

Article

Visibility Prediction and Diagnostic Interpretation Based on Comparative Modelling for Sustainable Urban Environmental Management: A Chengdu Study from November 2021 to October 2024

Bin Hu ^{1,*}, Haiming Fan ^{2,*}, Yushuai Wei ¹, Shangqing Zhang ^{2,3} and Hongyi Zhao ¹ 

¹ School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China

² Shanxi Province Key Laboratory of Metallogeny and Assessment of Strategic Mineral Resources, Taiyuan 030006, China

³ College of Geoscience and Surveying Engineering, China University of Mining and Technology, Beijing 100083, China

* Correspondence: bhu@cugb.edu.cn (B.H.); fanhming@aliyun.com (H.F.)

Abstract

Reliable visibility forecasting is important for transportation safety and sustainable urban environmental management, particularly in the Sichuan Basin, where poor visibility remains a persistent concern. This study developed a comparative diagnostic framework for one-day-ahead prediction of continuous daily visibility in Chengdu using meteorological and air-quality observations from November 2021 to October 2024. Grey relational analysis and temporal diagnostics were used to characterize variable associations and temporal dependence. Seven methods—seasonal autoregressive integrated moving average with exogenous variables (SARIMAX), CatBoost, long short-term memory (LSTM), Transformer, convolutional neural network–LSTM (CNN–LSTM), CNN–Transformer, and Transformer–LSTM—were evaluated under four input configurations, yielding 28 model–input combinations. SARIMAX achieved the best overall balance between predictive accuracy and generalization stability. CatBoost was the most accurate data-driven method and obtained the highest coefficient of determination ($R^2 = 0.644$) under meteorological-only inputs. Meteorological-only inputs outperformed pollutant-only and full multivariate inputs for all models. Among the deep-learning models, LSTM performed better under univariate, meteorological-only, and pollutant-only inputs, whereas Transformer benefited more from the full multivariate input. CNN pre-processing improved LSTM mainly under the full multivariate and pollutant-only configurations. The moderate maximum R^2 suggests that routine observations did not fully capture aerosol composition, particle number and size distributions, hygroscopic growth, aerosol–water interactions, and fog-related processes. These findings support model and input selection for daily visibility forecasting.

Keywords: visibility prediction; time-series forecasting; comparative modelling; diagnostic interpretation; Chengdu

Academic Editor: Stephan Havemann

Received: 27 June 2026

Revised: 3 August 2026

Accepted: 3 August 2026

Published: 5 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Atmospheric visibility is an important indicator of atmospheric transparency and near-surface environmental conditions. Low visibility can strongly affect road traffic, aviation, shipping, urban operation, and public safety, and is also closely related to air

quality degradation and public perception of atmospheric pollution [1–4]. However, visibility prediction remains challenging because visibility is jointly influenced by meteorological conditions, pollutant concentrations, aerosol-related properties, and local atmospheric processes, such as relative humidity, temperature, wind conditions, precipitation, particulate matter with aerodynamic diameters of $2.5\text{ }\mu\text{m}$ or less ($\text{PM}_{2.5}$) and $10\text{ }\mu\text{m}$ or less (PM_{10}), boundary-layer conditions, and fog or cloud processes [2,3,5–7]. Many of these factors and processes cannot be fully represented by routine meteorological and air-quality observations alone [7]. Therefore, visibility should be regarded as an integrated atmospheric variable, and its prediction is a complex nonlinear environmental time-series problem.

The Sichuan Basin has long experienced persistently poor visibility, making reliable visibility prediction particularly important despite recent improvements under air-pollution-control measures [8,9]. The relatively poor visibility of the Sichuan Basin has commonly been associated with its enclosed topography, humid climate, and unfavorable atmospheric ventilation, which facilitate the accumulation of moisture and air pollutants [10,11]. Within this regional setting, Chengdu represents a particularly policy-relevant case because it is the provincial capital and a major metropolitan and transportation center in western China, with officially designated international aviation- and comprehensive-transportation-hub functions [12]. Accurate visibility forecasting in Chengdu therefore has direct relevance to urban mobility, airport and road safety, and sustainable environmental management, providing the regional and practical motivation for selecting the city as the study area.

Visibility forecasting methods can be generally grouped into physically based numerical models, traditional empirical or statistical models, machine-learning and deep-learning models, and remote-sensing or multisource-data-based approaches [4,7]. Physically based numerical models, such as numerical weather prediction and atmospheric chemistry models, estimate visibility by representing meteorological fields, pollutant transport, aerosol processes, and related physical or chemical mechanisms [7,13,14]. These models have clear physical foundations and are useful for forecasting applications, but their performance can be affected by uncertainties in initial conditions, model resolution, emission inventories, and parameterization schemes [13,14]. Among these physically based numerical approaches, deterministic fog models further simulate fog formation, development, and dissipation by coupling atmospheric dynamics, radiation, cloud microphysics, and chemical processes. Some models incorporate multiphase chemistry, trace gases, aerosol particle activation and interactions, and aerosol–water interactions. These processes affect aerosol hygroscopic growth, fog-droplet formation, liquid water content, particle and droplet size, and light extinction, thereby linking air pollution, fog processes, and visibility [13–16]. However, such models require detailed information on aerosol chemical composition, particle number and size distributions, trace gases, emissions, and fog microphysics, which is often unavailable in routine monitoring datasets.

Traditional empirical and statistical methods, including regression-based models, model output statistics, and time-series models, are relatively simple and interpretable. However, because they often rely on predefined or simplified relationships, they may have limited ability to capture nonlinear interactions and complex temporal dependence in visibility variation [4]. Statistical post-processing has also been used to improve visibility forecasts derived from numerical models. For example, Chmielecki and Raftery [17] applied Bayesian model averaging to produce probabilistic visibility forecasts using ensemble forecasts and related meteorological variables.

With the increasing availability of environmental observations and computational

tools, data-driven methods have been widely used for visibility forecasting. Early studies applied neural networks as nonlinear post-processors of surface observations and numerical-model outputs for ceiling and visibility forecasting [18]. Machine-learning methods, such as support vector machines, random forests, gradient boosting models, and other tree-based algorithms, have improved the ability to capture nonlinear relationships between visibility and meteorological or pollutant variables [2,5]. For example, Yu et al. [2] developed a machine-learning-based fusion model using multisource information, Zhang et al. [5] compared several machine-learning algorithms for visibility prediction, and Zhen et al. [19] combined Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) through a stacking strategy. These studies illustrate the development of data-driven visibility forecasting from individual nonlinear algorithms toward ensemble and multisource fusion approaches. Deep-learning models, including recurrent neural networks, long short-term memory (LSTM) networks, gated recurrent units (GRUs), convolutional neural networks (CNNs), Transformer-based models, and hybrid architectures, further extend data-driven forecasting capabilities by learning temporal patterns and complex feature representations from sequential data [1,3,6,20].

In addition, remote sensing, image-based observations, numerical model outputs, and multisource fusion strategies have been introduced to provide richer information for visibility estimation and forecasting. For example, Babari et al. [21] developed a model-driven method to estimate atmospheric visibility using ordinary cameras, while Palvanov and Cho [6] used deep CNNs to extract visibility-related information from images. Yu et al. [2] further combined observations and numerical-model information in a machine-learning fusion framework. These approaches expand the information available beyond conventional station-based meteorological and air-quality observations [2,6,21].

Overall, the development of visibility forecasting methods has followed a general trend from physically based simulation and simple statistical relationships toward nonlinear, data-driven, and multisource modelling frameworks [4,7]. However, these approaches are complementary rather than mutually exclusive because physically based models describe atmospheric processes, whereas data-driven models identify predictive relationships from available observations.

Previous studies have substantially advanced visibility forecasting by developing machine-learning, deep-learning, hybrid, and fusion models and by incorporating richer meteorological, air-quality, remote-sensing, and spatial information [1–6,19,20,22–24]. Model comparison in these studies has mainly been used to improve prediction accuracy, identify the best-performing model, or demonstrate the advantages of a newly proposed framework [4,5,23]. Although such comparisons are valuable for operational forecasting, they provide limited insight into why different model classes perform differently on the same task or why the same model responds differently to alternative input-variable structures. Prediction performance depends not only on the sophistication of the forecasting algorithm, but also on how well the model architecture matches the information and temporal structures of the input data [23,25–27].

Historical visibility may contain persistence information inherited from previous meteorological, pollution, and optical conditions [1,20,26], whereas meteorological and pollutant covariates provide additional explanatory information [2,5]. However, additional variables may also introduce redundancy, collinearity, or noise [25], while temporal characteristics such as non-stationarity and short-term dependence may be captured differently by statistical, machine-learning, and deep-learning models [4,26]. Nevertheless, relatively few studies have jointly examined model architecture, input-variable composition, and temporal structure within a unified diagnostic framework [4].

Consequently, it remains unclear whether visibility predictability primarily originates from historical visibility persistence, meteorological or pollutant covariates, or the ability of particular model architectures to extract useful signals from complex and potentially redundant inputs.

To address this limitation, this study develops a comparative and diagnostic framework for one-day-ahead prediction of continuous daily visibility. Variable association analysis and temporal-structure diagnostics are first used to characterize the information contained in the candidate predictors. Statistical, machine-learning, deep-learning, and hybrid deep-learning models are then compared under four input configurations: historical visibility alone, all covariates, meteorological covariates only, and pollutant covariates only. This framework is used not only to diagnose how model architecture, input composition, and temporal structure jointly affect prediction performance, but also to evaluate the relative predictive contributions of historical persistence, meteorological conditions, and pollutant information. By linking these differences to the atmospheric processes governing visibility, the study provides process-oriented interpretation and practical implications for model and input selection in daily visibility forecasting and sustainable urban environmental management. The resulting one-day-ahead daily visibility forecasts may provide advance information for identifying potential low-visibility conditions and support preliminary operational planning, precautionary advisories, and risk-informed decision-making by airport, road-traffic, and urban environmental management authorities in Chengdu.

2. Data and Methods

2.1. Study Area and Data Sources

The study area is located in Chengdu, Sichuan Province, China. The Shuangliu meteorological station (No. 56294099999) and three surrounding air-quality monitoring stations were included in this study, as shown in Figure 1. The Shuangliu station was selected partly because Shuangliu hosts Chengdu Shuangliu International Airport, where reliable visibility monitoring and forecasting are particularly important for aviation operations and safety. The study period extended from 1 November 2021 to 31 October 2024, covering three full years and all four seasons.

The dataset used in this study consists of meteorological variables and air quality variables. The meteorological variables include dew point temperature (DPT), maximum temperature (T_{Max}), minimum temperature (T_{Min}), sea-level pressure (SLP), visibility (VIS), wind speed (WS), maximum wind speed (WS_{Max}), relative humidity (RH), and precipitation (PREC). The meteorological data were obtained from the Global Surface Summary of the Day (GSOD) archive, which is provided by the National Centers for Environmental Information (NCEI) of the National Oceanic and Atmospheric Administration (NOAA) (available at: <https://www.ncei.noaa.gov/data/global-summary-of-the-day/archive/> (accessed on 1 May 2026)). The air quality data were obtained from the China National Urban Air Quality Real-time Publishing Platform, which is operated by the China National Environmental Monitoring Centre (CNEMC) (available at: <https://air.cnemc.cn:18007/> (accessed on 1 May 2026)).

Because visibility was observed at the meteorological station and served as the target variable, the meteorological station was used as the prediction unit. The surrounding air quality monitoring stations were matched to the meteorological station based on proximity, and the air quality variables were then calculated as the arithmetic mean across the matched air quality monitoring stations.

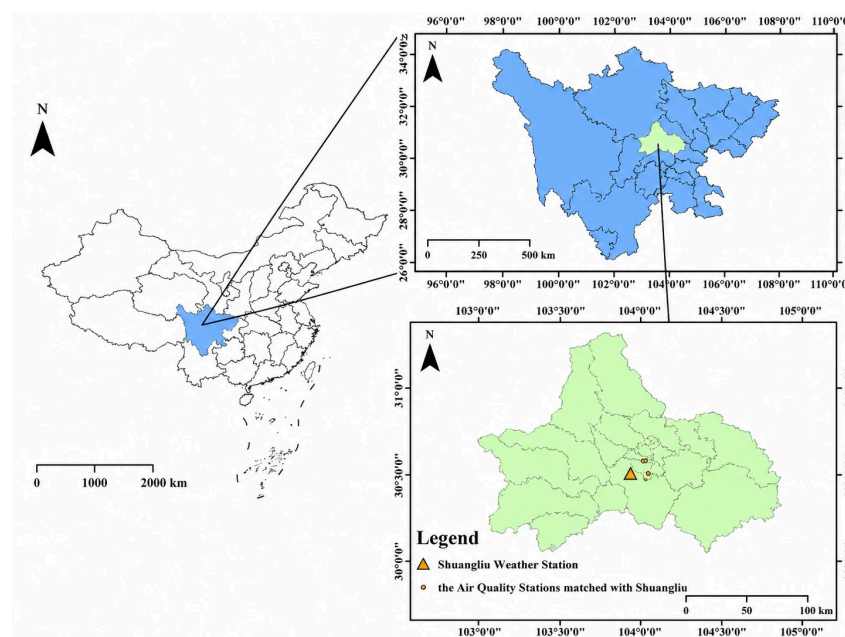


Figure 1. Study area and distribution of stations.

2.2. Descriptive Statistics

Descriptive statistics were calculated for 15 variables, including the target variable and 14 meteorological and air-quality covariates as presented in Section 2.1. The descriptive statistics are summarized in Table 1, including the mean, minimum (Min), maximum (Max), standard deviation (SD), and coefficient of variation (CV), where $CV = (SD / \text{Mean}) \times 100\%$.

Table 1. Descriptive statistics of variables.

Variable	Unit	Mean	Min~Max	Standard Deviation (SD)	Coefficient Variance (CV)
DPT	°C	13.06	−9.60~25.30	7.42	56.85%
T _{Max}	°C	22.84	3.00~41.00	8.78	38.44%
T _{Min}	°C	15.19	−2.00~32.00	7.70	50.67%
SLP	hPa	1059.53	1056.20~1063.20	1.67	0.16%
VIS	km	8.38	1.30~10.00	1.91	22.78%
WS	m/s	2.01	0.31~7.56	0.76	37.86%
WS _{Max}	m/s	4.04	0.98~11.99	1.66	41.14%
PREC	mm	4.43	0.05~44.06	6.95	156.72%
RH	%	69.75	32.88~94.55	11.44	16.40%
CO	mg/m ³	0.72	0.25~2.15	0.26	35.86%
NO ₂	μg/m ³	27.09	5.47~71.43	11.53	42.55%
O ₃	μg/m ³	67.97	7.26~153.00	25.96	38.20%
PM _{2.5}	μg/m ³	29.56	4.50~185.56	23.23	78.56%
PM ₁₀	μg/m ³	71.78	13.04~810.79	66.87	93.16%
SO ₂	μg/m ³	12.26	3.93~35.66	4.67	38.06%

Note: DPT: dew point temperature; SLP: sea-level pressure; VIS: visibility; WS: wind speed; WS_{Max}: maximum wind speed; T_{Max}: maximum temperature; T_{Min}: minimum temperature; PRE: precipitation; RH: relative humidity. For all variables except precipitation, $n = 1089$. For precipitation, only rainy records with $PRE > 0$ were included in the descriptive statistics ($n = 466$).

As shown in Table 1, visibility (VIS) had a mean of 8.38 km, ranging from 1.30–10.00 km with an SD of 1.91 km and a CV of 22.78%, indicating moderate variability.

Among meteorological covariates, DPT showed relatively high variability (CV = 56.85%), followed by T_{Min} (50.67%). T_{Max} , WS, and WS_{Max} exhibited moderate variability, whereas SLP was highly stable, with a CV of only 0.16%. Relative humidity showed comparatively low variability. Precipitation (PREC), calculated only for rainy days ($n = 466$), exhibited the highest variability, with a mean of 4.43 mm, a range of 0.05–44.06 mm, and a CV of 156.72%.

Among the air pollutants, PM_{10} and $PM_{2.5}$ showed substantial variability, with mean concentrations of $71.78 \mu\text{g}/\text{m}^3$ and $29.56 \mu\text{g}/\text{m}^3$ and CV values of 93.16% and 78.56%, respectively. The mean concentrations of O_3 , NO_2 , and SO_2 were $67.97 \mu\text{g}/\text{m}^3$, $27.09 \mu\text{g}/\text{m}^3$, and $12.26 \mu\text{g}/\text{m}^3$, respectively. CO had a mean concentration of $0.72 \text{ mg}/\text{m}^3$ and showed the lowest relative variability among the measured pollutants, with a CV of 35.86%.

2.3. Grey Relation

Grey relational analysis (GRA), first proposed within grey system theory by Deng [28], was used to quantify the association between the target variable and the candidate explanatory variables. In GRA, the target variable is defined as the reference sequence, and each candidate explanatory variable is defined as a comparative sequence. The grey relational grade measures the similarity in normalized variation patterns between the reference and comparative sequences, with a larger value indicating a stronger association.

Since these variables have different units and magnitudes, all sequences were first normalized before calculating the grey relational coefficients [29]. The normalization was performed as follows:

$$x'_i(k) = \frac{x_i(k) - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

where $x_i(k)$ is the original value of the i -th sequence at the k -th observation, and $x'_i(k)$ is the normalized value.

After normalization, the absolute difference between reference sequence and each comparative sequence was calculated as:

$$\Delta_i(k) = |x'_0(k) - x'_i(k)| \quad (2)$$

where $x'_0(k)$ represents the normalized value of reference, and $x'_i(k)$ represents the normalized value of the i -th comparative sequence.

The grey relational coefficient was then calculated as:

$$\xi_i(k) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_i(k) + \rho\Delta_{\max}} \quad (3)$$

where $\Delta_{\min}$ and $\Delta_{\max}$ are the minimum and maximum absolute differences among all sequences and observations, respectively. The distinguishing coefficient ρ was set to 0.5.

Finally, the grey relational grade was obtained by averaging the grey relational coefficients across all observations:

$$\gamma_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (4)$$

where n is the number of observations, and γ_i represents the grey relational grade between reference and the i -th comparative variable. A larger γ_i indicates a stronger association between the corresponding candidate explanatory variable and the target variable.

2.4. Temporal Stationarity and Dependence Diagnostics

Before model development, the temporal structure of all variables was evaluated using a two-stage diagnostic procedure [30]. The Augmented Dickey–Fuller (ADF) test first assessed stationarity and non-stationarity. The Ljung–Box test was then applied to stationary series to determine whether they exhibited temporal dependence or were consistent with white noise.

2.4.1. Augmented Dickey–Fuller (ADF) Test

The ADF test [31] was used to examine each series for a unit root. The regression model is expressed as:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^k \delta_i \Delta y_{t-i} + \epsilon_t \quad (5)$$

where y_t is the observation at time t ; Δ is the first-difference operator; α is the intercept; βt is the deterministic trend; γ is the coefficient of the lagged level; δ_i denotes the coefficients of the lagged first differences; k is the lag order; and ϵ_t is a white-noise error term.

The null hypothesis $H_0: \gamma=0$ indicates a unit root and non-stationarity, whereas the alternative hypothesis $H_1: \gamma<0$ indicates stationarity. At a significance level of 0.05, series with $p \leq 0.05$ were classified as stationary; otherwise, they were classified as non-stationary.

The ADF test was applied to all variables. Variables classified as stationary were subsequently evaluated using the Ljung–Box test.

2.4.2. Ljung–Box Test

For each stationary series, the Ljung–Box test [32] was used to assess short-term autocorrelation within the selected lag range.

The Ljung–Box statistic is defined as:

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \quad (6)$$

where n is the sample size, m is the maximum lag considered, and $\hat{\rho}_k$ is the sample autocorrelation at lag k .

The null hypothesis $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ indicates no autocorrelation up to lag m and is consistent with white noise. At a significance level of 0.05, $p \leq 0.05$ indicates significant short-term temporal dependence; otherwise, the series is considered consistent with white noise.

2.5. Time Series Analysis Methods

2.5.1. SARIMAX

Autoregressive Integrated Moving Average (ARIMA) models, developed within the Box–Jenkins framework, are widely used for time series forecasting by capturing temporal dependence in a target variable based on its historical observations [26,33,34]. Seasonal Autoregressive ARIMA (SARIMA) extends ARIMA framework by incorporating fixed-period components, referred to as seasonal components in standard time-series terminology, whereas ARIMAX incorporates exogenous explanatory variables through a regression model with ARIMA errors [26,33,35]. SARIMAX combines these two extensions, enabling the model to account simultaneously for temporal dependence, fixed-period patterns, and covariate effects [36].

The SARIMAX model can be expressed as:

$$\begin{cases} y_t = \beta^T \mathbf{x}_t + u_t \\ \Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D u_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \end{cases} \quad (7)$$

where y_t denotes the target variable to be predicted at time t ; $\mathbf{x}_t$ denotes the vector of exogenous explanatory variables at time t ; β is the corresponding vector of regression coefficients; and $\beta^T \mathbf{x}_t$ represents the linear contribution of the exogenous variables. The term u_t represents the residual component after removing the effects of the exogenous variables, and ε_t is the random error term, which is assumed to follow a white-noise process. B is the backshift operator, defined as $By_t = y_{t-1}$.

In Equation (7), p , d , and q denote the orders of the non-seasonal autoregressive, differencing, and moving-average components, respectively, while P , D , and Q denote the corresponding seasonal orders. The parameter s represents the seasonal period. Furthermore, $\phi_p(B)$, $\theta_q(B)$, $\Phi_P(B^s)$, and $\Theta_Q(B^s)$ denote the non-seasonal autoregressive polynomial, non-seasonal moving-average polynomial, seasonal autoregressive polynomial, and seasonal moving-average polynomial, respectively. The terms $(1-B)^d$ and $(1-B^s)^D$ represent the non-seasonal and seasonal differencing operators, respectively.

2.5.2. CatBoost

CatBoost (python 3.11.2) is a gradient boosting decision-tree algorithm that builds an additive ensemble of regression trees to approximate complex nonlinear relationships between predictors and a target variable [37,38]. Compared with conventional gradient boosting methods, CatBoost adopts ordered boosting and symmetric tree structures, which help reduce prediction shift and improve model stability.

Although CatBoost is not a native sequential model, it can be applied to time-series forecasting by transforming temporal observations into a supervised-learning format. CatBoost has also recently been applied to air-quality analysis to capture nonlinear relationships and interactions among meteorological and pollutant variables [39]. Let y_t denote the target variable at time t , and let $\mathbf{x}_t$ denote the vector of covariates. The com-

bined feature vector at time t can be written as:

$$z_t = \begin{bmatrix} y_t, x_t^T \end{bmatrix}^T \quad (8)$$

Given a look-back window of length L , the lagged input vector is constructed as:

$$Z_t^{(L)} = \begin{bmatrix} z_t^T, z_{t-1}^T, \dots, z_{t-L+1}^T \end{bmatrix}^T \quad (9)$$

The one-step-ahead prediction is then expressed as:

$$\hat{y}_{t+1} = F \left(Z_t^{(L)} \right) \quad (10)$$

where $F(\cdot)$ denotes the CatBoost regression function. The additive boosting structure can be written as:

$$F_M(Z) = F_0(Z) + \eta \sum_{m=1}^M f_m(Z) \quad (11)$$

where F_M is the final ensemble model after M boosting iterations, F_0 is the initial prediction, η is the learning rate, and $f_m(\cdot)$ denotes the regression tree added at iteration m . Through iterative residual fitting, CatBoost captures nonlinear effects and interactions among lagged target values and covariates.

2.5.3. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are recurrent neural networks designed to model sequential dependencies and long-term memory in time-series data [40,41]. By using input, forget, and output gates, LSTM selectively updates its memory state and alleviates the vanishing-gradient problem encountered in conventional recurrent neural networks. This makes LSTM suitable for forecasting tasks in which the future target value is influenced by historical observations and time-varying covariates. LSTM-based models have also been widely used in environmental and air-quality forecasting, where temporal persistence and delayed effects are important [42].

In the modelling framework, a fixed look-back sequence containing historical values of the target variable and associated covariates is used as the model input. The LSTM layer learns temporal dependencies from this sequence, and the learned representation is passed to a dense output layer to generate the next-step prediction. Figure 2 illustrates the structure of the LSTM model.

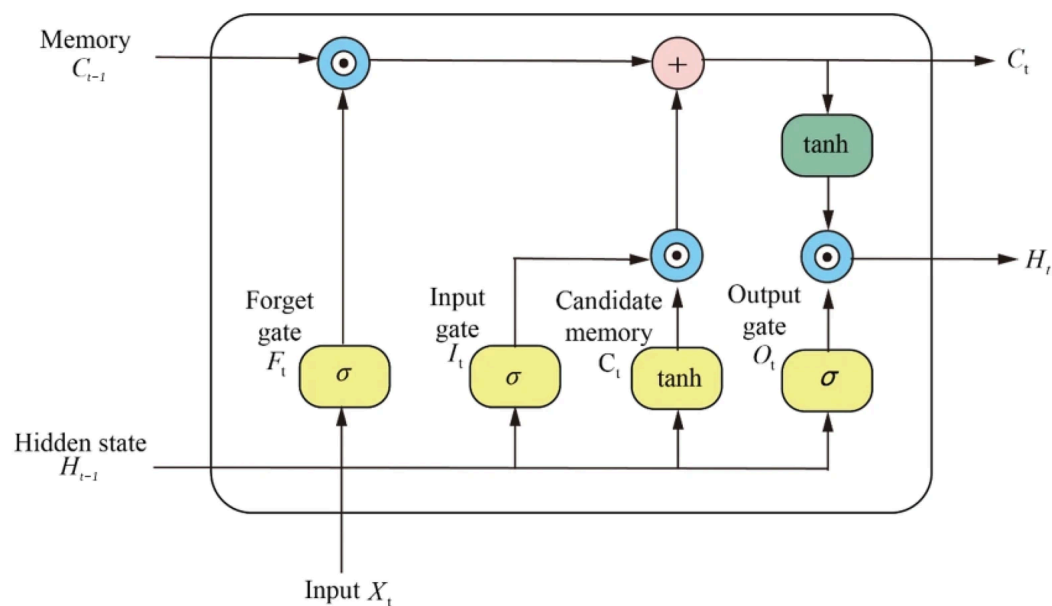


Figure 2. The architecture of the LSTM model [1].

2.5.4. Transformer

The Transformer is a deep learning model based on the self-attention mechanism. Unlike recurrent models, it processes the whole input sequence in parallel and uses attention weights to identify important time steps. This allows the model to capture long-range temporal dependencies more efficiently [43]. Transformer-based models have increasingly been applied to environmental time-series forecasting because they can represent complex temporal interactions between the target variable and covariates [30,42].

In the modelling framework, the input sequence is first embedded and combined with positional information to preserve temporal order. The multi-head attention module then learns relationships among different time steps, while the feed-forward layers transform the attention-based representation into the final predictive output. Figure 3 presents the overall Transformer structure.

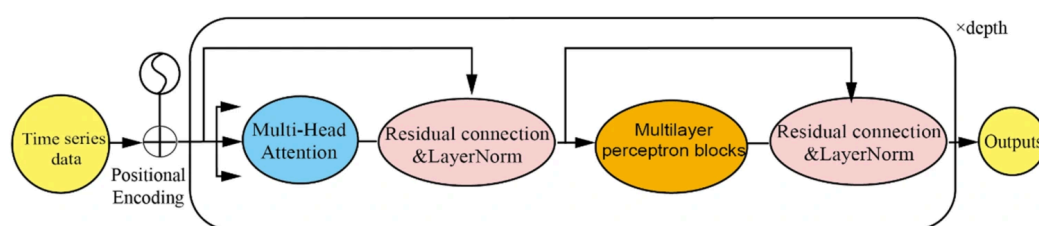


Figure 3. The architecture of the Transformer model [42].

2.5.5. CNN-LSTM

CNN-LSTM is a hybrid architecture that combines convolutional neural networks (CNNs) and LSTM networks for sequential modelling [40,42,44]. The CNN component is used to extract short-term local patterns from the input sequence, while the LSTM component learns temporal dependencies from the extracted features. This combination is useful for time-series forecasting because local fluctuations and sequential dependence can be modelled within the same framework.

In the modelling framework, convolutional layers first operate on the look-back sequence to identify local temporal patterns and interactions among covariates. The extracted feature maps are then passed to the LSTM layer, which captures sequential conti-

nuity and produces a compact temporal representation for the final prediction. Figure 4 shows the framework of the CNN-LSTM model.

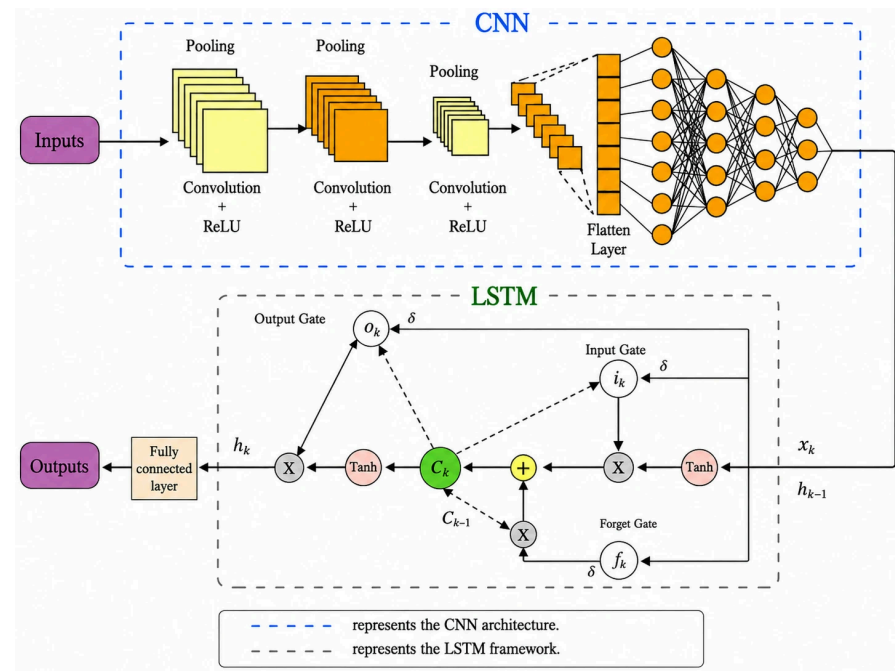


Figure 4. The architecture of the CNN-LSTM model [42].

2.5.6. CNN-Transformer

CNN-Transformer combines the local feature extraction ability of CNNs with the global dependency modelling capacity of Transformer networks [42,43]. The CNN module can identify localized temporal patterns and covariate interactions in the input sequence, whereas the Transformer module uses self-attention to learn long-range dependencies across the sequence [43]. This hybrid design is suitable for complex time-series forecasting tasks in which both short-term fluctuations and long-term temporal relationships may contribute to the predicted target value [42].

In the modelling framework, the CNN module first transforms the look-back sequence into higher-level local features. These features are then fed into the Transformer module, where attention weights are used to emphasize informative time steps and global sequence relationships. Figure 5 presents the architecture of the CNN-Transformer model.

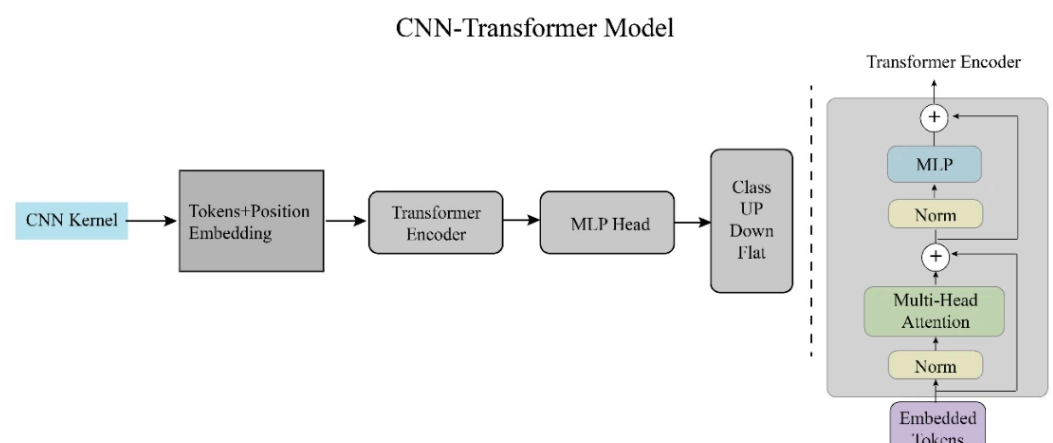


Figure 5. The architecture of the CNN-Transformer.

2.5.7. Transformer–LSTM

Transformer–LSTM is a hybrid model that integrates the attention-based feature extraction capacity of the Transformer with the sequential memory mechanism of LSTM [40,42,43]. The Transformer component is designed to capture global temporal relationships and identify important time steps through self-attention, whereas the LSTM component further model sequential continuity in the learned features.

In the modelling framework, the input sequence is first processed by the Transformer module to obtain attention-based temporal representations. These representations are then passed to the LSTM layer, which summarizes the sequential information and produces the final predictive representation. This design allows the model to combine global attention with recurrent memory for next-step time-series forecasting. Figure 6 shows the architecture of the Transformer–LSTM model.

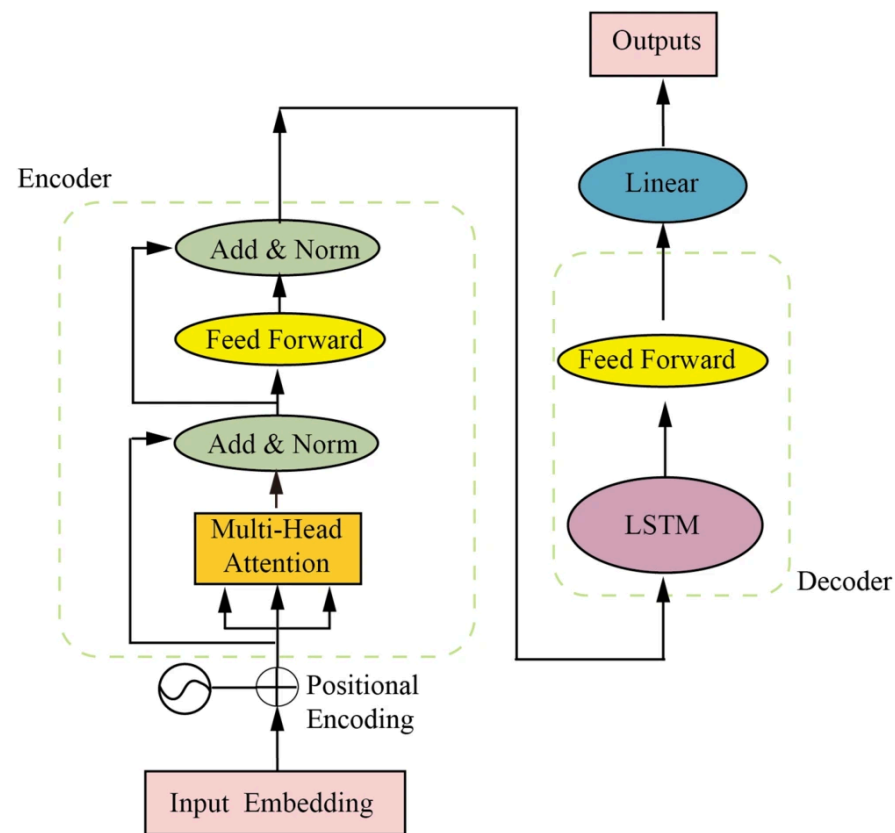


Figure 6. The architecture of the Transformer–LSTM model [38].

2.6. Evaluation Metrics

The prediction performance of different models was evaluated using three commonly used metrics: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) [30,39,45–47].

2.6.1. R-Square (R^2)

R^2 measures the proportion of variance in observed visibility explained by the model. It is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

where y_i is the observed visibility, $\hat{y}_i$ is the predicted visibility, $\bar{y}$ is the mean of observed visibility, and n is the number of observations. A higher R^2 indicates stronger explanatory ability.

2.6.2. Root Mean Square Error (RMSE) and RMSE%

RMSE measures the average magnitude of prediction errors and gives larger errors greater weight. It is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

where y_i , $\hat{y}_i$, and n are as defined above. A lower RMSE indicates smaller prediction error.

RMSE% normalizes RMSE by the mean of actual values, expressed as percentages, to account for varying $PM_{2.5}$ scales across scenarios. It is defined as:

$$RMSE\% = 100 \times RMSE / \bar{y} \quad (14)$$

where $\bar{y}$ is the mean of actual values, as defined above.

2.6.3. Mean Absolute Error (MAE) and MAE%

MAE measures the average absolute difference between observed and predicted values. It is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

Compared with RMSE, MAE is less sensitive to extreme errors and is easier to interpret. A lower MAE indicates better predictive accuracy.

MAE% normalizes MAE by the mean of actual values, expressed as percentages, to account for varying $PM_{2.5}$ scales across scenarios. It is defined as:

$$MAE\% = 100 \times MAE / \bar{y} \quad (16)$$

where $\bar{y} = (1/n) \sum_{i=1}^n y_i$ is the mean of actual values.

3. Results

3.1. Grey Relations Between Covariates and Visibility

Grey relational analysis (Section 2.3) was used to evaluate the associations between visibility and the candidate covariates using the dataset from Shuangliu meteorological station. In this analysis, visibility (VIS) was used as the reference sequence, while dew point temperature (DPT), maximum temperature (T_{Max}), minimum temperature (T_{Min}), relative humidity (RH), sea-level pressure (SLP), wind speed (WS), maximum wind speed (WS_{Max}), precipitation (PREC), $PM_{2.5}$, PM_{10} , CO, NO₂, O₃, and SO₂ were used as comparative sequences.

As shown in Table 2, the grey relational grades ranged from 0.411 to 0.728, indicating varying degrees of association between the selected covariates and visibility. DPT showed the highest grey relational grade with VIS (0.728), followed by T_{Max} (0.664), T_{Min}

(0.646), and RH (0.625). Furthermore, O₃, SLP, NO₂, and WS_{Max} all exhibited moderate associations with visibility, with grey relational grades above 0.5. SO₂, CO, WS, PM_{2.5}, PM₁₀, and PRE still showed relatively notable associations, with grey relational grades ranging between 0.4 and 0.5.

Table 2. Grey relational grades between visibility and candidate covariates.

	DPT	TMax	TMin	RH	O ₃	SLP	NO ₂	WSMax	SO ₂	CO	WS	PM _{2.5}	PM ₁₀	PREC
VIS	0.728	0.664	0.646	0.625	0.582	0.570	0.515	0.507	0.496	0.487	0.486	0.454	0.426	0.411

Overall, the results indicate that these candidate variables exhibit meaningful associations with visibility, supporting their use as covariates in subsequent predictive modeling.

3.2. Temporal Structure of Variables

Following the grey relational analysis in Section 3.1, the variables retained as candidate predictors were further examined to diagnose their temporal structures before being incorporated into the subsequent forecasting models. Specifically, the temporal structure of each retained variable was evaluated using the two-stage diagnostic procedure described in Section 2.4.

First, the ADF test was used to classify the variables as stationary or non-stationary, as shown in Table 3. Non-stationary variables may contain persistent stochastic behavior. Stationary variables were retained for further assessment of series dependence.

Second, the Ljung–Box test was applied to stationary variables identified by the ADF test to determine whether those stationary variables exhibited significant short-term dependence or behaved as white-noise processes, and the results are shown in Table 4.

Table 3. ADF test results for all variables.

Variable	ADF Statistic	<i>p</i> -Value	Stationarity
DPT	−1.956	0.306	Non-stationary
SLP	−2.041	0.269	Non-stationary
T _{Max}	−2.382	0.147	Non-stationary
T _{Min}	−1.822	0.369	Non-stationary
O ₃	−2.380	0.147	Non-stationary
PM _{2.5}	−2.735	0.068	Non-stationary (borderline)
VIS	−4.037	0.0012	Stationary
WS	−14.382	<0.001	Stationary
WS _{Max}	−4.781	<0.001	Stationary
PREC	−4.622	<0.001	Stationary
RH	−9.015	<0.001	Stationary
CO	−3.375	0.011	Stationary
NO ₂	−3.441	0.0096	Stationary
PM ₁₀	−4.157	<0.001	Stationary
SO ₂	−4.312	<0.001	Stationary

At the 5% significance level, DPT, SLP, T_{Max}, T_{Min}, O₃, and PM_{2.5} were classified as non-stationary; PM_{2.5} was borderline (*p* = 0.068). In contrast, VIS, WS, WS_{Max}, PREC, RH,

CO, NO₂, PM₁₀, and SO₂ were stationary and were therefore evaluated using the Ljung–Box test.

Table 4. Ljung–Box test results for stationary variables.

Variable	Q-Statistic	p-Value	Temporal Dependence
VIS	1879.87	<0.001	Significant
WS	301.10	<0.001	Significant
WS _{Max}	185.54	<0.001	Significant
PREC	287.91	<0.001	Significant
RH	1056.42	<0.001	Significant
CO	2067.64	<0.001	Significant
NO ₂	1806.93	<0.001	Significant
PM ₁₀	649.00	<0.001	Significant
SO ₂	451.53	<0.001	Significant

For all stationary variables, the Ljung–Box test rejected the null hypothesis of no autocorrelation (all $p < 0.001$). Thus, none of the stationary series behaved as white noise; instead, each exhibited significant short-term temporal dependence.

Overall, the two-stage analysis distinguished variables with non-stationary long-run temporal structures from stationary variables characterized by significant short-term dependence.

3.3. Prediction Performances of Different Models

The temporal diagnostics in Section 3.2 showed that visibility and the candidate co-variates contained either persistent or significant short-term temporal structure. Together with the grey relational results, these findings supported their use in the subsequent forecasting experiments.

Seven forecasting methods were developed and compared representing different modelling paradigms. SARIMAX was used as the statistical method; CatBoost represented the machine-learning method; and LSTM, Transformer, CNN–LSTM, CNN–Transformer, and Transformer–LSTM represented deep-learning or hybrid deep-learning methods.

Model development followed a chronological design. The first two study years were used for training and model selection, and the third year was retained as an independent test set. All models used a fixed seven-day look-back window for one-day-ahead forecasting. For the deep-learning models, the last 15% of the training sequences was used for time-ordered validation, and training employed the AdamW optimizer, mean squared error loss, a batch size of 16, a maximum of 300 epochs, and early stopping.

The basic framework of each deep-learning model was kept unchanged across the four input configurations, while the hyperparameter settings were adjusted according to input dimensionality and validation performance. The main differences in hyperparameter settings occurred between the univariate and full multivariate configurations, whereas the meteorological-only and pollutant-only configurations used the same settings for each model because their input dimensions were relatively similar. Early stopping was applied to all machine-learning and deep-learning models. For CatBoost, the number of boosting iterations was determined using a time-ordered validation subset, after which the model was refitted using the complete training period. SARIMAX specifications were selected according to the minimum corrected Akaike information criterion

(AICc) calculated from the training data, using candidate models with and without a fixed seven-day periodic component.

For each method, four input configurations were evaluated: a univariate configuration using only the historical time series of visibility, a full multivariate configuration using all meteorological and air-pollutant covariates, a meteorological-only covariate configuration, and an air-pollutant-only covariate configuration. Therefore, a total of 28 model configurations were tested.

Model performance was evaluated using test-set R^2 , RMSE%, and MAE%, together with the test-to-training RMSE ratio. For both RMSE% and MAE%, normalization was based on the mean observed visibility of the corresponding test set. Higher R^2 and lower RMSE% and MAE% indicate better predictive performance. The test-to-training RMSE ratio was used as an empirical indicator of the generalization gap and potential overfitting. A ratio of 0.90–1.10 was interpreted as a negligible generalization gap, 1.10–1.20 as a mild gap, 1.20–1.50 as a noticeable gap, and values above 1.50 as a substantial gap.

Table 5 compares the univariate and full multivariate configurations, whereas Table 6 compares the meteorological-only and pollutant-only configurations.

Table 5. Prediction performance under full multivariate and univariate input configurations.

Model	Multivariate Models with All Covariates				Univariate Models with Only VIS			
	R^2	RMSE%	MAE%	Test-to-Training RMSE Ratio	R^2	RMSE%	MAE%	Test-to-Training RMSE Ratio
SARIMAX	0.621	14.700	11.400	1.050	0.598	15.200	10.900	0.950
CatBoost	0.610	14.900	11.000	1.190	0.599	15.200	11.200	1.060
LSTM	0.575	15.500	10.800	1.190	0.559	15.800	11.600	0.990
Transformer	0.601	15.100	10.900	1.090	0.522	16.500	12.100	1.010
CNN-LSTM	0.596	15.200	11.200	1.090	0.539	16.200	11.900	1.010
Transformer-LSTM	0.530	16.300	11.400	1.100	0.553	15.900	11.400	1.020
CNN-Transformer	0.597	15.100	10.700	1.140	0.564	15.700	11.100	0.990

Table 6. Prediction performance under meteorological-only and pollutant-only input configurations.

Model	Meteorological Covariates Only				Pollutant Covariates Only			
	R^2	RMSE%	MAE%	Test-to-Training RMSE Ratio	R^2	RMSE%	MAE%	Test-to-Training RMSE Ratio
SARIMAX	0.635	14.500	11.100	0.970	0.573	15.700	11.500	1.010
CatBoost	0.644	14.300	10.300	1.090	0.562	15.900	11.600	1.160
LSTM	0.639	14.300	10.300	1.060	0.517	16.600	11.700	1.130
Transformer	0.630	14.500	10.300	1.010	0.332	19.500	14.000	1.060
CNN-LSTM	0.636	14.400	10.300	1.080	0.551	16.000	11.500	1.100
Transformer-LSTM	0.584	15.400	10.200	0.940	0.260	20.500	15.900	1.030
CNN-Transformer	0.623	14.600	10.300	1.060	0.382	18.700	13.400	1.050

Under the full multivariate configuration (Table 5), SARIMAX achieved the highest

R^2 (0.621) and the lowest RMSE% (14.7%), while CatBoost ranked second in terms of R^2 . Under the univariate configuration, SARIMAX and CatBoost showed nearly identical performance, with R^2 values of 0.598 and 0.599, respectively. The inclusion of all covariates improved predictive performance for six of the seven methods, with particularly clear improvements for Transformer and CNN-LSTM. Transformer-LSTM was the only method that did not benefit from the full multivariate input.

The test-to-training RMSE ratios ranged from 0.95 to 1.06 for the univariate models and from 1.05 to 1.19 for the full multivariate models. All univariate models showed negligible generalization gaps. Under the full multivariate configuration, mild gaps were observed for CatBoost, LSTM, and CNN-Transformer, but none of the models exhibited a noticeable or substantial generalization gap.

The meteorological-only configuration consistently outperformed the pollutant-only configuration for all seven methods (Table 6). CatBoost achieved the highest overall R^2 of 0.644, followed closely by LSTM, CNN-LSTM, and SARIMAX. The test-to-training RMSE ratios for the meteorological-only models ranged from 0.94 to 1.09, indicating that the improvement in predictive accuracy was not accompanied by an evident increase in the generalization gap.

Under the pollutant-only configuration, SARIMAX achieved the highest R^2 and the lowest RMSE%, followed by CatBoost and CNN-LSTM. The Transformer-based models generally showed a larger reduction in predictive accuracy under this configuration. Although mild generalization gaps were observed for CatBoost and LSTM, all test-to-training RMSE ratios remained below 1.20.

Across all 28 model-input combinations, the highest R^2 was obtained by CatBoost under the meteorological-only configuration. The meteorological-only input produced the highest R^2 for every method. Overall, SARIMAX provided the most consistent balance between predictive accuracy and generalization stability across the four configurations, whereas CatBoost was the most accurate machine-learning or deep-learning method. None of the evaluated model-input combinations exhibited a noticeable or substantial generalization gap according to the test-to-training RMSE ratio.

4. Discussion

4.1. Overall Robustness and Competitive Performance of SARIMAX

Across the four input configurations, SARIMAX achieved the best overall balance between predictive accuracy and generalization stability. Although some machine-learning or deep-learning models marginally outperformed SARIMAX in an individual metric under specific configurations, such as a slightly higher R^2 or a lower MAE%, these advantages were not consistently maintained across the different input settings. In contrast, SARIMAX remained highly competitive in predictive accuracy under all four configurations and consistently exhibited a small test-to-training RMSE ratio.

The strong overall performance of SARIMAX can be explained by the close correspondence between its model structure and the temporal characteristics of daily visibility. The stationarity and significant short-term autocorrelation of visibility indicate that recent observations contain substantial information for one-day-ahead forecasting. SARIMAX explicitly represents this temporal persistence through autoregressive and moving-average components, while its exogenous-variable component enables meteorological and pollutant information to be incorporated without requiring a highly parameterized nonlinear structure. This combination allows the model to use both the historical dependence of visibility and the additional information supplied by the covariates.

By comparison, the machine-learning and deep-learning models needed to learn temporal dependence, nonlinear relationships, and variable relevance simultaneously

from a relatively limited daily dataset. For one-day-ahead visibility prediction, their greater model complexity did not provide a clear additional advantage because much of the useful predictive information was already contained in the strong short-term dependence of visibility. These models may also have been more affected by noise and redundant covariates. Consequently, the machine-learning and deep-learning models did not consistently outperform SARIMAX in this study.

This result is noteworthy because SARIMAX is a traditional statistical time-series model, whereas the other methods are based on machine-learning, deep-learning, or hybrid architectures. Its strong overall performance shows that greater model complexity does not necessarily lead to better forecasting skill. The temporal structure of daily visibility was well matched by the autoregressive framework of SARIMAX, enabling it to maintain both high predictive accuracy and stable generalization. By contrast, the greater representational capacity of the more complex models did not consistently provide additional benefits under the present data and forecasting conditions. Therefore, forecasting methods should be selected according to the temporal characteristics and information structure of the data, rather than on the assumption that more complex architectures are inherently superior.

4.2. Interpretation and Implications of the Moderate R^2 and Low Prediction Error

Although the evaluated models achieved generally acceptable error levels, the highest R^2 was 0.644. This indicates that the available predictors captured a substantial proportion of daily visibility variation, while a considerable proportion remained unexplained. The moderate R^2 should therefore be interpreted in relation to the information represented by the input data rather than solely as a limitation of the modelling methods. Routine meteorological variables and bulk pollutant concentrations provide useful predictive signals, but they cannot fully describe all physical, chemical, and microphysical processes affecting atmospheric visibility.

Visibility is influenced not only by meteorological conditions and total pollutant mass concentrations, but also by aerosol optical properties, chemical composition, mixing state, hygroscopic growth, aerosol–water interactions, particle number and size distributions, boundary-layer evolution, radiation, and fog and cloud microphysics [7,13,14,48]. Bulk $PM_{2.5}$ and PM_{10} concentrations cannot distinguish among aerosol components with different water-uptake and light-extinction properties. For example, ammonium sulfate, ammonium nitrate, and water-soluble organic matter may contribute differently to visibility degradation, particularly under humid conditions [12,49,50]. Particle number concentration and size distribution can also affect aerosol extinction and visibility [51]. The absence of these variables from the available daily datasets may partly explain the remaining unexplained variation.

The coexistence of a moderate R^2 with relatively low RMSE% and MAE% is not contradictory. R^2 reflects the proportion of observed variability reproduced by a model, whereas RMSE% and MAE% describe the average magnitude of prediction errors. Historical visibility exhibited significant short-term temporal dependence and contained integrated information from preceding meteorological, pollution, and optical conditions. Consequently, the models could reproduce the general level and routine day-to-day variation in visibility with relatively small errors, even though some irregular or abrupt variations remained difficult to explain. Overall, the results indicate that daily visibility has meaningful predictability, but further improvement will depend not only on model development but also on the inclusion of more informative atmospheric variables.

Furthermore, the relatively low prediction errors enhance the practical applicability of the models by enabling reasonably accurate one-day-ahead predictions of continuous

daily visibility. Such prediction information can help airport and road-traffic authorities identify deteriorating visibility conditions in advance, assess the potential severity of low-visibility events, make preliminary operational arrangements, and issue precautionary advisories. It can also support local environmental authorities in targeted monitoring, management-resource allocation, and timely risk-informed responses to adverse atmospheric conditions. Although the framework does not directly improve air quality, it may indirectly contribute to local environmental management by providing more timely and specific decision-support information.

4.3. Predictive Contributions of Meteorological Versus Pollutant Covariates

The ablation results showed that, for all models, the meteorological-only configuration consistently outperformed both the pollutant-only and even full multivariate configurations, yielding higher R^2 and lower RMSE% and MAE%. This finding does not imply that air pollutants are unimportant for visibility. Rather, it suggests that pollutant concentrations alone provide incomplete information for predicting daily visibility.

Visibility degradation depends not only on aerosol loading but also on how meteorological conditions regulate its optical effects [48,52]. For example, at the same $PM_{2.5}$ concentration, high humidity can enhance aerosol hygroscopic growth and light extinction, whereas drier conditions may produce a weaker visibility reduction. Wind and precipitation also influence pollutant dispersion and removal. Meteorological variables therefore provide essential contextual information for relating pollutant concentrations to visibility.

Moreover, bulk $PM_{2.5}$ and PM_{10} concentrations do not fully represent aerosol chemical composition, hygroscopicity, particle properties, or light-extinction efficiency. Specific components, including ammonium sulfate, ammonium nitrate, and water-soluble organic matter, may contribute differently to aerosol water uptake and visibility impairment [49,50]. The superior performance of the meteorological-only configuration therefore suggests that the available meteorological variables provided more directly usable predictive information in this dataset, rather than that meteorology is inherently more important than aerosol pollution. Part of the information provided by pollutant concentrations may also overlap with that already contained in recent visibility and meteorological conditions, which may explain why adding all covariates did not consistently yield further improvements.

4.4. Architecture-Dependent Performance of Machine-Learning and Deep-Learning Models

Among the machine-learning and deep-learning methods, CatBoost maintained consistently competitive performance across all four input configurations. Its tree-based boosting structure can capture nonlinear relationships and variable interactions while prioritizing informative features, making it relatively adaptable to changes in input composition [27,37,38].

The relative performance of LSTM and Transformer was strongly configuration dependent. LSTM performed better under the univariate, meteorological-only, and pollutant-only configurations, indicating that recurrent memory was well suited to short sequences dominated by temporal persistence or coherent predictors [40,41]. Transformer was slightly better under the full multivariate configuration, suggesting that attention [43] became more useful when richer cross-variable and temporal information was available. Thus, the advantage of attention was conditional rather than universal.

CNN feature extraction improved LSTM mainly under the full multivariate and pollutant-only configurations. In these settings, convolution may help organize local temporal patterns, compress heterogeneous inputs, and reduce redundant information before

recurrent modelling [42,44]. Its benefit was limited for univariate and meteorological-only inputs, where the available signals were already relatively direct. CNN also provided inconsistent gains when combined with Transformer, possibly because convolution and attention partly overlapped in their feature-extraction functions.

When CNN and Transformer were used as front-end modules for LSTM, CNN–LSTM generally performed better than Transformer–LSTM under the multivariate configurations. CNN supplied localized and compact features that complemented recurrent memory, whereas attention-based feature mixing may have overlapped with the temporal modelling of LSTM and increased optimization complexity. These findings do not imply that one architecture is universally superior. Rather, model performance depends on whether the architecture matches the input structure and whether the combined modules provide complementary functions.

5. Conclusions

This study developed a comparative and diagnostic framework for daily visibility prediction in Chengdu by integrating variable association analysis, temporal-structure diagnostics, and multi-model comparisons under different input configurations.

The key findings are summarized as follows:

- (1) SARIMAX achieved the best overall balance between forecast accuracy and generalization across the four configurations. This result shows that model selection should be based on temporal dependence and data structure rather than model complexity alone.
- (2) The moderate R^2 values suggest that unmeasured aerosol properties, boundary-layer processes, and fog microphysics limited the explanatory power of the models. However, the relatively low error metrics indicate that daily visibility remained reasonably predictable. These forecasts can support the early identification of deteriorating visibility conditions and inform preliminary planning, precautionary advisories, and risk-informed decision-making by relevant authorities.
- (3) Meteorological-only inputs consistently outperformed pollutant-only inputs, indicating that the meteorological variables provided more direct predictive information, whereas the effects of bulk pollutant concentrations were strongly regulated by meteorological conditions and partly redundant with other inputs.
- (4) CatBoost was the strongest data-driven model because its tree-based structure effectively captured nonlinear interactions and reduced the influence of redundant or collinear covariates.

Author Contributions: B.H.: Writing—original draft, writing—review and editing, validation, methodology, resources, project administration, funding acquisition. H.F.: validation, formal analysis, funding acquisition, conceptualization, resources. Y.W.: validation, data curation, project administration. S.Z.: supervision, visualization, investigation. H.Z.: validation, visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Deep Earth Probe and Mineral Resources Exploration—National Science and Technology Major Project 2024ZD1001200.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are derived from public domain resources. The meteorological data were obtained from the Global Surface Summary of the Day (GSOD) archive provided by the National Centers for Environmental Information (NCEI) of the National Oceanic and Atmospheric Administration (NOAA), available at: <https://www.ncei.noaa.gov/data/global-summary-of-the-da>

y/archive/ (accessed on 1 May 2026). The air quality data were obtained from the China National Urban Air Quality Real-time Publishing Platform operated by the China National Environmental Monitoring Centre (CNEMC), available at: <https://air.cnemc.cn:18007/> (accessed on 1 May 2026).

Acknowledgments: The authors thank Dandan Shuai and Daichi Xu for editing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Deng, T.; Cheng, A.; Han, W.; Lin, H.-X. Visibility forecast for airport operations by LSTM neural network. In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019)*; SciTePress: Setúbal, Portugal, 2019; pp. 466–473. [Google Scholar] [CrossRef]
- Yu, Z.; Qu, Y.; Wang, Y.; Ma, J.; Cao, Y. Application of machine-learning-based fusion model in visibility forecast: A case study of Shanghai, China. *Remote Sens.* **2021**, *13*, 2096. [Google Scholar] [CrossRef]
- Chen, H.; Xu, Y.; Gao, Z.; Kang, J.; Jiang, Y.; Li, Z.; Shen, H. Visibility forecast in Jiangsu province based on the GCN-GRU model. *Sci. Rep.* **2024**, *14*, 12599. [Google Scholar] [CrossRef] [PubMed]
- Salcedo-Sanz, S.; Guijo-Rubio, D.; Pérez-Aracil, J.; Peláez-Rodríguez, C.; Gomez-Orellana, A.M.; Gutiérrez-Peña, P.A. Artificial intelligence-based methods and algorithms in fog and atmospheric low-visibility forecasting. *Atmosphere* **2025**, *16*, 1073. [Google Scholar] [CrossRef]
- Zhang, Y.; Wang, Y.; Zhu, Y.; Yang, L.; Ge, L.; Luo, C. Visibility prediction based on machine learning algorithms. *Atmosphere* **2022**, *13*, 1125. [Google Scholar] [CrossRef]
- Palvanov, A.; Cho, Y.I. VisNet: Deep convolutional neural networks for forecasting atmospheric visibility. *Sensors* **2019**, *19*, 1343. [Google Scholar] [CrossRef] [PubMed]
- Gultepe, I.; Tardif, R.; Michaelides, S.C.; Cermak, J.; Bott, A.; Bendix, J.; Müller, M.D.; Pagowski, M.; Hansen, B.; Ellrod, G.; et al. Fog research: A review of past achievements and future perspectives. *Pure Appl. Geophys.* **2007**, *164*, 1121–1159. [Google Scholar] [CrossRef]
- Chen, Y.; Xie, S. Temporal and spatial visibility trends in the Sichuan Basin, China, 1973 to 2010. *Atmos. Res.* **2012**, *112*, 25–34. [Google Scholar] [CrossRef]
- Zhao, G.; Hu, M.; Zhang, Z.; Tang, L.; Shang, D.; Ren, J.; Meng, X.; Zhang, Y.; Feng, M.; Luo, Y.; et al. Current challenges in visibility improvement in Sichuan basin. *Geophys. Res. Lett.* **2022**, *49*, e2022GL098836. [Google Scholar] [CrossRef]
- Liao, T.; Gui, K.; Jiang, W.; Wang, S.; Wang, B.; Zeng, Z.; Che, H.; Wang, Y.; Sun, Y. Air stagnation and its impact on air quality during winter in Sichuan and Chongqing, southwestern China. *Sci. Total Environ.* **2018**, *635*, 576–585. [Google Scholar] [CrossRef] [PubMed]
- Zhang, L.; Guo, X.; Zhao, T.; Xu, X.; Zheng, X.; Li, Y.; Luo, L.; Gui, K.; Zheng, Y.; Shu, Z. Effect of large topography on atmospheric environment in Sichuan Basin: A climate analysis based on changes in atmospheric visibility. *Front. Earth Sci.* **2022**, *10*, 997586. [Google Scholar] [CrossRef]
- Wang, Q.; Cao, J.; Tao, J.; Li, N.; Su, X.; Chen, L.W.; Wang, P.; Shen, Z.; Liu, S.; Dai, W. Long-term trends in visibility and at Chengdu, China. *PLoS ONE* **2013**, *8*, e68894. [Google Scholar] [CrossRef] [PubMed]
- Lee, H.H.; Chen, S.H.; Kleeman, M.J.; Zhang, H.; DeNero, S.P.; Joe, D.K. Implementation of warm-cloud processes in a source-oriented WRF/Chem model to study the effect of aerosol mixing state on fog formation in the Central Valley of California. *Atmos. Chem. Phys.* **2016**, *16*, 8353–8374. [Google Scholar] [CrossRef]
- Shao, N.; Lu, C.; Jia, X.; Wang, Y.; Li, Y.; Yin, Y.; Zhu, B.; Zhao, T.; Liu, D.; Niu, S.; et al. Radiation fog properties in two consecutive events under polluted and clean conditions in the Yangtze River Delta, China: A simulation study. *Atmos. Chem. Phys.* **2023**, *23*, 9873–9890. [Google Scholar] [CrossRef]
- Pandis, S.N.; Seinfeld, J.H.; Pilinis, C. The smog-fog-smog cycle and acid deposition. *J. Geophys. Res. Atmos.* **1990**, *95*, 18489–18500. [Google Scholar] [CrossRef]
- Forkel, R.; Seidl, W.; Dlugi, R.; Deigle, E. A one-dimensional numerical model to simulate formation and balance of sulfate during radiation fog events. *J. Geophys. Res. Atmos.* **1990**, *95*, 18501–18515. [Google Scholar] [CrossRef]
- Chmielecki, R.M.; Raftery, A.E. Probabilistic visibility forecasting using Bayesian model averaging. *Mon. Weather Rev.* **2011**, *139*, 1626–1636. [Google Scholar] [CrossRef]
- Marzban, C.; Leyton, S.; Colman, B. Ceiling and visibility forecasts via neural networks. *Weather Forecast.* **2007**, *22*, 466–479. [Google Scholar] [CrossRef]

19. Zhen, M.; Yi, M.; Luo, T.; Wang, F.; Yang, K.; Ma, X.; Cui, S.; Li, X. Application of a fusion model based on machine learning in visibility prediction. *Remote Sens.* **2023**, *15*, 1450. [Google Scholar] [CrossRef]
20. Ortega, L.C.; Otero, L.D.; Otero, C.E.; Fabregas, A. Visibility forecasting with deep learning. In *2020 IEEE International Systems Conference (SysCon)*; IEEE: New York, NY, USA, 2020; pp. 1–7. [Google Scholar] [CrossRef]
21. Babari, R.; Hautiere, N.; Dumont, E.; Brémond, R.; Paparoditis, N. A model-driven approach to estimate atmospheric visibility with ordinary cameras. *Atmos. Environ.* **2011**, *45*, 5316–5324. [Google Scholar] [CrossRef]
22. Lu, Z.; Zheng, C.; Yang, T. Application of offshore visibility forecast based on temporal convolutional network and transfer learning. *Comput. Intell. Neurosci.* **2020**, *2020*, 8882279. [Google Scholar] [CrossRef] [PubMed]
23. Ortega, L.C.; Otero, L.D.; Solomon, M.; Otero, C.E.; Fabregas, A. Deep learning models for visibility forecasting using climatological data. *Int. J. Forecast.* **2023**, *39*, 992–1004. [Google Scholar] [CrossRef]
24. Li, W.; Yang, X.; Yuan, G.; Xu, D. ATCNET: A novel approach for predicting highway visibility using attention-enhanced transformer–capsule networks. *Electronics* **2024**, *13*, 920. [Google Scholar] [CrossRef]
25. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [Google Scholar] [CrossRef]
26. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021. [Google Scholar]
27. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6638–6648. [Google Scholar]
28. Deng, J.L. Control problems of grey systems. *Syst. Control Lett.* **1982**, *1*, 288–294. [Google Scholar] [CrossRef]
29. Hu, B.; Zeng, L.; Fan, H. Comparative Study of Four Hybrid Spatiotemporal Models for Daily PM_{2.5} Prediction in the Chengdu–Chongqing Region. *Sustainability* **2026**, *18*, 3126. [Google Scholar] [CrossRef]
30. Hu, B.; Zeng, L.; Fan, H. Interpretable data-driven ozone prediction using statistical diagnostics, XGBoost, SHAP and temporal fusion transformers. *Sustainability* **2026**, *18*, 1009. [Google Scholar] [CrossRef]
31. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. [Google Scholar] [CrossRef]
32. Ljung, G.M.; Box, G.E. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303. [Google Scholar] [CrossRef]
33. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, Revised ed.; Holden-Day: San Francisco, CA, USA, 1976. [Google Scholar]
34. Zhang, L.; Lin, J.; Qiu, R.; Hu, X.; Zhang, H.; Chen, Q.; Tan, H.; Lin, D.; Wang, J. Trend analysis and forecast of PM_{2.5} in Fuzhou, China using the ARIMA model. *Ecol. Indic.* **2018**, *95*, 702–710. [Google Scholar] [CrossRef]
35. Pankratz, A. *Forecasting with Dynamic Regression Models*; John Wiley & Sons: New York, NY, USA, 1991. [Google Scholar] [CrossRef]
36. Lee, C.Y.; Lee, J.Y.; Han, S.H.; Kang, J.G.; Lee, J.B.; Choi, D.R. Performance evaluation of PM_{2.5} forecasting using SARIMAX and LSTM in the Korean Peninsula. *Atmosphere* **2025**, *16*, 524. [Google Scholar] [CrossRef]
37. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [Google Scholar] [CrossRef]
38. Dorogush, A.V.; Ershov, V.; Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv* **2018**, arXiv:1810.11363. [Google Scholar]
39. Zeng, L.; Shuai, D.; Xu, D.; Jing, L. Identifying Meteorological and Gaseous Pollutant Factors Across PM_{2.5} Pollution Levels for Sustainable Air Quality Management in the Beijing–Tianjin–Hebei Region Using CatBoost–SHAP: A 2021–2024 Analysis. *Sustainability* **2026**, *18*, 5611. [Google Scholar]
40. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [Google Scholar]
41. Graves, A. Supervised sequence labelling. In *Supervised sequence labelling with recurrent neural networks*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 5–13. [Google Scholar]
42. Zeng, L.; Dong, R.; Yuan, M.; Jing, L.; Jiao, S. Evaluating deep learning time series models for PM_{2.5} forecasting across diverse horizons. *IScience* **2026**. [Google Scholar]
43. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008. [Google Scholar]

44. Qin, D.; Yu, J.; Zou, G.; Yong, R.; Zhao, Q.; Zhang, B. A novel combined prediction scheme based on CNN and LSTM for urban PM_{2.5} concentration. *IEEE Access* **2019**, *7*, 20050–20059. [Google Scholar] [CrossRef]
45. Sun, X.; Zhao, T.; Liu, D.; Gong, S.; Xu, J.; Ma, X. Quantifying the Influences of PM_{2.5} and Relative Humidity on Change of Atmospheric Visibility over Recent Winters in an Urban Area of East China. *Atmosphere* **2020**, *11*, 461. [Google Scholar] [CrossRef]
46. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [Google Scholar]
47. Nagelkerke, N.J.D. A note on a general definition of the coefficient of determination. *Biometrika* **1991**, *78*, 691–692. [Google Scholar]
48. Won, W.S.; Oh, R.; Lee, W.; Ku, S.; Su, P.C.; Yoon, Y.J.; Bae, M.S. Hygroscopic properties of particulate matter and effects of their interactions with weather on visibility. *Sci. Rep.* **2021**, *11*, 16401. [Google Scholar]
49. Chen, J.; Qiu, S.; Shang, J.; Wilfrid, O.M.; Liu, X.; Tian, H.; Boman, J. Impact of relative humidity and water soluble constituents of PM_{2.5} on visibility impairment in Beijing, China. *Aerosol Air Qual. Res.* **2014**, *14*, 260–268. [Google Scholar]
50. Park, S.M.; Song, I.H.; Park, J.S.; Oh, J.; Moon, K.J.; Shin, H.J.; Lee, G. Variation of PM_{2.5} chemical compositions and their contributions to light extinction in Seoul. *Aerosol Air Qual. Res.* **2018**, *18*, 2220–2229. [Google Scholar]
51. Young, L.H.; Hsu, C.S.; Hsiao, T.C.; Lin, N.H.; Tsay, S.C.; Lin, T.H.; Jung, C.R. Sources, transport, and visibility impact of ambient submicrometer particle number size distributions in an urban area of central Taiwan. *Sci. Total Environ.* **2023**, *856*, 159070. [Google Scholar]
52. Wang, X.; Zhang, R.; Yu, W. The effects of PM_{2.5} concentrations and relative humidity on atmospheric visibility in Beijing. *J. Geophys. Res. Atmos.* **2019**, *124*, 2235–2259. [Google Scholar]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.