MINISTRY OF EDUCATION AND TRAINING
HANOI UNIVERSITY OF MINING AND GEOLOGY

LE TUAN NGOC

DEVELOPMENT OF ARTIFICIAL INTELLIGENCE MODELS FOR REAL-TIME AIR-POLLUTION FORECASTING DURING MINING OPERATIONS AT THE SIN QUYEN COPPER MINE, LAO CAI

SUMMARY OF DOCTORAL DISSERTATION IN ENGINEERING

Hanoi - 2026

The dissertation was completed at: Department of Surface Mining
Faculty of Mining, Hanoi University of Mining and Geology

Scientific supervisors:

1. **Dr. Hoang Nguyen**, Hanoi University of Mining and Geology
2. **Prof. Dr. Bui Xuan Nam**, Ton Duc Thang University

Reviewer 1:

Reviewer 2:

Reviewer 3:

The dissertation will be defended before the University-level Doctoral Dissertation Assessment Council, convened
at...
at........hour........on........day........month........2026.

The dissertation is available at the National Library of Vietnam
or the Library of Hanoi University of Mining and Geology

INTRODUCTION

1. Rationale of the study

In Vietnam's socio-economic development strategy, the mining industry plays a key role by supplying essential raw materials and fuels for energy security and infrastructure development, including minerals extracted by open-pit methods. Open-pit mining requires the removal of large volumes of overburden and surrounding waste rock to recover the orebody, typically using drilling and blasting together with high-capacity, mechanized mining equipment fleets. Consequently, the environmental impacts are substantial, particularly those affecting air quality.

Most large open-pit mines in Vietnam, including the Sin Quyen copper mine in Lao Cai Province, are entering deeper stages of extraction. As the pit geometry progressively develops into a deep basin, a distinctive and complex microclimate is created. Natural ventilation may be substantially weakened, while thermal inversion and recirculating airflow can prevent dust generated by blasting and haulage from dispersing, causing pollutants to be retained at the pit bottom and on working benches. This condition may produce severe localized pollution, directly affect workers' health and equipment productivity, influence mining output, and degrade ambient air quality in surrounding areas.

Current inspection and control of air-environment parameters at open-pit mines mainly rely on periodic field measurements using mobile instruments, commonly conducted only once per quarter. This is essentially a post-event approach, because managers become aware of pollution levels only after an event has occurred. Such reactive management is no longer appropriate for modern mining. The industry therefore needs to shift toward forecasting and proactive control, so that the time and location of potential exceedances of dust or toxic-gas limits can be identified in advance, enabling the mine to adjust blasting and haulage activities or activate dust-suppression systems before adverse environmental impacts occur.

Advances in digital technologies and the Fourth Industrial Revolution have created favorable conditions for applying artificial intelligence models to smart, automated, and real-time mine management. Artificial intelligence (AI), particularly deep-learning algorithms combined with modern optimization techniques, can model and forecast the temporal evolution of dust and gaseous pollutants in mine environments, addressing limitations of conventional statistical methods in representing complex nonlinear behavior.

In Vietnam, studies on AI-based air-quality forecasting for open-pit mining remain limited, especially for deep mines with complex microclimatic conditions such as the Sin Quyen copper mine. In response to the urgent operational need at Sin Quyen and international technological trends, the dissertation entitled "Development of artificial intelligence models for real-time air-pollution forecasting during mining operations at the Sin Quyen copper mine, Lao Cai" has scientific and practical significance. It provides a forecasting tool to support safer, more efficient, and more sustainable mine operation.

2. Objective

To develop and validate artificial intelligence models for forecasting fine-particulate $PM_{2.5}$ and CO concentrations at three elevation-stratified monitoring locations in the Sin Quyen copper mine. The forecasts provide information to support mine managers in selecting appropriate production-control and environmental-management measures. The results are validated within the scope of field measurements from the Sin Quyen mine and may be used as a reference, subject to recalibration, for deep open-pit mines with similar conditions.

3. Research contents

The dissertation focuses on the following principal research contents:

- Reviewing the generation and dispersion of air pollutants in deep open-pit mining, the current status of air-environment monitoring and forecasting, and trends in the application of artificial intelligence to mining;

- Investigating the theoretical foundations of artificial intelligence models, machine-learning and deep-learning algorithms, and optimization methods for real-time air-quality forecasting;

- Developing an air-environment, meteorological, and production database for the Sin Quyen copper mine, and developing artificial intelligence models suitable for forecasting air quality under deep open-pit mining conditions;

- Evaluating and comparing the performance of forecasting models; analyzing the effects of mining operations and mine microclimate on air quality; and proposing approaches for applying the forecasting models to environmental management and mine-production control.

4. Research objects and scope

4.1. Research objects

The research objects of the dissertation are:

- Mining-technological, microclimatic, and air-environmental factors associated with deepening operations at the Sin Quyen copper mine, Lao Cai. Within the experimental scope of the dissertation, the two selected output variables are fine-particulate $PM_{2.5}$ and CO; PM_{10}, TSP, NO_x, SO_2, noise, and vibration are not model outputs and are identified as extensions for future research.

- Artificial intelligence models and optimization algorithms capable of forecasting air quality in real time under deep open-pit mining conditions.

4.2. Research scope

- Spatial scope: The study was conducted within the open-pit workings of the Sin Quyen copper mine, Lao Cai Province. Time-series forecasts were developed for three fixed locations at elevations +102 m, +68 m, and -58 m. The results do not represent a continuous three-dimensional concentration field throughout the pit; extrapolation to sheltered areas or newly developed benches requires validation using additional measurements.

- Temporal scope: Mine surveys began in December 2023, when the candidate was admitted to the doctoral program. Real-time data were collected from 10 April 2024 to 31 March 2025.

5. Research methods

The principal research methods used to achieve the stated objectives include:

- Synthesis, analysis, and evaluation of domestic and international literature concerning air pollution, mine microclimate, and artificial intelligence applications for forecasting mine-environment conditions;

- Field surveys and the collection and processing of air-environmental, meteorological, and production data related to machinery and equipment used at the Sin Quyen copper mine, Lao Cai;

- Statistical and data-analytical methods for evaluating distributions, temporal trends, and relationships among factors influencing air quality in the mine;

- Modeling and artificial intelligence methods, including machine-learning, deep-learning, and ensemble-learning algorithms, to develop real-time air-quality forecasting models suitable for conditions at the Sin Quyen copper mine;

- Hyperparameter optimization using modern optimization algorithms to improve forecasting accuracy and model generalization;

- Model-performance evaluation using statistical indicators including MAE, RMSE, and the coefficient of determination R^2, together with inter-model comparisons to identify suitable models for the Sin Quyen copper mine and deep open-pit mines with similar conditions.

6. Novel contributions of the dissertation

- A stratified monitoring procedure was established at three elevations, together with a processing workflow for $PM_{2.5}$, CO, meteorological, and production-activity time series. CEEMDAN, interaction features, and lagged variables were used to reduce the influence of local noise and represent the temporal accumulation of pollutants in a deep open-pit mine.

- The XGBoost model was optimized using the Marine Predators Algorithm (MPA), and a Stacking Ensemble model was constructed. On the Sin Quyen test dataset, MPA-XGBoost achieved low RMSE and MAE values and high R^2 values relative to the evaluated benchmark models, while improving peak tracking compared with the non-optimized RF model.

- Temporal relationships among drilling-blasting cycles, loading-haulage operations, microclimatic conditions, and variations in $PM_{2.5}$ and CO were clarified using time-series analysis, lagged variables, and an attention mechanism. The results show that incorporating information from the mining-production chain into the feature set enables the models to better represent sudden emission episodes and prolonged accumulation periods.

- A procedure was proposed for converting 1–3 h forecasts into monitoring, alert, and operational-recommendation levels subject to approval by authorized mine personnel, thereby supporting a transition from post-event environmental management to preventive management within the Sin Quyen case-study scope.

7. Defended propositions

Proposition 1:

In a deep open-pit mine, pit geometry, thermal stratification, and recirculating airflow zones can make $PM_{2.5}$ and CO dispersion nonlinear and temporally persistent. Combining CEEMDAN decomposition, microclimatic interaction variables, and lagged features produces a dataset that more effectively represents this mechanism and reduces the influence of local noise on the forecasting models.

Proposition 2:

For the monitoring dataset collected at the Sin Quyen copper mine, the MPA-XGBoost and Stacking Ensemble models outperformed the evaluated benchmark models according to R^2, RMSE, and MAE. In particular, MPA-XGBoost reduced underprediction of concentration peaks relative to non-optimized RF, thereby improving the alert value of the forecasts during blasting and high-intensity haulage periods.

Proposition 3:

An environmental-management framework integrating forecasting models can predict $PM_{2.5}$ and CO concentrations at each monitoring station in advance, thereby supporting adjustments to production activities and air-pollution mitigation measures during mining operations.

8. Scientific and practical significance

8.1. Scientific significance

- The dissertation proposes an approach for digitizing the emission and microclimatic characteristics of deep open-pit mines through stratified monitoring data, CEEMDAN, interaction features, and lagged variables. This approach helps preserve the physically meaningful trends of $PM_{2.5}$ and CO time series before model training.

- The dissertation provides additional empirical evidence for the application of MPA-XGBoost and Stacking Ensemble to $PM_{2.5}$ and CO forecasting in deep open-pit mines, based on R^2, RMSE, and MAE and on the models' ability to track extreme values.

- The research results provide a basis for shifting from static, post-event assessment toward forecasting, decision support, and proactive prevention.

8.2. Practical significance

- Air-quality forecasts, particularly forecasts of $PM_{2.5}$ and CO concentrations, support proactive production control and environmental safety at the mine;

- Replacing periodic, generalized water spraying with targeted, quantitatively scheduled dust suppression can directly reduce industrial-water use and fuel consumption by auxiliary equipment. Tight control of gas and dust concentrations at the pit bottom also helps maintain safe visibility for mobile equipment and reduce occupational-accident and disease risks;

- The dissertation framework, extending from stratified monitoring to the MPA-XGBoost forecasting core, is organized as a complete technical procedure. It can be transferred and adapted to other deep open-pit mines in Vietnam with comparable conditions, supporting safe and sustainable mining.

9. Database

The database used in the dissertation was organized into the following groups:

- Primary data from three automatic monitoring stations at +102 m, +68 m, and -58 m, sampled at 5-min intervals from 10 April 2024 to 31 March 2025. The recorded fields include time, $PM_{2.5}$, CO, temperature, relative humidity, pressure, wind speed, and wind direction.
- Mining-technology and production-organization data: drilling-blasting times, shift-change cycles, loading-haulage-dumping activities, equipment inventories, and equipment operating times. These data were used for event alignment and interpretation, not as direct measurements of emissions from individual machines.
- Spatial data and mine records: current mine plans, elevation sections, field/UAV survey results, mining plans, and mine environmental documents used to select station locations and interpret dispersion conditions.
- Secondary data: scientific publications, monitoring reports, environmental-impact assessment reports, and relevant technical standards and regulations on ambient and occupational air quality.
- Data-quality control: verification against certificates of origin and quality and equipment-acceptance records; timestamp synchronization; checks of continuity, plausible measurement ranges, and consistency among stations; and flagging of missing, disconnected, and outlying observations before processing.
- Intellectual-property boundary: the monitoring-station system was developed under national project KC.05/21-25, in which the doctoral candidate participated as a project leader. The independent contributions of the dissertation focus on research-dataset design, feature-processing techniques, forecasting-model optimization, and the decision-support procedure.

10. Dissertation structure

The dissertation comprises 174 pages, including 9 tables and 33 figures. In addition to the Introduction, Conclusions, References, and Appendices, the dissertation contains four chapters as follows:

Introduction;

Chapter 1: Overview of air-environment control in open-pit mining and air-quality forecasting models

Chapter 2: Scientific basis and research methodology

Chapter 3: Data collection, analysis, and feature extraction at the Sin Quyen copper mine, Lao Cai

Chapter 4: Development of artificial intelligence models for real-time air-pollution forecasting during mining operations and proposal of a proactive operational decision-making procedure for the Sin Quyen copper mine, Lao Cai

Conclusions and recommendations

List of publications related to the doctoral dissertation.

CHAPTER 1: OVERVIEW OF AIR-ENVIRONMENT CONTROL IN OPEN-PIT MINING AND AIR-QUALITY FORECASTING MODELS

1.1. Emission and pollutant-dispersion characteristics in deep open-pit mines

Pollutant concentrations in open-pit mines vary under the simultaneous effects of emission sources, shift organization, pit geometry, and meteorological conditions. Drilling and blasting generate point sources with short-duration high-intensity peaks; loading generates dust at working faces; haulage forms line sources along mine roads and produces both resuspended dust and diesel exhaust; and dumping generates area sources over waste-dump surfaces. These sources may overlap, producing cyclic, nonlinear time series with short-lived extreme values.

As a mine deepens, bench faces and stepped topography obstruct airflow, increasing the likelihood of thermal stratification and recirculating eddies. Fine particles may remain suspended, while CO accumulates along haul roads or in areas where diesel equipment operates under high load. This retention mechanism implies that the concentration at time t also depends on previous states at t-1, t-3, t-6, or longer horizons, providing the basis for lagged variables and time-series models.

Table 1. Classification of emission sources by mining operation

Mining operation	Source type	Time-series characteristic	Dominant equipment/activity
Drilling-blasting	Point source	Abrupt, high-intensity emission peaks	Drilling rigs, delay blasting network, explosive type and charge
Loading	Point source	Sustained dust background at the working face	Electric and hydraulic excavators
Haulage	Line source	Continuous emission background with peaks after shift change	Large-capacity haul trucks and inclined roads
Dumping	Area source	Persistent background pollution over a broad area	Dump trucks and bulldozers

1.2. Current status of air-environment monitoring and forecasting in deep open-pit mines

A review of domestic and international studies shows that manual monitoring has three principal limitations: it does not capture emission cycles; it is difficult to identify dust-retention processes associated with thermal inversions; and it does not provide information rapidly enough for operational control. Recent studies increasingly use IoT sensors, multi-point stations, UAVs, CFD simulation, machine learning, and deep learning. However, many studies use urban or shallow-mine datasets and do not represent elevation-dependent stratification or the overlap of mining operations in deep pits.

In Vietnam, environmental monitoring and control requirements are specified in technical regulations, but continuous data are rarely connected directly to forecasting and operational decisions. The dissertation therefore adopts an integrated approach comprising continuous measurement, time-aware data processing, model development, and the conversion of forecasts into conditional operational recommendations.

1.3. Current applications of Artificial Intelligence (AI) and optimization in environmental management at deep open-pit mines

RF and XGBoost are suitable for tabular data and can represent nonlinear relationships and interactions among variables. LSTM, TCN, and AttnLSTM are suited to temporal dependence, while Attention assigns greater weight to influential time steps. Stacking Ensemble combines forecasts from base models through a meta-model to reduce the bias of individual learners. MPA is used to identify hyperparameter configurations that minimize validation RMSE.

Four research gaps were identified: a lack of stratified monitoring data from deep open-pit mines; a lack of feature-engineering workflows linked to pollutant-accumulation mechanisms; insufficient simultaneous evaluation of average forecasting accuracy and peak-capturing ability; and a lack of procedures for translating forecasts into operational decisions. The dissertation addresses these gaps within the Sin Quyen case study.

1.4. Chapter 1 summary

Chapter 1 establishes the basis for selecting $PM_{2.5}$ and CO, classifies emission sources by mining operation, clarifies the roles of pit geometry and microclimate, and demonstrates the need to combine continuous monitoring with AI models. The review provides the foundation for the scientific basis and research procedure developed in Chapter 2.

CHAPTER 2: SCIENTIFIC BASIS AND RESEARCH METHODOLOGY

2.1. Emission dynamics and microclimatic mechanisms at the Sin Quyen copper mine

Emission dynamics are represented according to mining-operation cycles. Blasting produces high-amplitude $PM_{2.5}$ pulses; loading and haulage maintain background emissions, with CO mainly associated with the diesel fleet; and dumping contributes dust over broad areas. When several operations occur simultaneously, assigning a measured concentration to a single machine is inappropriate. The dissertation uses production-event timestamps to align and interpret concentration peaks, but does not treat them as direct emission measurements from individual machines.

A thermal inversion is defined as a state in which the air temperature above is higher than that below, weakening vertical convection. At the +102 m pit-rim station, natural wind promotes rapid dispersion; at +68 m, airflow is influenced by bench slopes and haulage routes; and at the -58 m pit bottom, confined geometry and recirculation increase pollutant residence time. These differences require station-specific lag selection and separate model training.

Table 2. Relationship between spatial stratification and time-series processing requirements

Elevation	Microclimatic characteristic	Dispersion capacity	Historical information range to be considered
+102 m	Relatively stable natural ventilation	Rapid, short residence time	Short lags of 1–3 h
+68 m	Airflow interaction and local recirculation	Moderate, with short-term accumulation	Lags of 3–6 h
-58 m	Deep confined space, thermal inversion, and recirculation	Low, with potential accumulation across a shift	Longer lags; 6–24 h evaluation

2.2. Feature engineering and rationale for model selection

Mine-environment data contain local fluctuations caused by vehicles passing near a station, equipment operating close to sensors, communication interruptions, and measurement noise. CEEMDAN decomposes the series into intrinsic oscillatory components at multiple time scales, facilitating separation of high-frequency noise from the underlying trend. CEEMDAN is not the only possible method; EMD, EEMD, STL, or statistical filters may be used instead. It was selected because it is suitable for nonlinear, non-stationary series and reduces mode mixing.

Meteorological interaction features were constructed from wind speed, wind direction, and the orientation of haulage routes and bench slopes. High wind speed does not always reduce pollution: when wind crosses a bench face, flow separation and recirculation may occur. The interaction variables enable the model to distinguish favorable from unfavorable ventilation. Lags of 1, 3, and 6 h represent short emission pulses, post-blast cycles, and accumulation over half a shift; longer windows were considered for the pit-bottom station.

RF and XGBoost were used as tabular-data models; LSTM and TCN represented temporal dependence; AttnLSTM focused on influential time steps; and Stacking combined base predictions using a Ridge meta-learner. MPA updates a population of hyperparameter configurations through Brownian and Lévy motions, balancing global exploration and local refinement. In the dissertation, “Prey” denotes a candidate configuration and “Elite” denotes the current best configuration; these biological terms are not physical mine variables.

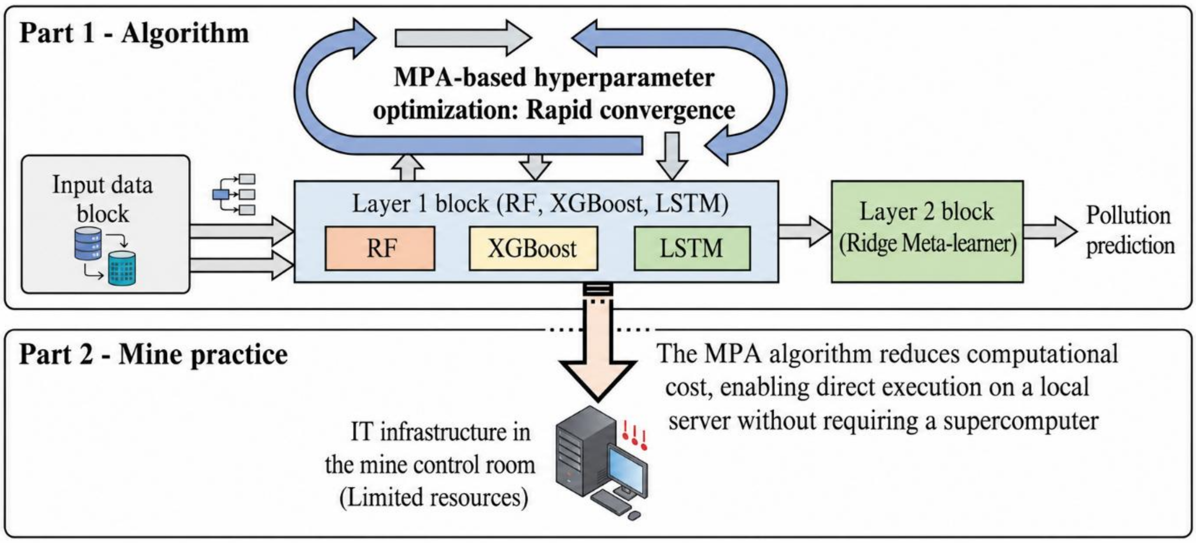


Figure 1. Stacking Ensemble architecture and MPA-based hyperparameter optimization mechanism

2.3. Research procedure

The procedure consists of four steps:

(1) establishing the monitoring network and collecting elevation-stratified data;

(2) preprocessing, decomposition, and construction of interaction and lagged features;

(3) optimizing the hyperparameter search space with MPA using the validation set;

(4) training, testing, and converting forecasts into operational decision-support information. The procedure was designed to ensure reproducibility, prevent data leakage, and clearly define the domain of model validity.

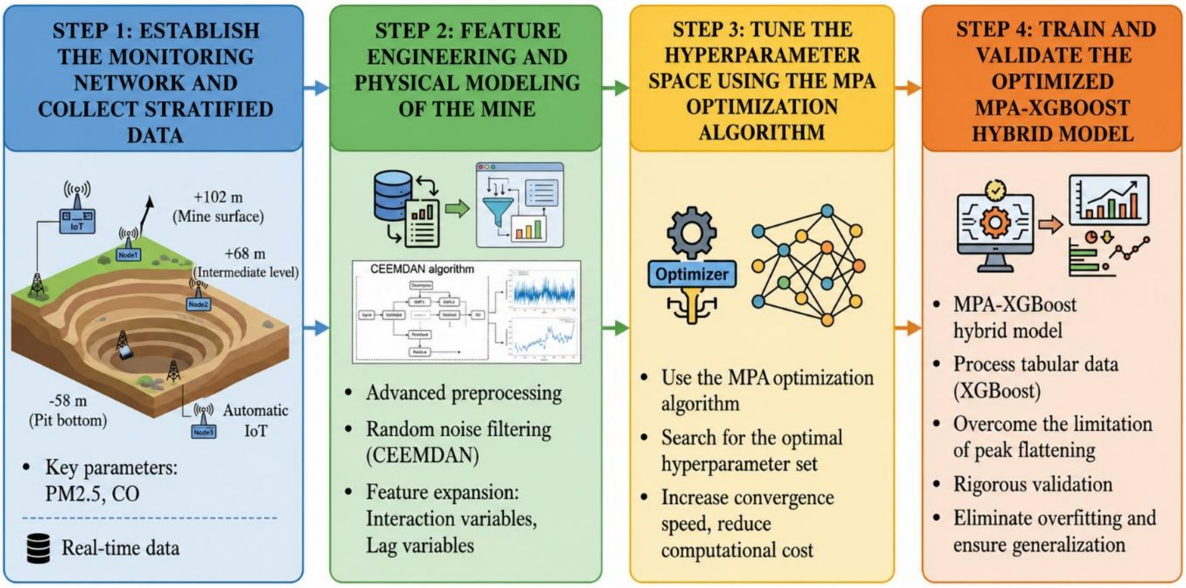


Figure 2. Research procedure of the dissertation

2.4. Chapter 2 summary

Chapter 2 establishes the link between emission-microclimate mechanisms and data-algorithm structures. This scientific basis guides station placement, input-variable selection, time-window design, and model evaluation in Chapters 3 and 4.

CHAPTER 3. DATA COLLECTION, PREPROCESSING, NORMALIZATION, ANALYSIS, AND FEATURE EXTRACTION AT THE SIN QUYEN COPPER MINE, LAO CAI

3.1. Design of the elevation-stratified monitoring network

The Sin Quyen copper mine has stepped topography, a deep pit bottom, and extended haulage routes. Three stations were installed to represent three aerodynamic zones: Station 1 at +102 m represents background conditions and relatively effective ventilation; Station 2 at +68 m represents airflow interaction and haulage activity; and Station 3 at -58 m represents the pit bottom, where thermal inversion, recirculation, and emission peaks are more likely. Locations were selected using mine plans, cross-sections, field surveys, installation safety, and power-supply and data-transmission feasibility.

Figure 3. View of the Sin Quyen copper mine and locations of the three automatic monitoring stations

Table 3. Technical characteristics and functions of the monitoring locations

Station	Elevation	Aerodynamic characteristic	Characteristic emission source	Role in the model
1	+102 m	Relatively stable natural ventilation	Background environment and post-blast effects	Reference background concentration
2	+68 m	Airflow interaction and haulage routes	Resuspended dust and haulage exhaust	Line-source representation
3	-58 m	Thermal inversion and local recirculation	Dust peaks and equipment exhaust	Lagged features and extreme-value forecasting

3.2. Statistical analysis of the time series

Data were recorded at 5-min intervals. Descriptive statistics show that mean $PM_{2.5}$ increased from 56.69 µg/m³ at +102 m to 68.60 µg/m³ at +68 m and 75.06 µg/m³ at -58 m.

The minimum $PM_{2.5}$ concentration at the pit bottom was substantially higher than at the pit rim, indicating a persistent pollution background. Mean temperature increased with depth in the surveyed dataset, while wind speed was lower at the deeper stations; these results are consistent with increased pollutant retention.

Mean CO concentrations were approximately 1.00–1.34 mg/m³, with relatively greater variability at the two deeper stations. CO distribution differed from that of $PM_{2.5}$ because CO was concentrated along haulage routes and near diesel-engine operations, whereas $PM_{2.5}$ was also affected by wall dust, blasting, and dumping. Separate models were therefore trained for each target variable and each station.

Table 4. Descriptive statistics of environmental parameters at the three elevations

Parameter	**Statistic**	**+102 m**	**+68 m**	**-58 m**
$PM_{2.5}$ (µg/m³)	Mean	56.69	68.60	75.06
$PM_{2.5}$ (µg/m³)	Standard deviation	11.26	13.58	13.67
$PM_{2.5}$ (µg/m³)	Maximum	117.12	107.53	112.64
$PM_{2.5}$ (µg/m³)	Minimum	14.21	32.77	39.34
CO (mg/m³)	Mean	1.34	1.00	1.00
CO (mg/m³)	Standard deviation	0.23	0.30	0.30
Temperature (°C)	Mean	17.69	19.72	21.20
Relative humidity (%)	Mean	76.60	80.00	80.00
Wind speed (m/s)	Mean	1.51	1.32	1.32

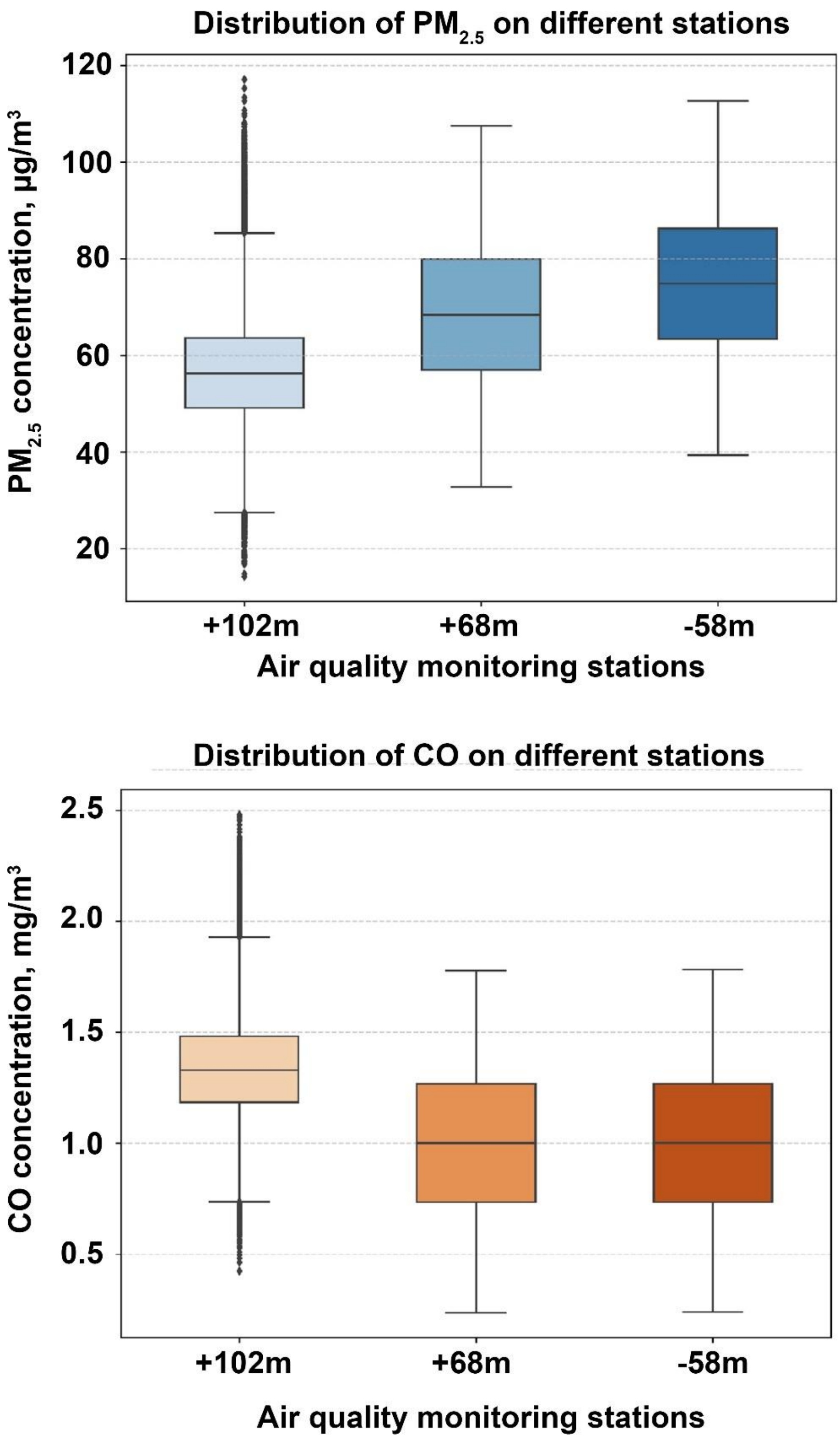


Figure 4. Distribution of $PM_{2.5}$ and CO across monitoring elevations

3.2.2. Identification of abrupt emission events by shift cycle

Timestamp alignment shows that $PM_{2.5}$ commonly increased sharply around blasting periods. At +102 m, peaks decayed more rapidly because of better ventilation, whereas at +68 m and -58 m the decay period was prolonged. Following shift changes, the simultaneous start-up of haul trucks and diesel excavators could generate secondary CO and $PM_{2.5}$ peaks. Causes were assigned by comparing production logs, meteorological conditions, and simultaneous behavior across stations; the dissertation does not claim that every peak was generated by only one mining operation.

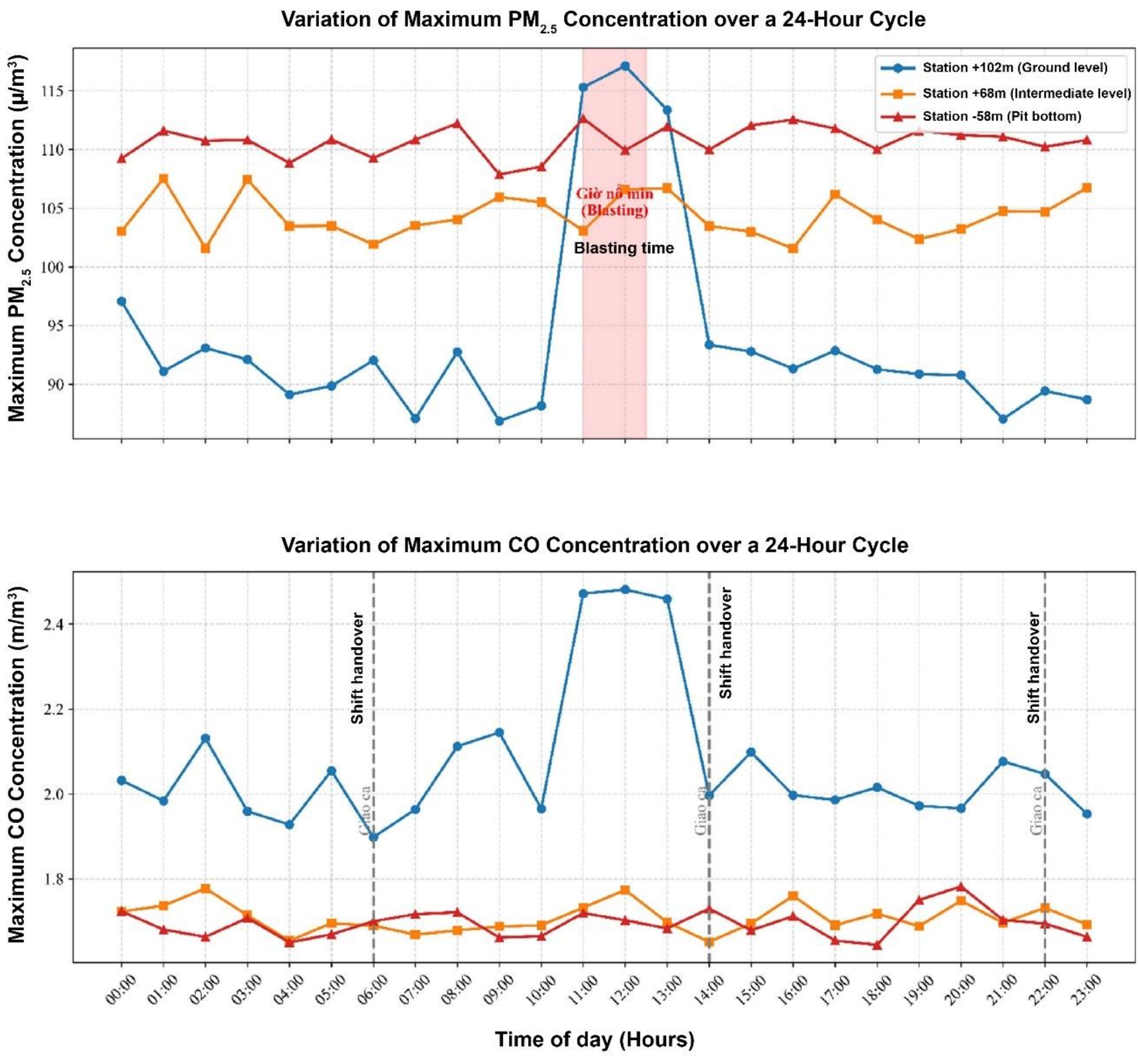


Figure 5. Maximum $PM_{2.5}$ and CO variation over a 24-h cycle

3.3. Data processing and feature-set structure

Processing included continuity checks, flags for communication losses, removal or correction of values outside plausible ranges, rolling-window smoothing, CEEMDAN decomposition, and creation of lagged variables. A 3-h moving average represented the post-blast trend, while a 6-h window represented accumulation over half a shift. $PM_{2.5}$ and CO were processed using analogous workflows, but station- and variable-specific parameters were selected to avoid suppressing meaningful peaks.

Lagged features were created from historical $PM_{2.5}$/CO values and meteorological variables. Data were ordered chronologically and divided into 70% training, 15% validation, and 15% testing subsets. MinMaxScaler was fitted only on the training data and then applied to the remaining subsets. This procedure prevents future information from entering training and ensures that test performance reflects genuine generalization.

3.4. Chapter 3 summary

Chapter 3 developed an elevation-stratified empirical dataset, clarified differences among the three elevations, and established relationships between production events and $PM_{2.5}$ and CO time series. The preprocessed dataset served as input for model development in Chapter 4. A major limitation is that the data represent three monitoring points and do not replace a three-dimensional concentration map of the entire pit.

CHAPTER 4. DEVELOPMENT OF ARTIFICIAL INTELLIGENCE MODELS FOR REAL-TIME AIR-POLLUTION FORECASTING AND PROPOSAL OF A PROACTIVE OPERATIONAL DECISION-MAKING PROCEDURE FOR THE SIN QUYEN COPPER MINE

4.1. Forecasting-system architecture

The system comprises the tabular-data models RF and XGBoost; the time-series models LSTM, TCN, and AttnLSTM; the optimized models MPA-RF and MPA-XGBoost; and a Stacking Ensemble. Each model was trained separately for each station and target variable. RF provided an interpretable machine-learning benchmark; XGBoost captured nonlinear interactions; LSTM, TCN, and AttnLSTM modeled temporal dependence; and Stacking combined predictions from multiple architectures.

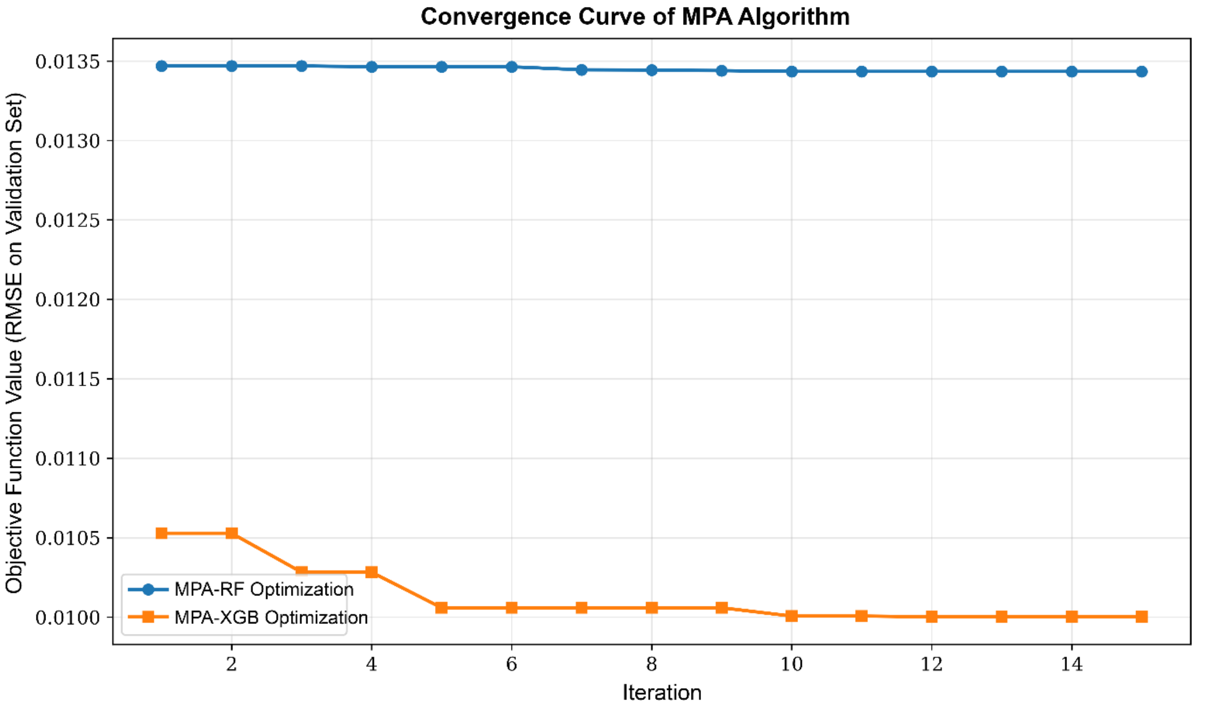
Table 5. Roles of the developed and compared models

Model	Primary role	Strength in the forecasting task	Limitation to be controlled
RF	Machine-learning benchmark	Stable; handles tabular data and supports feature-importance assessment	May smooth peaks and overfit
XGBoost	Boosted-tree model	Low error; handles nonlinearity and interactions	Sensitive to hyperparameters
LSTM	Time-series model	Retains long-term dependence	Performance depends on data volume and configuration
TCN	Temporal convolution	Wide receptive field and parallel training	May be less stable for highly noisy data
AttnLSTM	LSTM + Attention	Focuses on influential time steps	Complex architecture and high training cost
MPA-RF/MPA-XGB	Optimized model	Searches hyperparameters using validation RMSE	Does not guarantee that every test metric is better than the original model
Stacking	Ensemble learning	Compensates for errors of the base models	Depends on diversity among component models

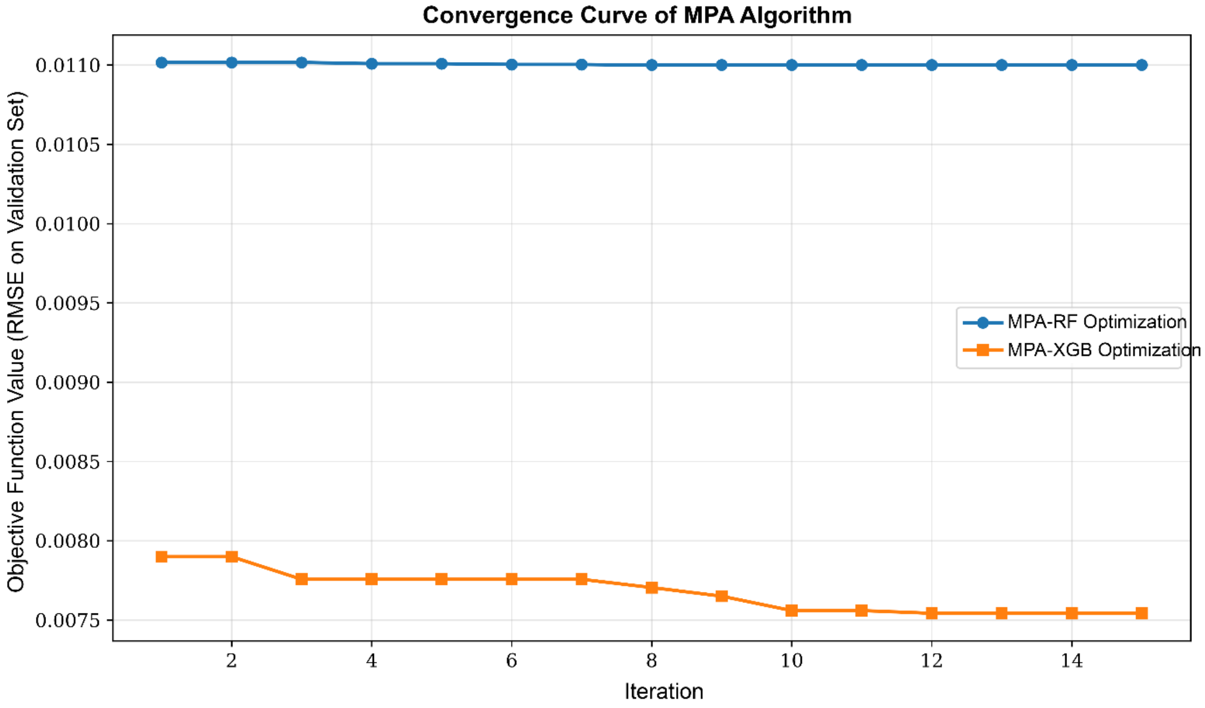
4.2. MPA optimization and development of hybrid models

The RF/XGBoost search spaces included the number of trees, tree depth, learning rate, subsampling ratios, and regularization parameters. MPA initialized multiple candidate configurations, evaluated them using validation RMSE, and updated them through three search phases. The convergence curves show that the objective decreased rapidly during early iterations and stabilized thereafter. After configuration selection, the models were retrained using the combined training and validation data and evaluated once on the test set.

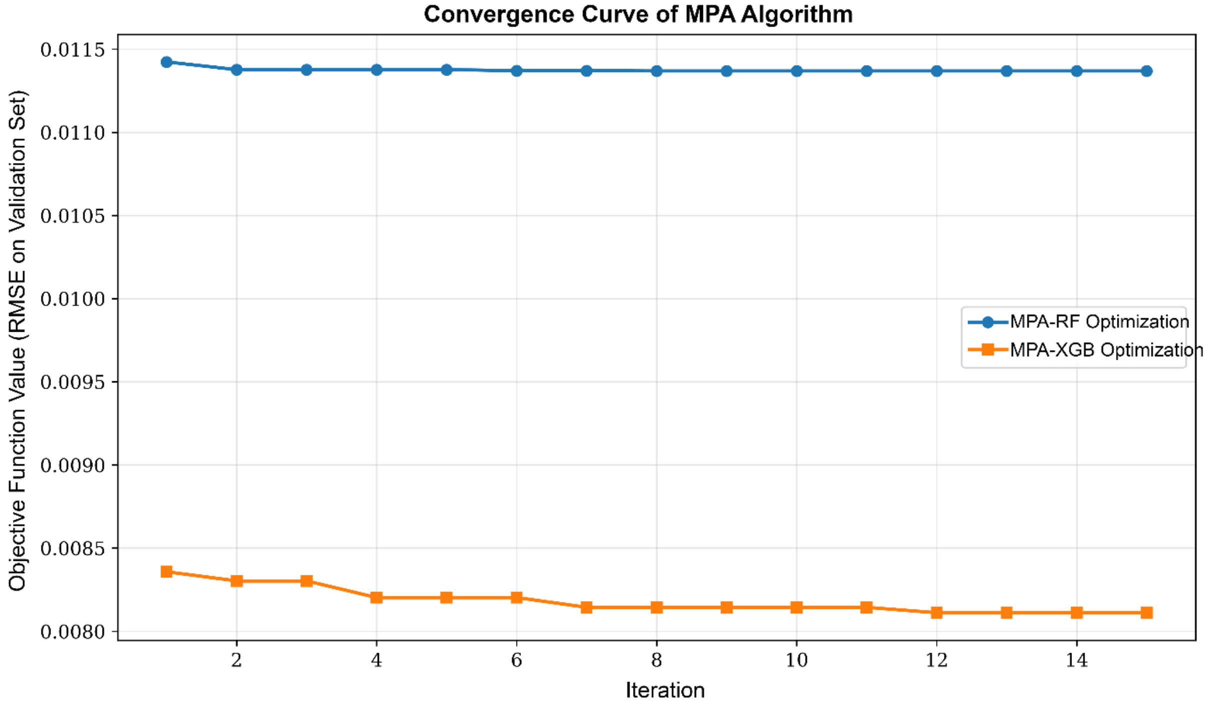
The results show that XGBoost and MPA-XGBoost were the most stable model group across all three stations. MPA-XGBoost showed clearer advantages during optimization/validation and in tracking extremes; on the test set, differences between XGBoost and MPA-XGBoost were small in some cases, and XGBoost occasionally achieved slightly lower RMSE or MAE. The contribution of MPA is therefore interpreted as an improvement in configuration search, stability, and peak-capturing ability, rather than a guarantee that every test metric will always be superior to the original model.



(a) Data from the +102 m station



(b) Data from the +68 m station



(c) Data from the -58 m station

Figure 6. MPA convergence curves for optimization of the RF and XGBoost models

4.3. Development of the Stacking Ensemble

Stacking used out-of-sample predictions from RF, XGBoost, TCN, and AttnLSTM as inputs to a Ridge meta-learner. This architecture combines the strengths of tree-based models and temporal neural networks. Stacking achieved high R^2 for $PM_{2.5}$ at all three stations and showed good stability, but did not always provide the lowest error. The results demonstrate that model selection should consider multiple metrics and operational objectives rather than a single coefficient.

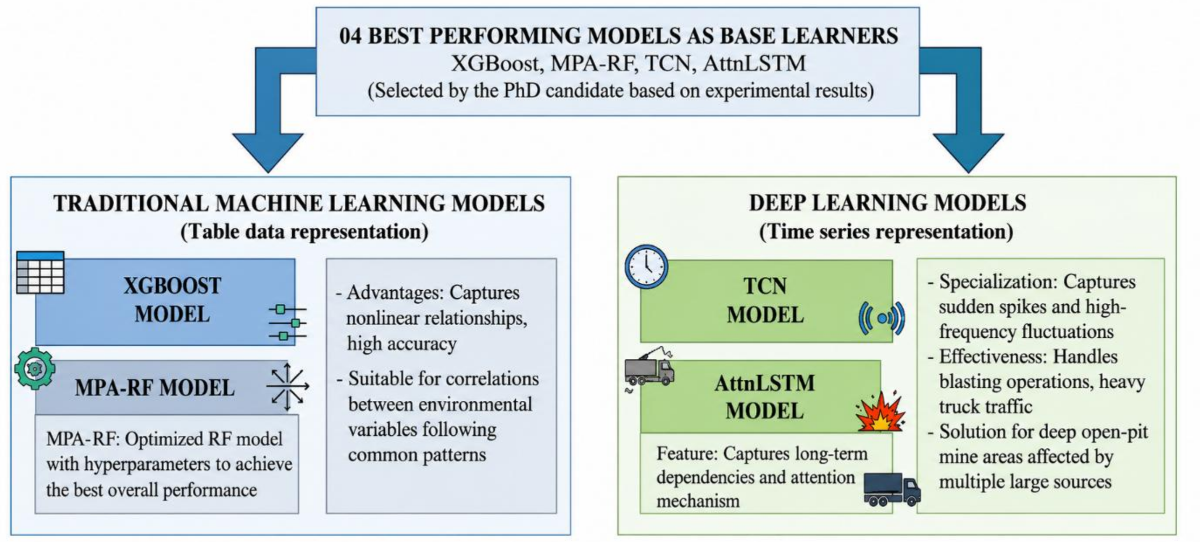


Figure 7. Grouping and roles of the four base models selected to develop the Stacking Ensemble for air-quality forecasting at the Sin Quyen copper mine

4.4. Performance and reliability evaluation

At the +102 m station, XGBoost and MPA-XGBoost achieved test R^2 values of 0.996 for $PM_{2.5}$ and 0.997 for CO. At +68 m, the two models achieved R^2 of approximately 0.997 for $PM_{2.5}$ and approximately 1.000 for CO. At -58 m, XGBoost/MPA-XGBoost achieved R^2 of approximately 0.997 for $PM_{2.5}$ and 1.000 for CO. Individual deep-learning networks performed less effectively for $PM_{2.5}$ at the intermediate and pit-bottom stations, particularly TCN at -58 m, indicating that appropriately engineered tabular data can outperform more complex deep-learning architectures for the present dataset.

R^2 represents the proportion of variance explained and must be interpreted together with RMSE and MAE. Actual-versus-predicted correlation analyses and the final 200-h test window show that MPA-XGBoost and XGBoost tracked most fluctuations closely, whereas RF and AttnLSTM smoothed some segments. Peak-capturing ability is particularly important for alerts following blasting and during high-intensity haulage.

Table 6. Representative test results of the high-performing model group

Station/target	**High-performing model group**	**Representative test RMSE**	**Representative test MAE**	**Test R^2**
+102 m – $PM_{2.5}$	XGBoost / MPA-XGBoost	0.346	0.210–0.217	0.996
+102 m – CO	XGBoost / MPA-XGBoost	0.007	0.004–0.005	0.997
+68 m – $PM_{2.5}$	XGBoost / MPA-XGBoost	0.414–0.416	0.325–0.328	0.997
+68 m – CO	XGBoost / MPA-XGBoost	0.005	0.004	≈1.000
-58 m – $PM_{2.5}$	XGBoost / MPA-XGBoost	0.428–0.440	0.342–0.353	0.997
-58 m – CO	XGBoost / MPA-XGBoost	0.006	0.004	≈1.000

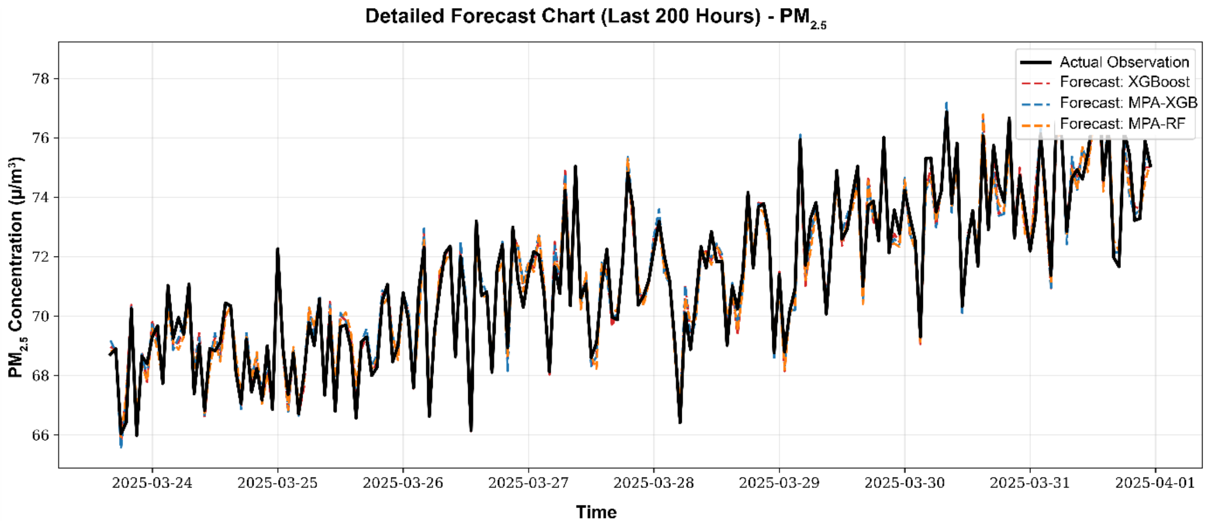
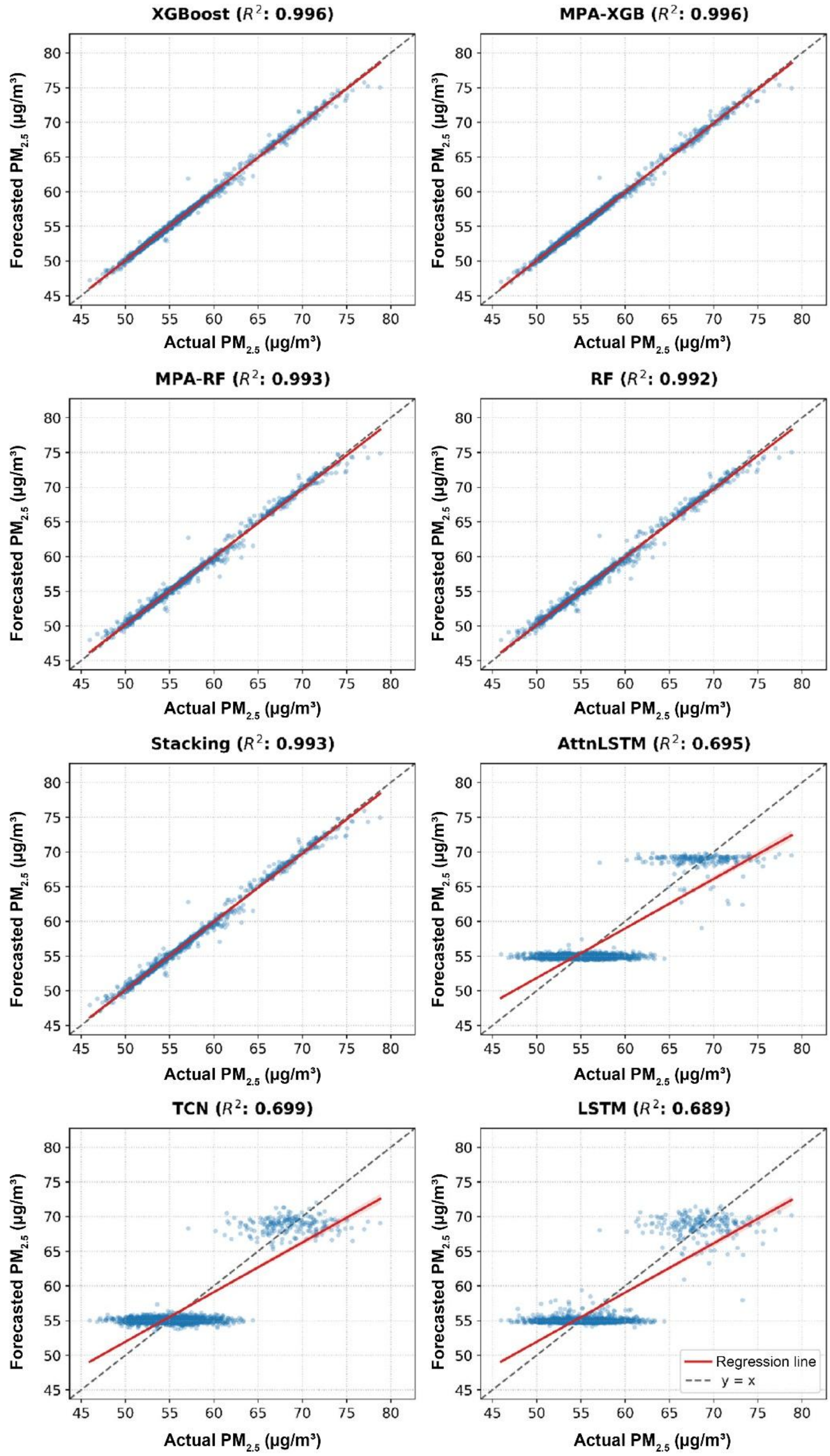
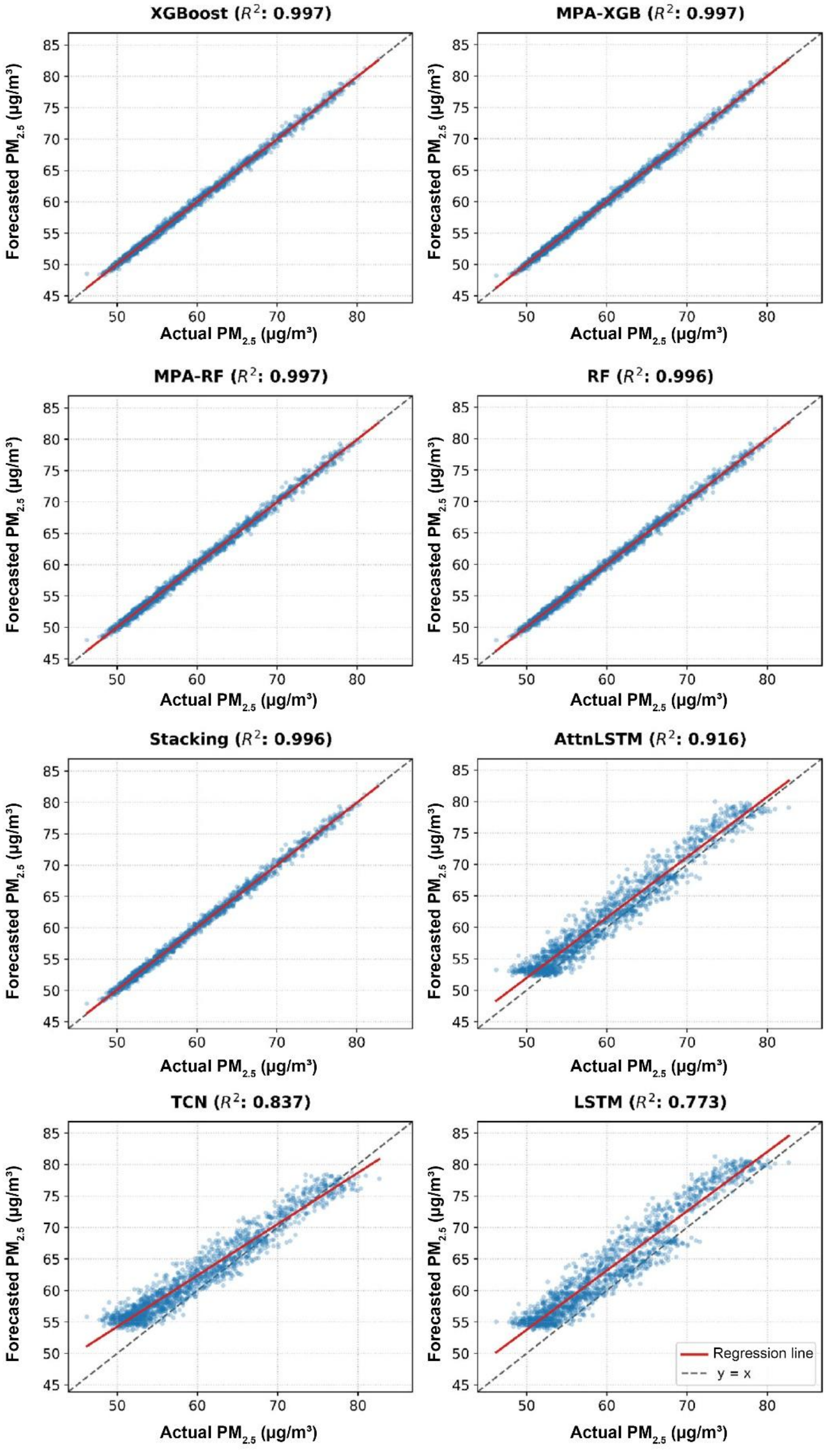


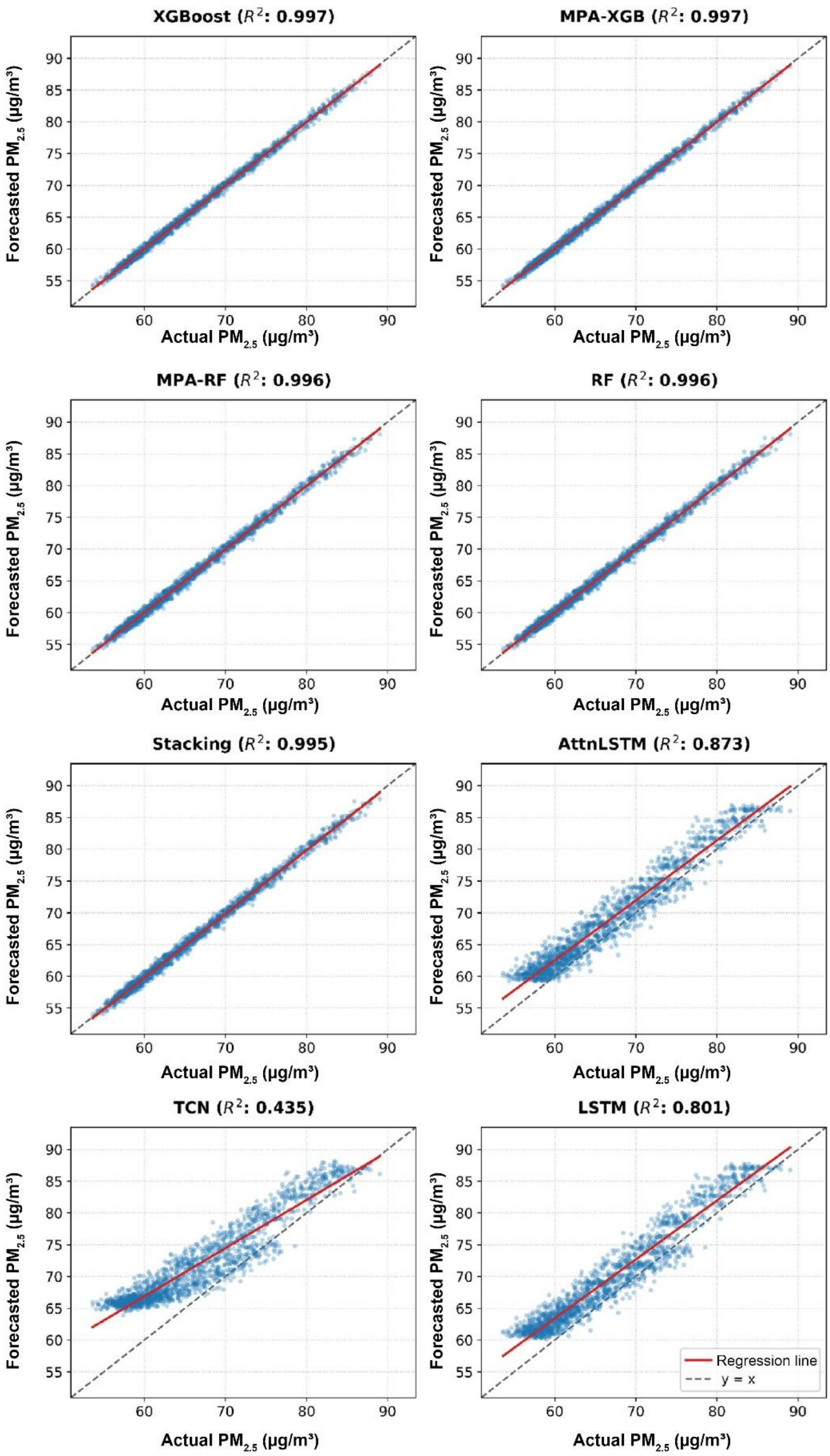
Figure 8. Tracking of $PM_{2.5}$ peaks at the -58 m station during the final 200 test hours



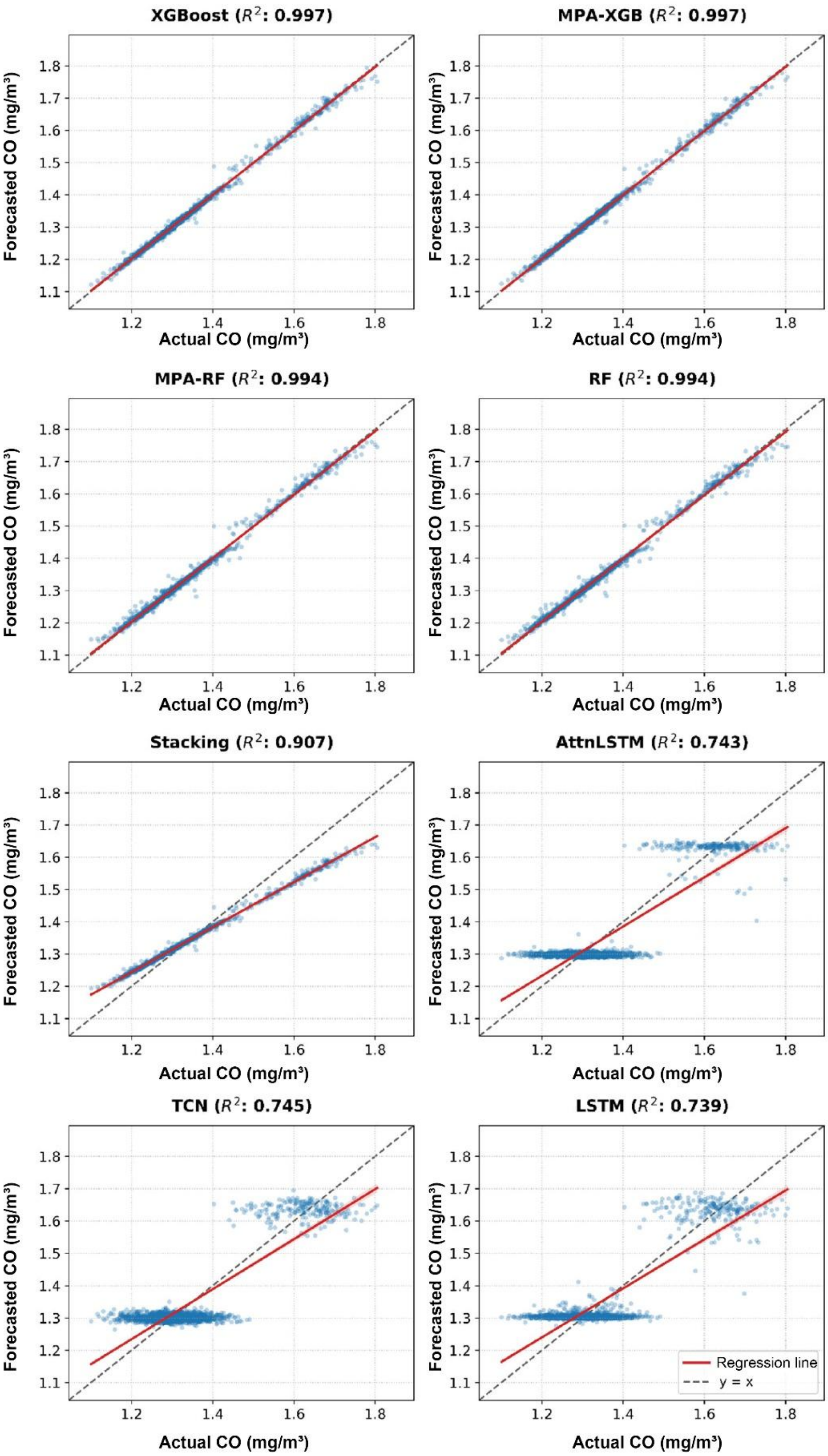
(a) Correlation between actual and predicted $PM_{2.5}$ values at the +102 m station



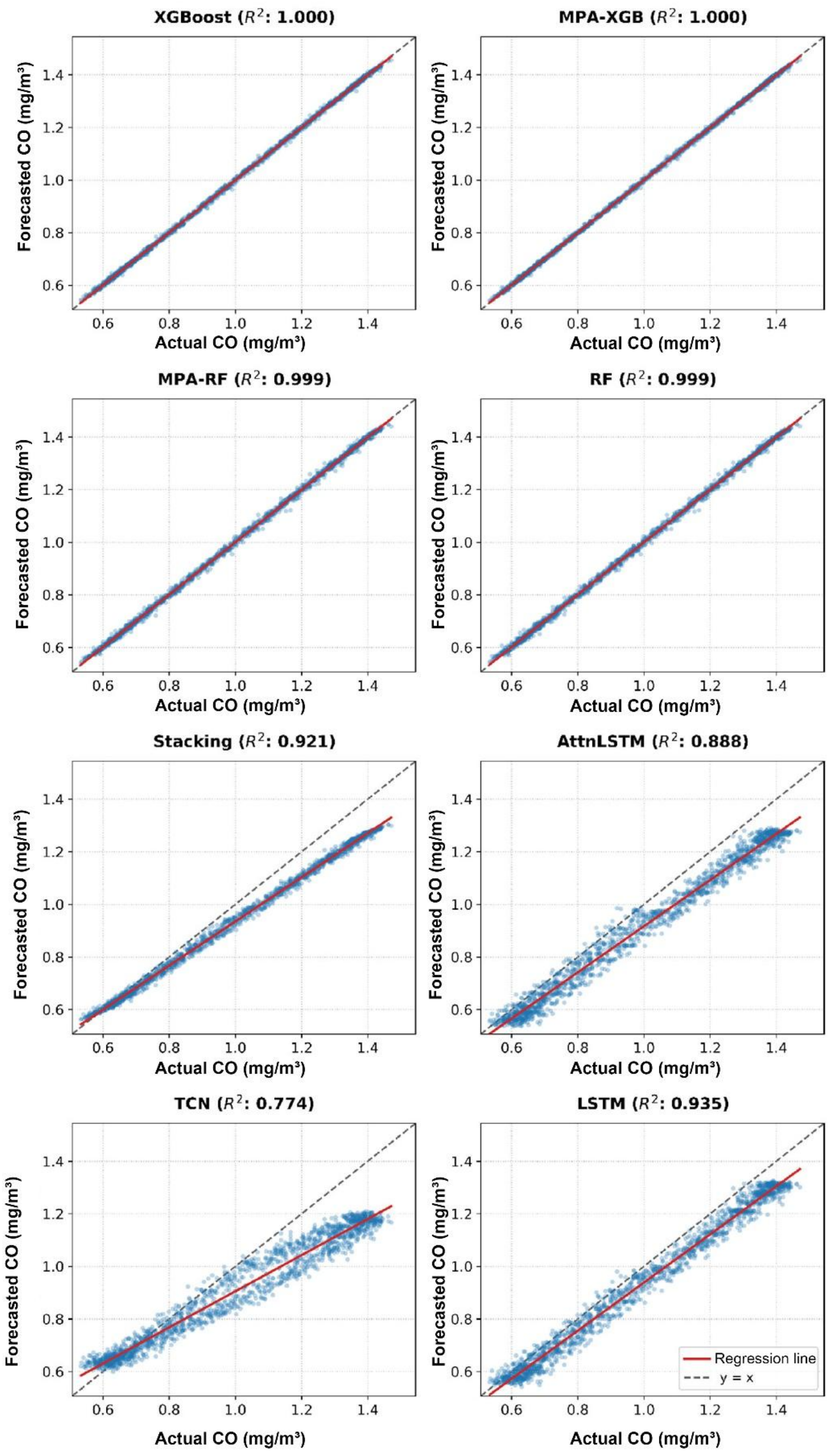
(b) Correlation between actual and predicted $PM_{2.5}$ values at the +68 m station



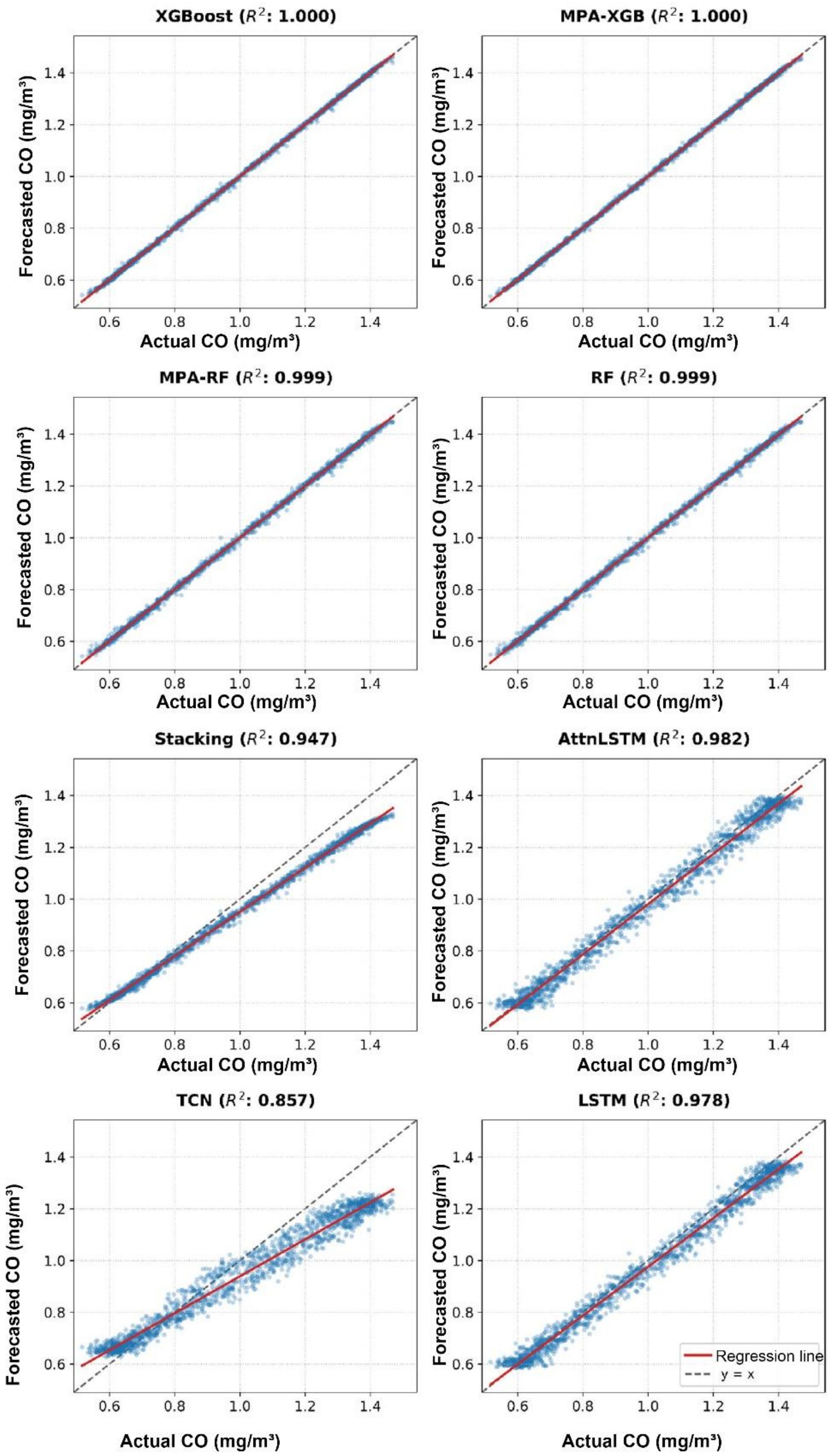
(c) Correlation between actual and predicted $PM_{2.5}$ values at the -58 m station



(d) Correlation between actual and predicted CO values at the +102 m station



(e) Correlation between actual and predicted CO values at the +68 m station



(f) Correlation between actual and predicted CO values at the -58 m station

Figure 9. Correlations between actual and predicted $PM_{2.5}$ and CO values at the monitoring stations

4.5. Procedure for proactive operational decision support

Forecasts are not used to control equipment fully automatically. The system follows four steps: data-quality verification; concentration forecasting over a 1–3 h horizon; comparison with appropriate limits according to averaging time and applicability; and issuance of recommendations for review by authorized personnel. When an increase is forecast at +68 m, priority is given to enhanced dust suppression and reduced truck density on the relevant route. When accumulation is forecast at the -58 m pit bottom, access to the affected area should be restricted, loading-haulage operations adjusted, and high-emission activities reviewed. Before blasting, forecasts are combined with meteorological information to select the timing and scope of control measures.

Table 7. Alert-level and operational-recommendation matrix

Level	Activation condition	Operational recommendation	Responsibility
Normal	Valid data; forecast below the pre-alert threshold	Maintain production; continue monitoring and logging	Operations control room
Watch	Concentration approaches the limit, trend increases, or wind is unfavorable	Increase inspections; target dust suppression to the relevant bench/route; reduce truck density; review the blasting schedule	Safety-environment and operations teams
Alert	Forecast reaches/exceeds the limit after adjustment for averaging time	Adjust or postpone high-emission activities; reroute traffic; restrict access; intensify mitigation	Shift manager/authorized person
Data fault	Connection loss, abnormal sensor, or model outside its validity domain	Do not use the forecast; switch to backup equipment or manual measurement	Instrumentation and safety teams

To maintain the system, the dissertation proposes a contingency procedure: short data gaps are flagged and imputed only when they do not span a blasting event; prolonged communication loss requires suspension of automated recommendations and a transition to backup measurements; and changes in pit geometry or equipment fleets require data-drift assessment, retraining, and independent testing. Each model version must retain its dataset, configuration, library versions, test metrics, and deployment approver.

4.6. Chapter 4 summary

Chapter 4 developed and validated an AI modeling system using field measurements. XGBoost and MPA-XGBoost achieved the most stable performance; Stacking provided a combined approach with high R^2 but was not always optimal in terms of error. The results were converted into a human-approved decision-support procedure consistent with the mining-engineering focus of the dissertation.

CONCLUSIONS AND RECOMMENDATIONS

1. Conclusions

– The dissertation systematized the mechanisms of $PM_{2.5}$ and CO generation across the mining-production chain and clarified the effects of thermal stratification, recirculating airflow, and pit geometry on pollutant accumulation in deep open-pit mines.
– A monitoring network was established at elevations +102 m, +68 m, and -58 m, and a 5-min database was developed for 10 April 2024–31 March 2025, integrating meteorological data, production events, mine plans, and equipment records.
– A workflow was developed for quality control, smoothing, CEEMDAN decomposition, interaction-feature and lag-feature construction, chronological data splitting, and leakage-free normalization.
– Eight model architectures were developed and compared. The XGBoost/MPA-XGBoost group achieved high and stable test performance at the three stations; MPA supported configuration search and peak tracking. Stacking Ensemble performed well but did not always outperform XGBoost in RMSE/MAE.
– A 1–3 h forecasting procedure and an operational-recommendation matrix were proposed, comprising normal, watch, alert, and data-fault levels and emphasizing approval by authorized mine personnel.
– The results are valid only for the dataset and three monitoring locations investigated. Application to other mines or new mining elevations requires site-specific data, recalibration, and retesting.

2. Recommendations

– Expand the monitoring network across seasons, wind directions, and newly developed elevations; add NO_x, SO_2, PM_{10}, and TSP to establish a multi-pollutant alert system.
– Combine MPA-XGBoost with Kriging or graph neural networks to develop grid-based forecast maps and model three-dimensional concentration fields.
– Add explainable-AI tools such as SHAP to quantify the time-specific contributions of wind speed, wind direction, haulage load, explosive charge, and lagged variables.
– Extend comparisons to ARIMA/SARIMA and physical models; quantitatively evaluate economic benefits, dust-suppression water use, and occupational-health risks.
– Develop an MLOps framework for data-drift monitoring, automatic versioning, and controlled retraining as the mine deepens or the equipment system changes.

LIST OF PUBLICATIONS RELATED TO THE DISSERTATION

1. **Ngoc, L. T.**, Nguyen, H., Bui, X. N., & Tran, T. T. (2026). *Dynamic weighting and genetic algorithm-based hybrid SVR-Ridge model for improving $PM_{2.5}$ forecasting in open-pit copper mines*. International Journal of Mining, Reclamation and Environment, 40(4), 349–379. https://doi.org/10.1080/17480930.2025.2492086

2. **Le Tuan Ngoc**, Hoang Nguyen, Xuan-Nam Bui, Le Thi Thu Hoa (2025). *RF-PM_{10}Hybrid Model for Real-Time PM_{10} Forecasting in Open-Pit Copper Mine: A Case Study at the Sin Quyen Copper Mine*. Journal of Mining and Earth Sciences, 66(4), 48–69. https://doi.org/10.46326/JMES.2025.66(4).04

3. **Tuan-Ngoc Le**, Nguyen Truc Anh, Hoang Nguyen, Xuan-Nam Bui (2025). *Predicting $PM_{2.5}$ Concentrations in an Open-Pit Copper Mine Using Advanced AI Techniques: Insights from the Sin Quyen Copper Mine, Vietnam*. Proceedings of the International Workshop on Advances in Surface Mining for Environmental Protection and Sustainable Development, 17 October 2025, Hanoi, Vietnam, 249–268.

4. Hoang Nguyen, Tran Dinh Bao, Xuan-Nam Bui, Pham Van Viet, Dinh-An Nguyen, Ngoc-Hoan Do, Le Thi Thu Hoa, Qui-Thao Le, **Tuan-Ngoc Le** (2024). *Measuring and Predicting Blast-Induced Flyrock Using Unmanned Aerial Vehicles and Lévy Flight Technique-Based Jaya Optimization Algorithm Integrated with Adaptive Neuro-Fuzzy Inference System*. Natural Resources Research, 34(9). https://doi.org/10.1007/s11053-025-10455-4.