M., Sun, W., Li, Q. (2020): Human activities and the natural environment have induced changes in the $PM_{2.5}$ concentrations in Yunnan Province, China, over the past 19 years. – Environmental Pollution 265. http://dx.doi.org/10.1016/j.envpol.2020.114878.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

[25] Yang, X., Wang, Q., Wang, L., Liu, F., Li, C., Xu, G. (2025): Meteorological influences and potential source areas analysis of O_3 pollution in Xi'an from 2014 to 2022. – Research of Environmental Science 38(3): 486-496. http://dx.doi.org/10.13198/j.issn.1001-6929.2024.11.14.

[26] Zeng, J. N., Li, H. X., Sun, B., Chen, H. P., Wang, H. J., Zhou, B. T., Duan, M. K. (2024a): Summertime compound heat wave and drought events in China: interregional and subseasonal characteristics, and the associated driving factors. – Environmental Research Letters 19(7). http://dx.doi.org/10.1088/1748-9326/ad5576.

[27] Zeng, W., Chen, X., Tang, K., Qin, Y. (2024b): Does COVID-19 lockdown matter for air pollution in the short and long run in China? A machine learning approach to policy evaluation. – Journal of Environmental Management 370. http://dx.doi.org/10.1016/j.jenvman.2024.122615.

[28] Zhang, D., Jin, X., Shi, P., Chew, X. (2023): Real-time load forecasting model for the smart grid using bayesian optimized CNN-BiLSTM. – Frontiers in Energy Research 11. http://dx.doi.org/10.3389/fenrg.2023.1193662.

[29] Zhang, H., Jiang, H., Gao, J., Li, H. (2022): Review on causes and influencing factors of O3 pollution in China. – Research of Environmental Sciences 35(12): 2657-2665. http://dx.doi.org/10.13198/j.issn.1001-6929.2022.09.01.

[30] Zhang, L., Wang, L., Ji, D., Xia, Z., Nan, P., Zhang, J., Li, K., Qi, B., Du, R., Sun, Y., Wang, Y., Hu, B. (2024a): Explainable ensemble machine learning revealing the effect of meteorology and sources on ozone formation in megacity Hangzhou, China. – Science of the Total Environment 922. http://dx.doi.org/10.1016/j.scitotenv.2024.171295.

[31] Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L. (2021): Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. – Geoscience Frontiers 12(1): 469-477. http://dx.doi.org/10.1016/j.gsf.2020.03.007.

[32] Zhang, Y., Zhou, W., Zhang, R. H. (2024b): Decadal changes in dry and wet heatwaves in eastern China: spatial patterns and risk assessment. – Advances in Atmospheric Sciences 41(10): 2011-2026. http://dx.doi.org/10.1007/s00376-024-3261-4.

[33] Zhao, Q., Qiu, F., Yang, J. (2019): Application of NARX neural network model in environmental air quality prediction in Kunming. – Environmental Monitoring in China 35(3): 42-48.

[34] Zhou, J., Qiu, Y., Zhu, S., Armaghani, D. J., Khandelwal, M., Mohamad, E. T. (2021): Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. – Underground Space 6(5): 506-515. http://dx.doi.org/10.1016/j.undsp.2020.05.008.

[35] Zhu, R., Cao, F., Fan, M. (2024): Chemical characteristics and source analysis of water-soluble inorganic in PM_(2.5) in Kunming. – Environmental Chemistry 43(6): 2005-2016.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

APPENDIX

Table A1. *List of air quality monitoring points*

Serial number	Geographic name	Name	City	Longitude	Latitude
1	Chenggongxinqu	1450A	Kunming	102.8204	24.8886
2	Xishansenlingognyuan	1451A	Kunming	102.6251	24.9608
3	Longquanzhen	1452A	Kunming	102.7252	25.0902
4	Dongfengdonglu	1453A	Kunming	102.7266	25.0375
5	Jindingshan	1454A	Kunming	102.6742	25.0669
6	Bijiguangchang	1455A	Kunming	102.6644	25.0401
7	Shihuanjingjiancezhan	2594A	Baoshan	99.1678	25.1081
8	Shihuanbaoju	2595A	Baoshan	99.1711	25.1328
9	Jiancezhan	2596A	Zhaotong	103.7667	27.3736
10	Huanbaoju	2597A	Zhaotong	103.7225	27.3361
11	Lijianggucheng	2599A	Lijiang	100.2497	26.8802
12	Shizhongxin	2600A	Lijiang	100.2203	26.8906
13	Shihuanbaoju	2601A	Puer	100.98	22.7633
14	Puerdierzhongxue	2602A	Puer	100.9817	22.8322
15	Shihuanbaoju	2603A	Lincang	100.0781	23.8982
16	Shiqixiangju	2604A	Lincang	100.0894	23.8822
17	Zhouhuanjingjiancezhan	2605A	Chuxiong	101.5482	25.0441
18	Shijingjikaifaqu	2606A	Chuxiong	101.538	25.0492
19	Wushuichulichang	2608A	Honghe	103.3772	23.3993
20	Jiancezhan	2609A	Honghe	103.3861	23.35
21	Zhoushuiwuju	2610A	Wenshan	104.2533	23.3594
22	Shibianminfuwuzhognxin	2611A	Wenshan	104.2319	23.3892
23	Jinghongshijiangnan	2612A	Xishuangbanna	100.7939	22.0019
24	Jinghongshijiangbei	2613A	Xishuangbanna	100.8017	22.0225
25	Dalishihuanjingjiancezhan	2614A	Dali	100.1542	25.7054
26	Daligucheng	2615A	Dali	100.2171	25.5811
27	Dehongzhoujiancezhan	2616A	Dehong	98.5842	24.428
28	Zhoujiancezhan	2618A	Nujianng	98.8601	25.8567
29	Lushuiyizhong	2619A	Nujianng	98.8546	25.8417
30	Diqingzhouzhan	2621A	Diqing	99.7056	27.8317
31	Huanjingjiancezhan	1916A	Qujing	103.7897	25.5035
32	Yanchangbangongqu	1917A	Qujing	103.8	25.5364
33	Yuxiyizhong	2882A	Yuxi	102.5389	24.3703
34	Wentizhongxin	2883A	Yuxi	102.5381	24.3389

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

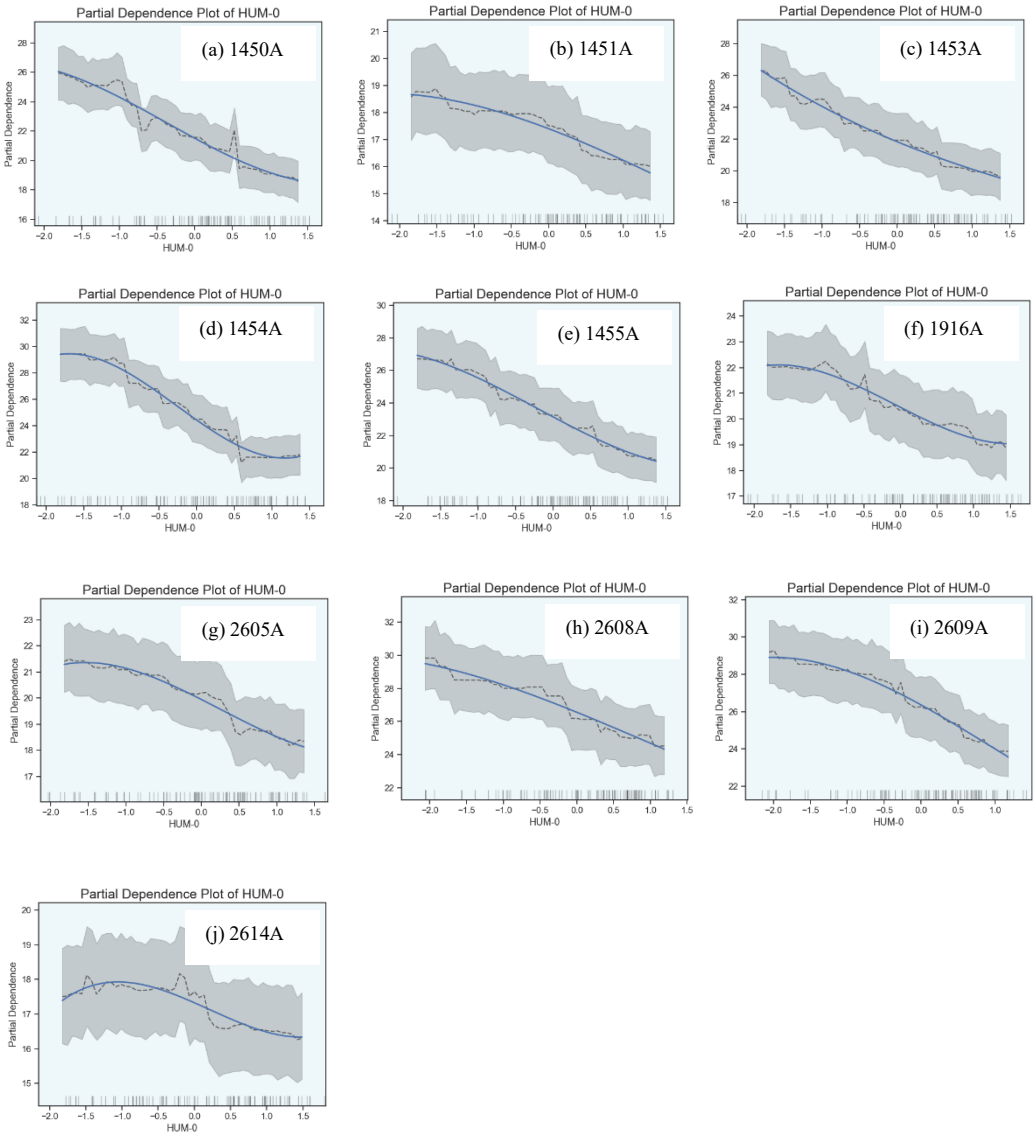

Figure A1. *Effect of HUM-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

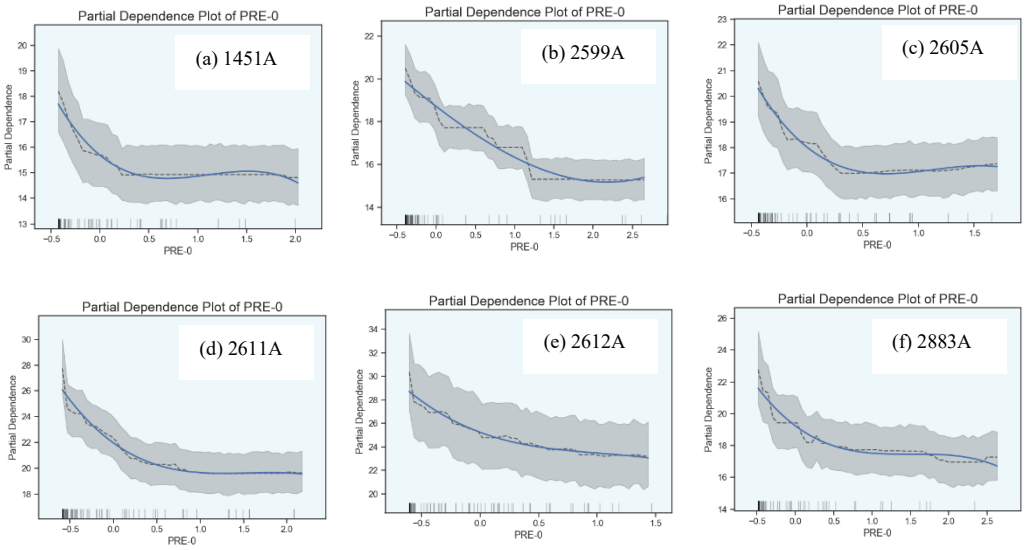

Figure A2. *Effect of PCP-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

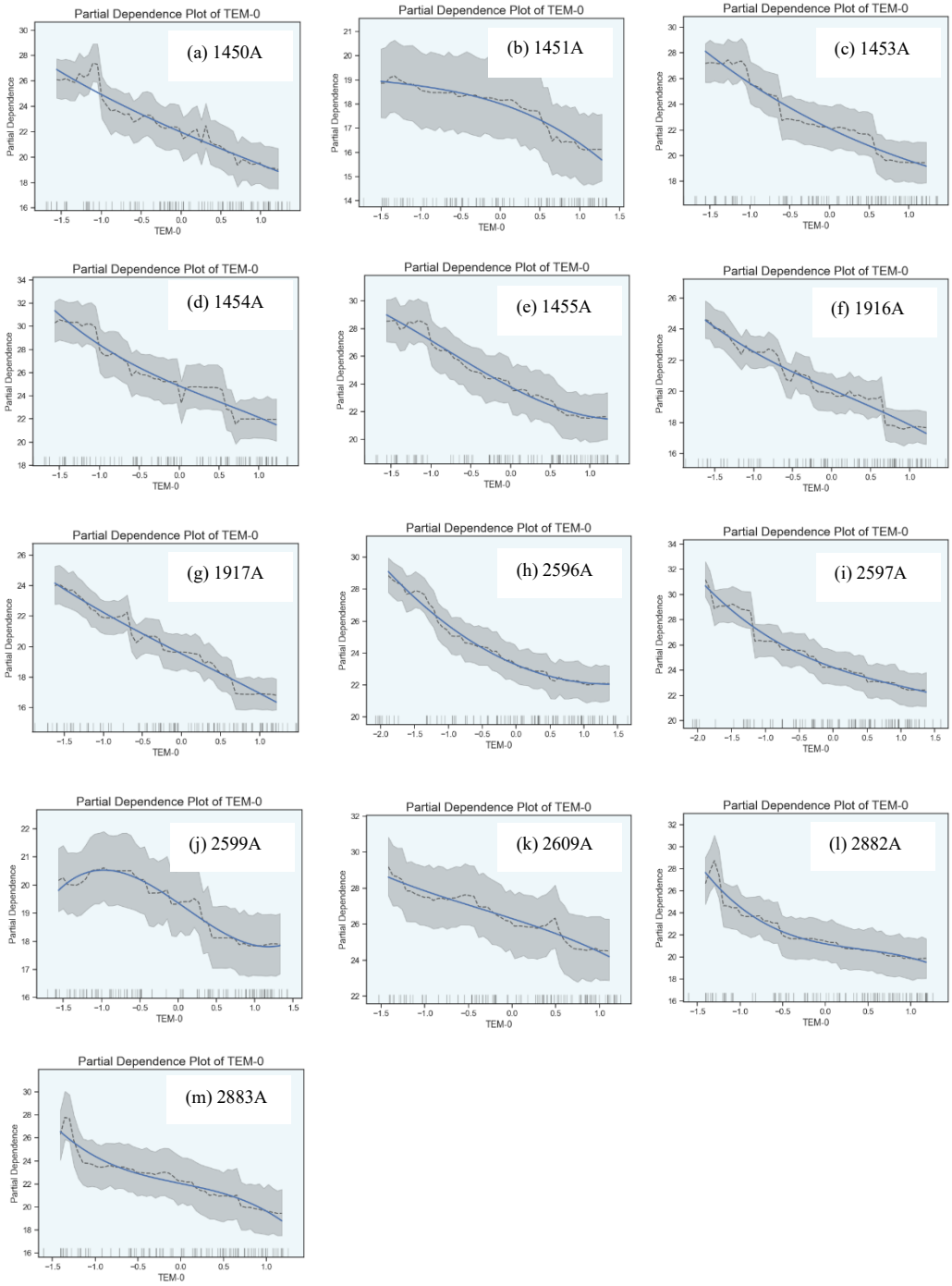

Figure A3. *Effect of TEM-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

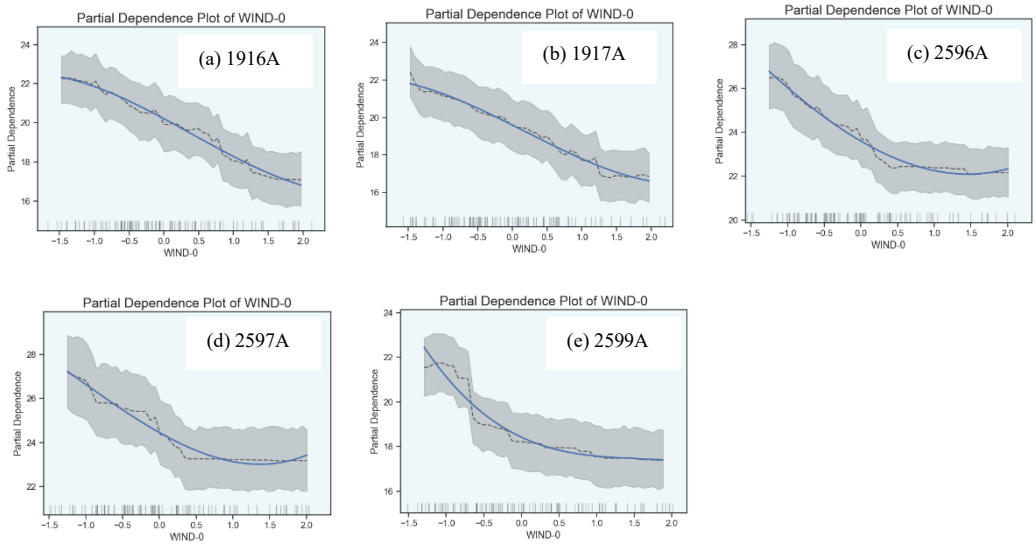

Figure A4. *Effect of WIND-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

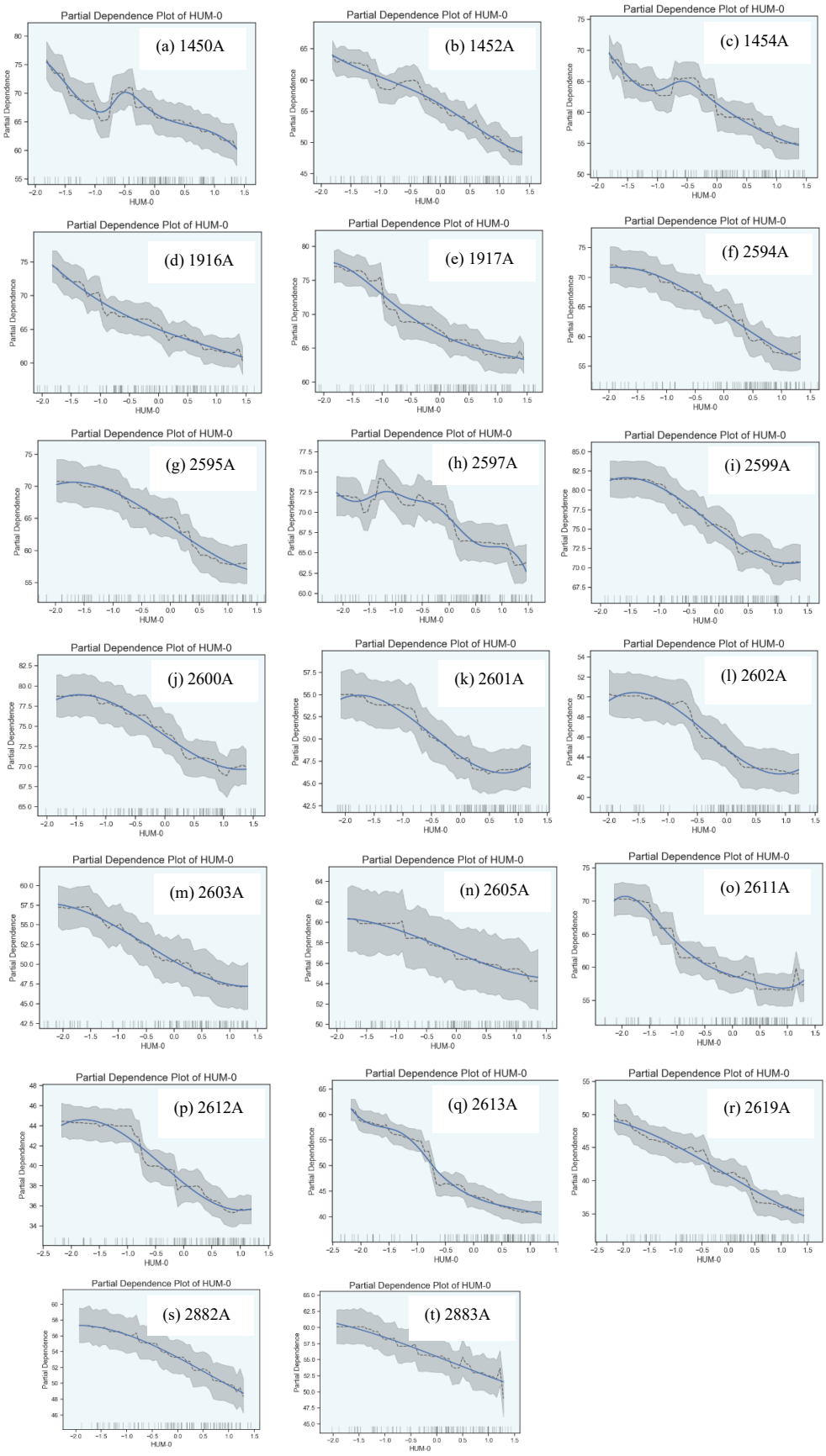

Figure A5. *Effect of HUM-0 on current O_3 concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

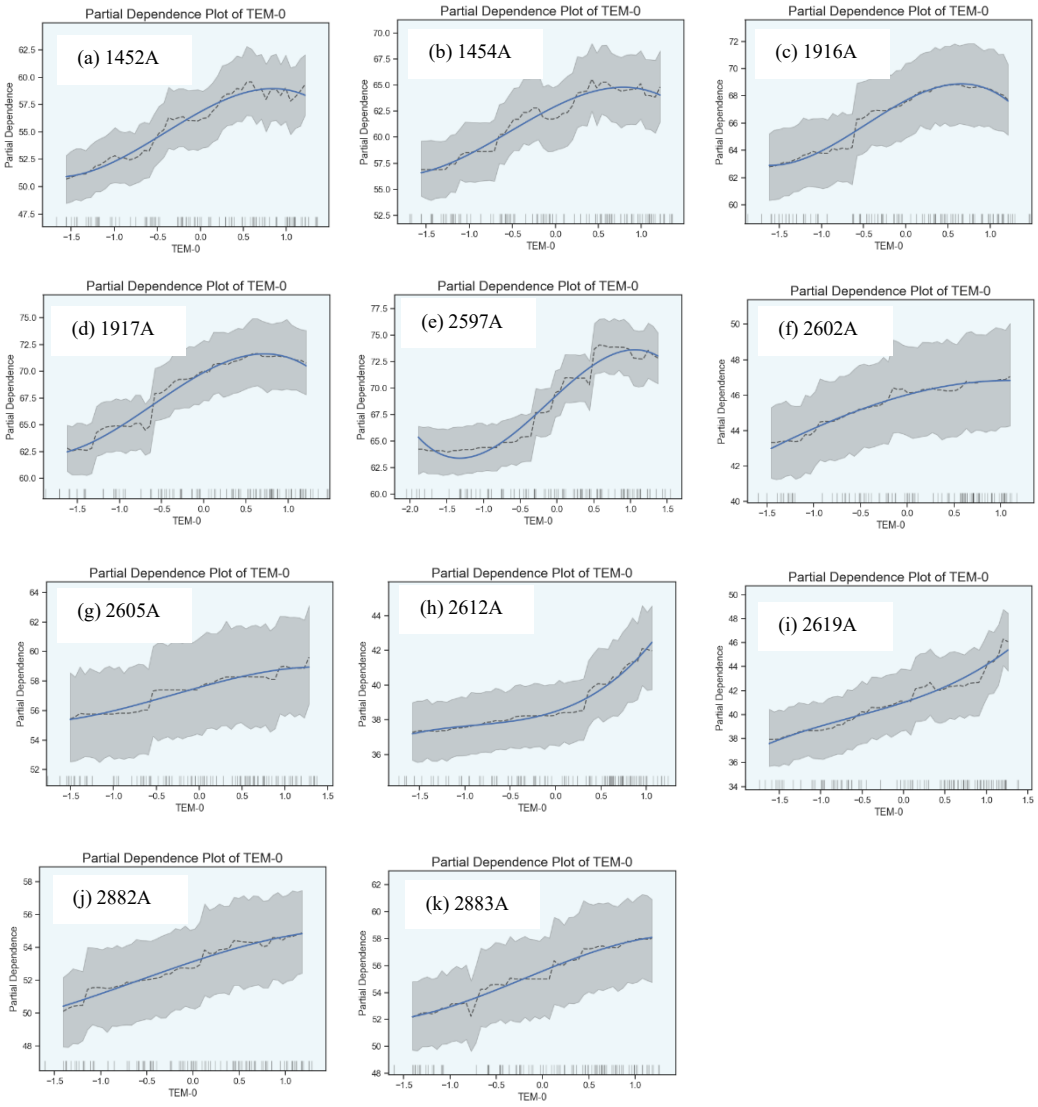

Figure A6. *Effect of TEM-0 on current O_3 concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

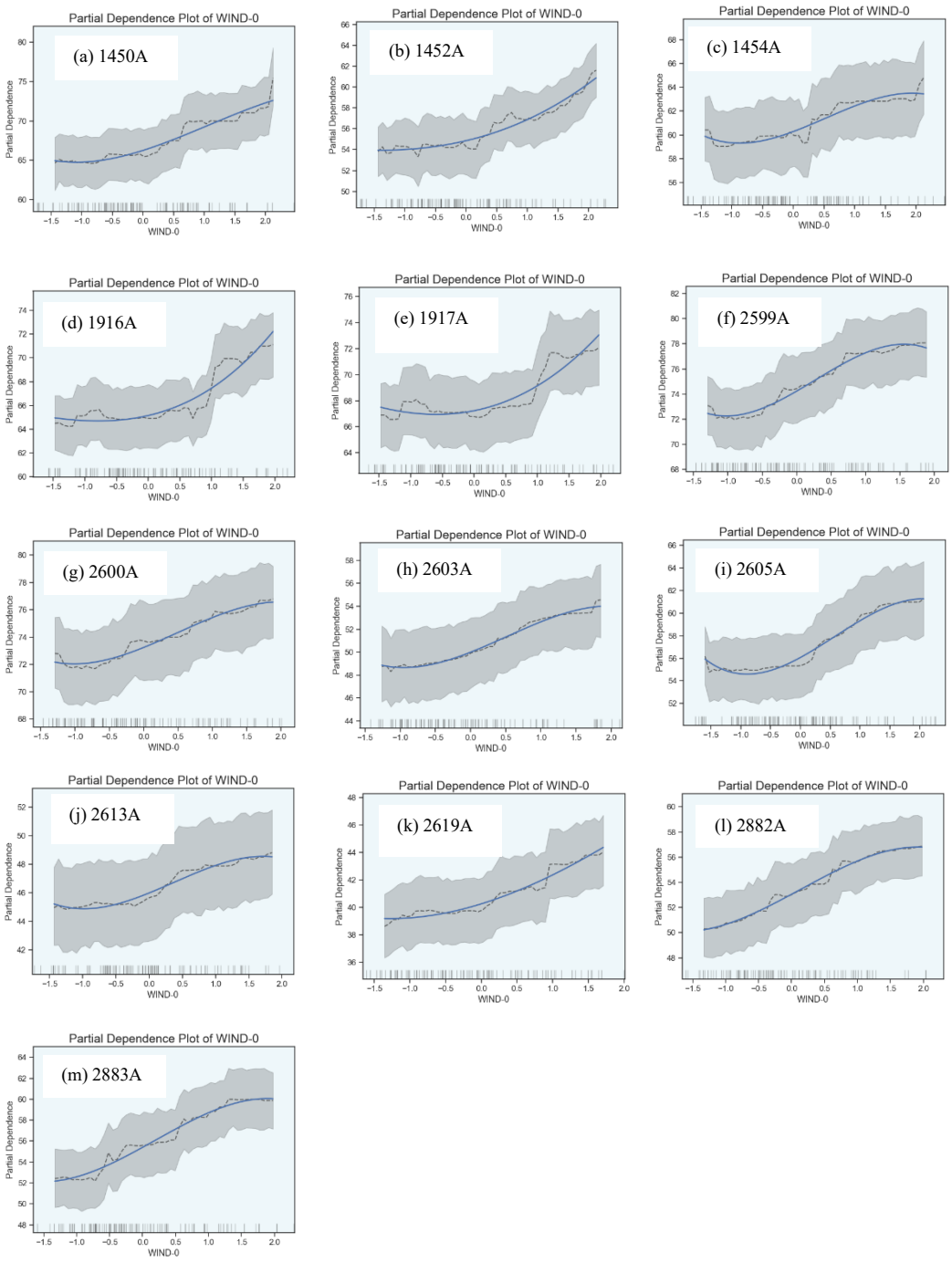

Figure A7. *Effect of WIND-0 on current O_3 concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

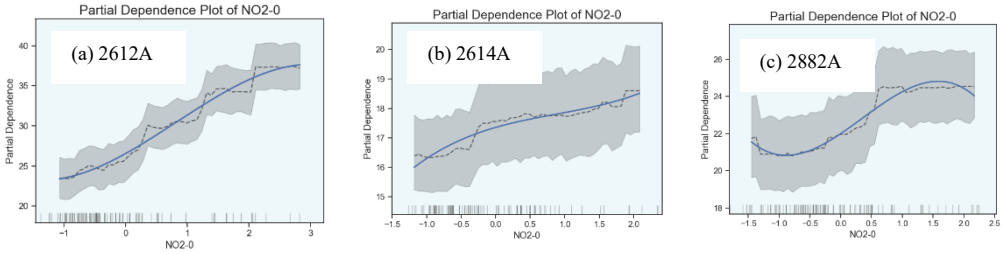

Figure A8. *Effect of NO2-0 on current $PM_{2.5}$ concentration*

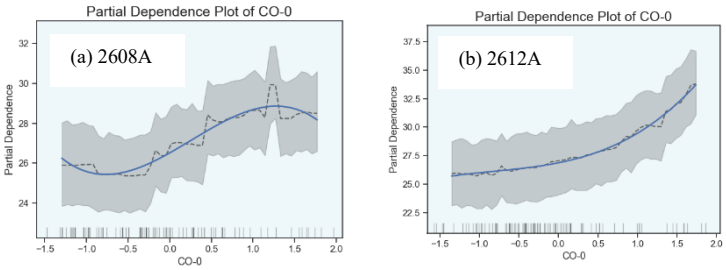

Figure A9. *Effect of CO-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

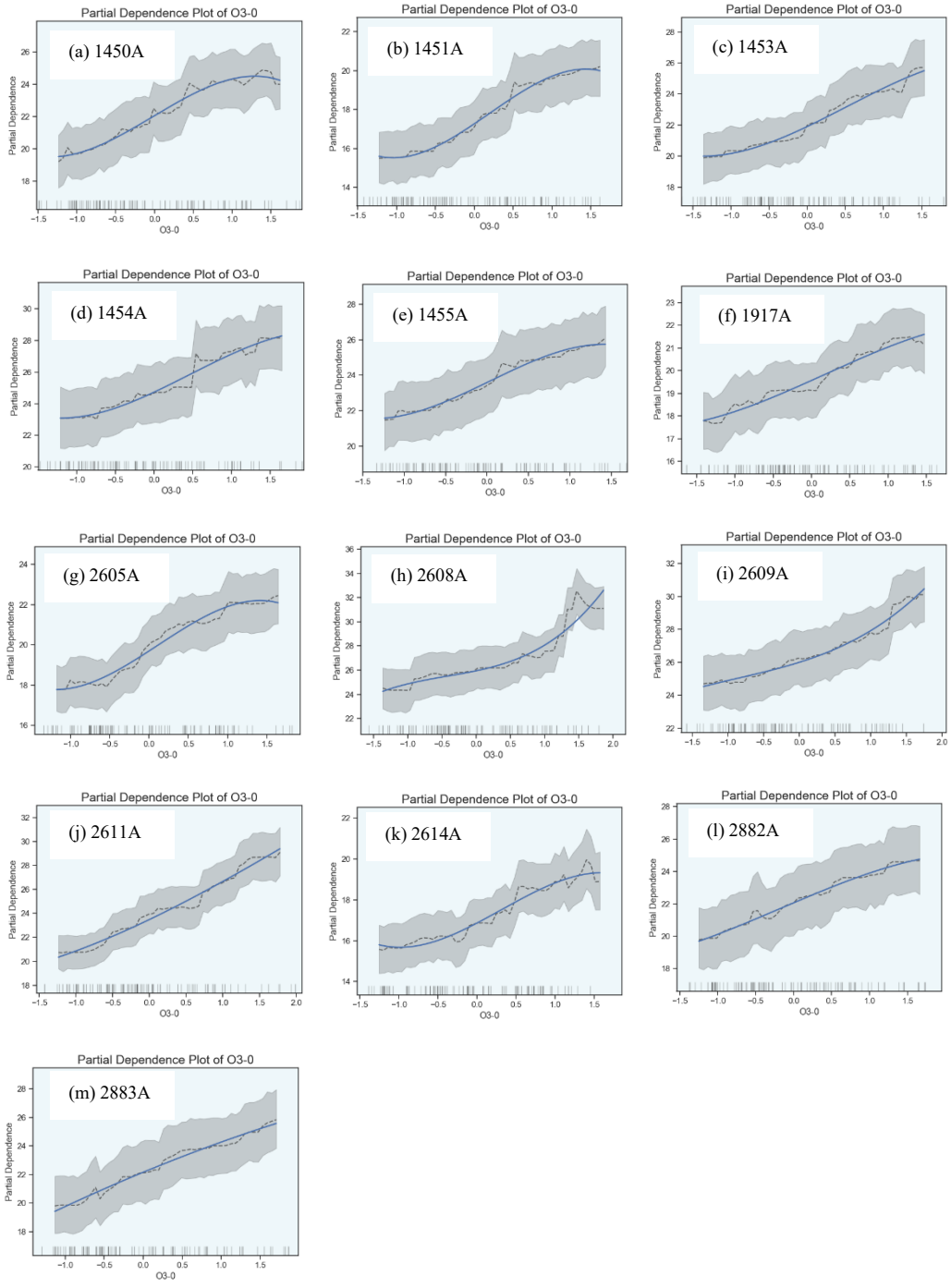

Figure A10. *Effect of O3-0 on current $PM_{2.5}$ concentration*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835

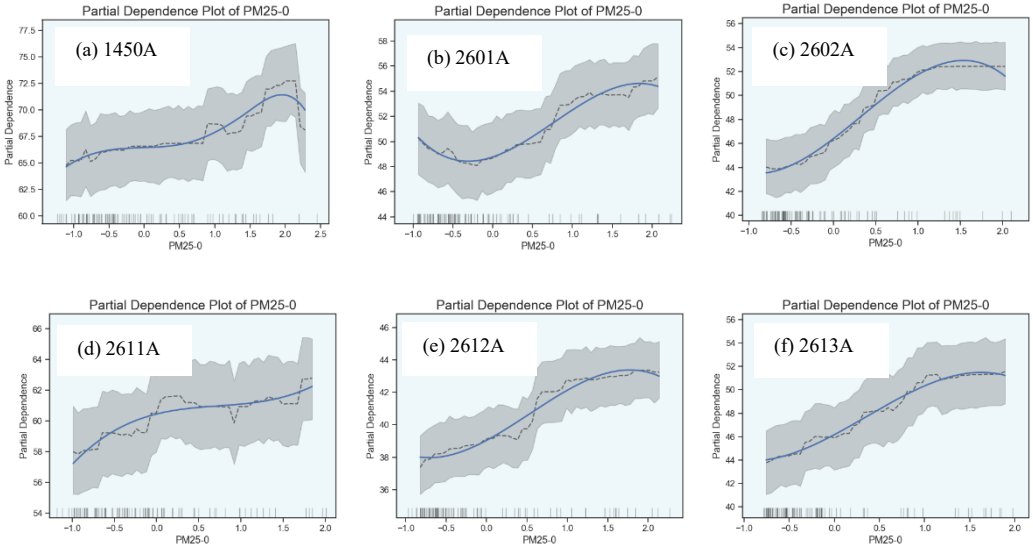

Figure A11. *Effect of PM25-0 on current O_3 concentration*

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 24(2):2811-2835.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2402_28112835