# CAP: Causal Air Quality Index Prediction Under Interference with Unmeasured Confounding

Huayi Yang
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
202224131078@stu.cqu.edu.cn

Chunyuan Zheng
School of Mathematical Sciences
Peking University
Beijing, China
cyzheng@stu.pku.edu.cn

Guorui Liao
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
guoruiliao@stu.cqu.edu.cn

Shanshan Huang
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
shanshanhuang@cqu.edu.cn

Jun Liao
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
liaojun@cqu.edu.cn

Zhili Gong
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
gongzl@stu.cqu.edu.cn

Haoxuan Li
Center for Data Science
Peking University
Beijing, China
hxli@stu.pku.edu.cn

Li Liu*
School of Big Data and Software Engineering, Chongqing University
Chongqing, China
dcsliuli@cqu.edu.cn

## Abstract

A significant challenge in air quality index (AQI) prediction is to accurately evaluate the potential outcomes after conducting interventions in pollutant factors such as industrial emissions for each enterprise. Existing methods often suffer from spurious correlations caused by unmeasured confounders and are lack of interpretability of the model, leading to sub-optimal prediction performance. This motivates us to propose a causal AQI prediction framework (CAP) that employs a structural causal model (SCM) to characterize the causal structural variability of various AQI factors for robust AQI prediction. Specifically, we employ the front-door adjustment to explicitly eliminate unmeasured confounders by intervening in industrial emissions from the target enterprise. Meanwhile, we take industrial emissions of neighboring enterprises into account when intervening in the target enterprise and simulate the dispersion of industrial emissions through a Gaussian plume model based on meteorological factors. Experiments on two real-world datasets validate the superior performance of our model on AQI prediction compared to the state-of-the-art baselines.

## CCS Concepts

• **Computing methodologies** → **Machine learning approaches**.

*Li Liu is the corresponding author.


*WWW'25, April 28-May 2, 2025, Sydney, NSW, Australia*

ACM ISBN 979-8-4007-1274-6/25/04
https://doi.org/10.1145/3696410.3714482

## Keywords

Air Quality Index, front-door adjustment, unmeasured confounders, neighbor effects

**ACM Reference Format:**
Huayi Yang, Chunyuan Zheng, Guorui Liao, Shanshan Huang, Jun Liao, Zhili Gong, Haoxuan Li, and Li Liu. 2025. CAP: Causal Air Quality Index Prediction Under Interference with Unmeasured Confounding. In *Proceedings of the ACM Web Conference 2025 (WWW'25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3696410.3714482

## 1 Introduction

Industrial emissions from enterprises have become one of the most serious environmental issues in many industrial cities and regions [4, 18, 35]. However, when an area experiences severe pollution, the government often roughly shuts down all the industrial enterprises within this area to control the air pollution, resulting in significant labor costs and economic losses. In fact, not all enterprises have an impact on current pollution. For example, emissions from industrial enterprises located in the downwind direction do not contribute to current pollution. Therefore, to effectively ameliorate such a strategy of "one-size-fits-all" [8] on pollution prevention and control, it is necessary to accurately predict potential air quality index (AQI), a standard metric for measuring air quality, when an intervention (or treatment) is conducted on the target enterprise.

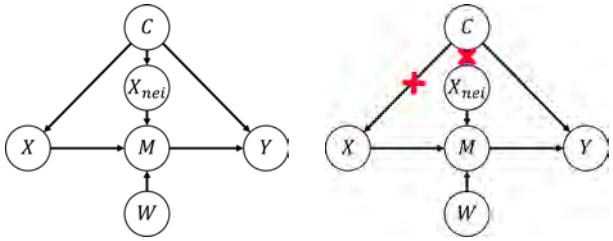
Existing AQI prediction models are usually categorized into three types [11]: statistical models, deep learning models, and hybrid models. From the data perspective, most existing studies usually leverage climate and environmental data to improve AQI prediction accuracy. Although statistical methods [26] work well when the relation between climate and AQI is linear, they are weak in handling massive and complex real-world data. On the other hand, deep learning models and hybrid models [6, 36, 38] can deal with

such cases by learning a spatio-temporal feature representation, but they still suffer from spurious correlation caused by unmeasured confounders, thus leading to sub-optimal prediction performance and cannot discover the further interpretation of pollution causes.

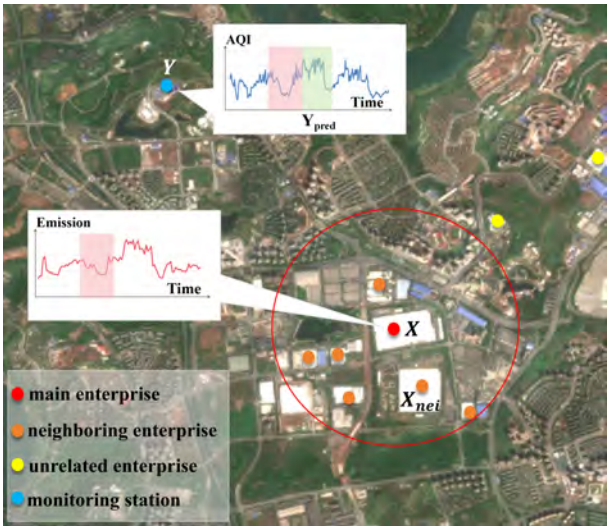
One effective way to address the spurious correlation problem is to learn the true causal relationship between the covariates and the outcome variable. Jiang et al. [13] leverage the Granger causality test (GC) to examine whether one AQI related variable is the cause of another variable, but it still cannot eliminate confounding effects since GC is a statistical approach that cannot account for unmeasured confounders, resulting in unsatisfied prediction accuracy of AQI. A more reasonable causal modeling framework is the structural causal model (SCM) [22, 23], which interprets causal relationships among data to endow models with the capability to quantify causal effects. However, to the best of our knowledge, there is no work that predicts AQI based on SCM. More significantly, existed studies rarely pay attention to industrial emission data, which are often unavailable or indirect, and also may increase confounding biases. Consequently, it is hard to accurately predict the potential AQI after intervening in a certain industrial enterprise (*e.g.*, *shut down or not*).

To fill this gap, we aim to build a causal model based on the SCM framework for AQI prediction that is capable of evaluating the potential AQI after conducting interventions in industrial emissions for each enterprise. Particularly, we consider five key variables when building a causal graph as shown in Fig. 1(a), including industrial emissions $X$, meteorological factors $W$ (*e.g.*, wind vectors), AQI values $Y$, unmeasured confounders $C$ (*e.g.*, complex policy conditions), and mediator variables $M$. Since pollutants emitted by enterprises diffuse in the atmosphere under the influence of the meteorological factors [19], the mediator variable $M$ is required to describe the concentration of pollution in the atmosphere after industrial emissions have diffused to a certain point in the vicinity. In such cases, we can apply a front-door adjustment that could eliminate unmeasured confounders $C$ to predict potential AQI after intervention in industrial emissions. Moreover, in reality, simultaneous emissions from multiple enterprises can collectively affect the AQI measured at the local air quality monitoring station, thus it is necessary to consider the emitting situations of its neighboring enterprises $X_{nei}$ at the same time [10].

In this paper, we propose a causal AQI prediction framework (CAP for short) based on SCM. In particular, our approach considers a principled way of dealing with the inherent causal structural variability [42] to eliminate unmeasured confounders and to achieve robust AQI prediction. Briefly speaking, as shown in Fig. 1(a) and Fig. 1(b), to accurately measure the potential outcomes of intervening in industrial emissions, we employ the front-door adjustment based on causal intervention [22, 23] by cutting off $C \rightarrow (X, X_{nei})$ under an unbiased condition. In particular, complex policy conditions are mentioned as unmeasured confounders $C$. Due to the difficulty of measuring policy conditions and the seriousness of policy implementation, it is almost impossible for us to collect data from a randomized controlled trial by intervening in industrial emissions $X$. However, in order to verify the validity of our approach employing causal intervention to eliminate the effect of unmeasured confounders $C$ on industrial emissions $X$, we partitioned the dataset according to different time periods for this purpose. Taking




**(a) causal graph for AQI with unmeasured confounders** **(b) causal intervention for removing unmeasured confounders**




**(c) Enterprises provide emissions data and monitoring stations provide AQI readings. When the meteorological factors are given, the causal relationship between industrial emissions and air quality (red part) is explained by the model, which leads to a causal prediction of the air quality index (green part).**

**Figure 1: AQI prediction from the perspective of causality. $C$: hidden confounders, $X$: main industrial emissions, $X_{nei}$: neighboring industrial emissions, $W$: meteorological factors, $M$: mediator variables, $Y$ : AQI values.**

$PM_{2.5}$ pollutants as an example, the government tends to tighten the regulation of $PM_{2.5}$ emissions during the fall and winter [9], which is the high period of $PM_{2.5}$ pollutants every year, instead of relatively lax management in the summer. We seize this feature and divide the $PM_{2.5}$ industrial emission data in fall and winter into the training set and the data in summer into the testing set. This gives us a significant distributional difference between the training and test set data, and the predictive performance of the test set will prove the effectiveness and necessity of our method to eliminate the unmeasured confounding factor $C$ [16, 17, 31].

To achieve this, a causal learning model is introduced to estimate the causal effects among observational variables based on the causal graph. Specifically, an AutoEncoder is employed to aggregate industrial emissions information from neighboring enterprises by calculating the prior joint distribution between each pair of neighbors $X$ and $X_{nei}$ with a normal distribution probability density function. It is worth noting that a Gaussian plume model (GPM) is used to

estimate the mediator variable $M$, which denotes the diffusion of industrial emissions. In addition, a temporal convolutional deep network (TCN) is devised to learn the high-level representations of the multi-source information in those time series variables (*e.g.*, wind vectors in meteorological factors $W$). In this way, our causal-based framework is capable of robustly predicting potential AQI after causal intervention on industrial emissions, which is also verified during empirical evaluations on two real-world datasets collected by ourselves from a megacity to be detailed in later sections.

## 2 Related Works

### 2.1 Correlation Methods for AQI

Correlation method for AQI can generally be categorized into three types: statistical models, deep learning models, and hybrid models.

Traditional statistical models have been widely used in air quality prediction due to their simplicity. For example, the ARIMA model [15, 20, 24] has been effective in capturing the temporal dependencies in time-series data, making it suitable for short-term AQI predicting. Similarly, regression models [26] have been applied to model the relationship between AQI and various environmental factors, such as temperature, humidity, and pollutant concentrations. Moreover, BP neural networks [12, 14, 34], a type of artificial neural network, have also been employed to model AQI by learning the nonlinear relationships between input features and the target AQI values.

In contrast, deep learning models and hybrid models have demonstrated remarkable representational capabilities, particularly in capturing the complex patterns underlying air quality data. For instance, Cui et al. [5] design a Transformer-based method for predicting hourly PM2.5 concentrations, leveraging the self-attention mechanism to effectively model long-term temporal dependencies. Similarly, Ni et al. [21] introduce the Gaussian-TCN, which innovatively replaces the traditional activation function in TCN with a Gaussian error current unit, improving the model's ability to handle temporal variations in AQI-related data.

However, one of the main limitations of standalone deep learning models is their focus on capturing only temporal characteristics, often neglecting the equally important spatial correlations present in air quality data. To address this limitation, several studies [7, 39]have proposed hybrid models that integrate spatio-temporal correlations by combining multiple architectures.

For example, Du et al. [6] propose a hybrid framework DAQFF, which integrates CNN and Bi-LSTM. This model processes multivariate air quality-related time-series data from multiple monitoring stations, enabling shared representation learning across spatially distributed features. In addition, Zhao et al. [38] introduce the MASTGN (Multi-Attention Spatio-Temporal Graph Network), which constructs spatio-temporal structural graphs by integrating atmospheric data from multiple monitoring stations. This model employs a multi-attention mechanism to more efficiently mine the intricate spatio-temporal dependencies within the data.

Although the above methods can capture complex patterns in rich spatio-temporal information and show excellent performance in AQI prediction, they neglect unmeasured confounders that may lead the model to learn spurious spatio-temporal correlations, resulting in biased prediction.

### 2.2 Causal Methods for AQI

Causal study is valuable for policy formulation and policy evaluation. In recent years, causal methods have gained increasing attention in the context of Air Quality Index (AQI) analysis. For example, Jiang et al. [13] and Zhu et al. [41] apply Granger causality tests (GC) to identify causal relationships between data, addressing the challenges posed by data diversity. Similarly, Yu et al. [37] employ a Difference-in-difference approach to evaluate the causal impact of policies on AQI outcomes. While these statistical causal methods are effective for analyzing the relationships between observational variables (e.g., dependent and independent variables), they are inherently limited in accounting for unmeasured confounders. As a result, the validity of their findings may be compromised by confounding effects.

To address these limitations, advanced causal inference frameworks have been developed. One effective framework is the Rubin Causal Model (RCM) [25, 28–30, 32, 33, 40]. Weather2vec[27] is a pioneered RCM-based model that successfully eliminates non-local confounding by spatial-varying intervention. While this model treats the meteorology factor as a confounder in studies of air pollution, it is unfortunately rather limited to characterizing dedicated causal relationships compared with treating it as a mediator. Another significant approach is the Structural Causal Model (SCM) [22, 23], which provides a robust framework for interpreting causal relationships within data. Despite its potential, the application of SCM in AQI prediction remains underexplored, particularly in addressing confounding effects through causal intervention.

To bridge this gap, we propose a novel SCM-based framework for AQI prediction that systematically mitigates confounding effects via causal intervention. By explicitly intervening in industrial emissions, our approach enables causal AQI prediction. This methodology not only enhances predictive accuracy but also offers actionable insights for policymakers aiming to mitigate pollution through targeted interventions.

## 3 Task Formulation

Our study introduces a new framework for predicting potential AQI after intervening in industrial emissions. We denote the set of industrial emission data from enterprises in an area as $X = \{X_1, X_2, \ldots, X_n\} \in \mathbb{R}^{n\times t}$, and $X_i = \{x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(t)}\}$, where $x_i^{(t)}$ represents the emission scale from the $i$-th enterprise at time step $t$. The meteorological factors, denoted by $W$, comprises the wind velocity $W_v \in \mathbb{R}^t$ and wind direction $W_d \in (0°, 360°)$, which significantly influence the dispersion of pollutants. The outcome variable, $Y \in \mathbb{R}^t$, corresponds to the AQI, a critical metric for assessing air quality.

Our observational data encompasses temporal sequences of industrial emissions, meteorological conditions, and AQI values, alongside the geographical coordinates of both industrial enterprises and air quality monitoring stations. We formalize the historical input dataset as $\mathcal{X} = \{(X_i, W_v, W_d)|t = 1, 2, \ldots, t_{obs}\}$, where $t_{obs}$ signifies the last observed time step. Subsequently, the future prediction is defined as $\mathcal{Y} = \{Y|t = t_{obs}+1, t_{obs}+2, \ldots, t_{obs}+t_{pred}\}$.

From the probabilistic perspective, the target of predicting is to estimate $P(Y|X, W)$. Conventional data-driven methods parameterize the target distribution as a predicting model $f_\Theta(X, W)$ where $\Theta$

denotes model parameters. These methods learn model parameters from historical data.

**Causal prediction:** Unmeasured confounders often cause models to obtain spurious correlations. To mitigate confounding effects, causal prediction views the predicting problem of likelihood as estimate $P(Y|do(X), W)$, where $do(\cdot)$ denotes the intervention operation [22, 23]. This implies that if we want to know the effect of intervening in industrial emissions of an enterprise on the AQI, causal prediction may be an excellent solution. For example, suppose that at time $t$, the meteorological factors $W = w^t$ and industrial emissions from the enterprise $\mathcal{A}$ are $X_{\mathcal{A}} = x^t_{\mathcal{A}}$, the causal prediction on the AQI values $Y$ by intervening in industrial emissions of the enterprise $\mathcal{A}$ can be computed by $do(X = x^t)$ as $P(Y|do(X = x^t_{\mathcal{A}}), W = w^t)$. However, it cannot be ignored that when an enterprise is emitting pollutants that affect AQI, its neighboring enterprises may be emitting pollutants at the same time, *i.e.*, if the location of the enterprise $\mathcal{B}$ is in the vicinity of the enterprise $\mathcal{A}$, and assuming that industrial emissions of the neighboring enterprise $\mathcal{B}$ are $X_{\mathcal{B}} = x^t_{\mathcal{B}}$, then it is possible that industrial emissions of enterprise $\mathcal{B}$ could have an impact on AQI values $Y$ at the same time as industrial emissions from enterprise $\mathcal{A}$. If we consider only the intervention in industrial emissions of enterprise $\mathcal{A}$, then the calculated $P(Y|do(X = x^t_{\mathcal{A}}), W = w^t)$ will be biased at this point. Therefore, causal prediction takes the intervention in industrial emissions of neighboring enterprises into account and corrects $P(Y|do(X), W)$ to $P(Y|do(X, X_{nei}), W)$.

## 4 Methodology

The overall framework of CAP is shown in Fig.2.

### 4.1 Structural Causal Model

We formulate the causalities among target industrial emissions $X$, neighboring industrial emissions $X_{nei}$, unmeasured confounders $C$ (*e.g.* complex policy conditions), meteorological factors $W$ (*e.g.* wind vectors), mediator variables $M$ and AQI-values $Y$ based on SCM, as shown in Fig. 1(a). The directed edges represent the causal relationships between nodes. Since pollutants are continuously diffused in the atmosphere by meteorological factors, and therefore $M$ is used as the mediator variable to characterize the concentration of pollution in the atmosphere after industrial emissions have diffused to a certain point (air quality monitoring stations) under given meteorological factors $W$.

### 4.2 Causal Intervention via Front-door Adjustment

Unmeasured confounders $C$ is represented as a common cause for $X$, $X_{nei}$ and $Y$ which is often unobservable, we cannot use back-door adjustment according to the path $(X, X_{nei}) \leftarrow C \rightarrow Y$. So we introduce the mediator variable $M$ to apply the front-door adjustment based on the path $C \rightarrow (X, X_{nei}) \rightarrow M$, $M \rightarrow Y \leftarrow C$, and cut off the link $C \rightarrow (X, X_{nei})$. Formally, we have:

$$p(Y|do(X,X_{nei}),W) \overset{(a)}{=} \int_m p(M|do(X,X_{nei}),W)p(Y|do(M),W)dM$$
$$\overset{(b)}{=} \int_m p(M|X,X_{nei},W)dM \int_x \int_{x_{nei}} p(Y|X,X_{nei},M,W)p(X,X_{nei})dXdX_{nei}. \tag{1}$$

In Eq. (1)(a), $p(M|do(X, X_{nei}), W)$ denotes the causal effect of $X$ and $X_{nei}$ on $M$, $p(Y|do(M), W)$ denotes the causal effect of $M$ on $Y$. According to Fig. 1(b), the back-door path $(X, X_{nei}) \leftarrow C \rightarrow Y \leftarrow M$ is $d$-separated by the collider $Y$. Therefore, $p(M|do(X, X_{nei}), W) = p(M|X, X_{nei}, W)$ where $X$, $X_{nei}$, $W$ are all observable values which can be used to calculate $M$. As to $p(Y|do(M), W)$, we can block the back-door path $M \leftarrow (X, X_{nei}) \leftarrow C \rightarrow Y$ without measuring $C$. This is because controlling $C$ is equal to controlling $(X, X_{nei})$ [22]. As such, we can achieve $p(Y|do(M), W)$ by conducting a back-door adjustment over the observable industrial emissions $(X, X_{nei})$. Thus we obtain the causal effect free from the unmeasured confounders $C$, as shown in Eq. (1)(b).

Up to this point, we have liberated $p(Y|do(X, X_{nei}), W)$ from $C$. We then try to consider the prediction on potential AQI from the observational data. According to Eq. (1)(b), to obtain $p(Y|do(X, X_{nei}), W)$, we need to build three suitable backbone models to represent $p(M|X, X_{nei}, W)$, $p(Y|X, X_{nei}, M, W)$ and $p(X, X_{nei})$.

*4.2.1 Modeling $p(X, X_{nei})$.* We initially identify neighboring enterprises within a 5-kilometer radius of the target enterprise $X$. These neighboring enterprises, denoted as $X_{nei} = \{X_1, X_2, ..., X_k\} \in \mathbb{R}^{k \times t}$, where $k$ signifies the nearest $k$ neighboring enterprises and $t$ represents the selected temporal resolution, with each time step equating to one hour.

Subsequently, we aggregate the emission data from these neighboring enterprises. As illustrated in Fig. 2, an AutoEncoder is engaged to distill the essence of the industrial emission features from neighboring enterprises, yielding a low dimensional representation $Z \in \mathbb{R}^t$. We then use the Reconstruction Error (RE) as the loss function to train the AutoEncoder to efficiently extracts features of the input data, expressed as:

$$\mathcal{L}_{recon} = \frac{1}{N} \sum_{n=1}^{N} \| X_{nei} - \hat{X}_{nei} \|^2 . \tag{2}$$

By utilizing $Z$, the computation of the prior probability is reformulated as $p(X, Z)$. Assuming a bivariate normal distribution for the joint distribution of $X$ and $Z$, denoted as $(X, Z) \sim \mathcal{N}(\mu, \Sigma)$, the prior probability is modeled by the probability density function of the bivariate normal distribution:

$$f(X, Z) = \frac{1}{2\pi|\Sigma|} exp\left(-\frac{1}{2}\begin{pmatrix} X - \mu_X \\ Z - \mu_Z \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} X - \mu_X \\ Z - \mu_Z \end{pmatrix}\right), \tag{3}$$

where $\mu = \begin{pmatrix} \mu_X \\ \mu_Z \end{pmatrix}$ and $\Sigma$ are model parameters denoting the mean vector and covariance matrix of the binary normal distribution, respectively, and $f(X, Z) \in \mathbb{R}^t$.

*4.2.2 Modeling $p(M|X, X_{nei}, W)$.* To estimate the mediator variable $M$, we parameterize $p(M|X, X_{nei}, W)$ as $h(X, X_{nei}, W)$. Since $M$ is unobservable, if we use a deep learning approach to estimate $M$, it will increase the parameters that the model needs to learn, which may affect the prediction performance of the model. Therefore, since $M$ represents the pollution level of industrial emission $X$ after it spreads to the vicinity of the air quality monitoring station, we decided to use a physics-based model to simulate the diffusion of industrial emission $X$. Specifically, we use GPM. Assuming uniform and direct wind flow, conservation of mass of pollutants during diffusion and continuous and uniform emissions from the source [1].

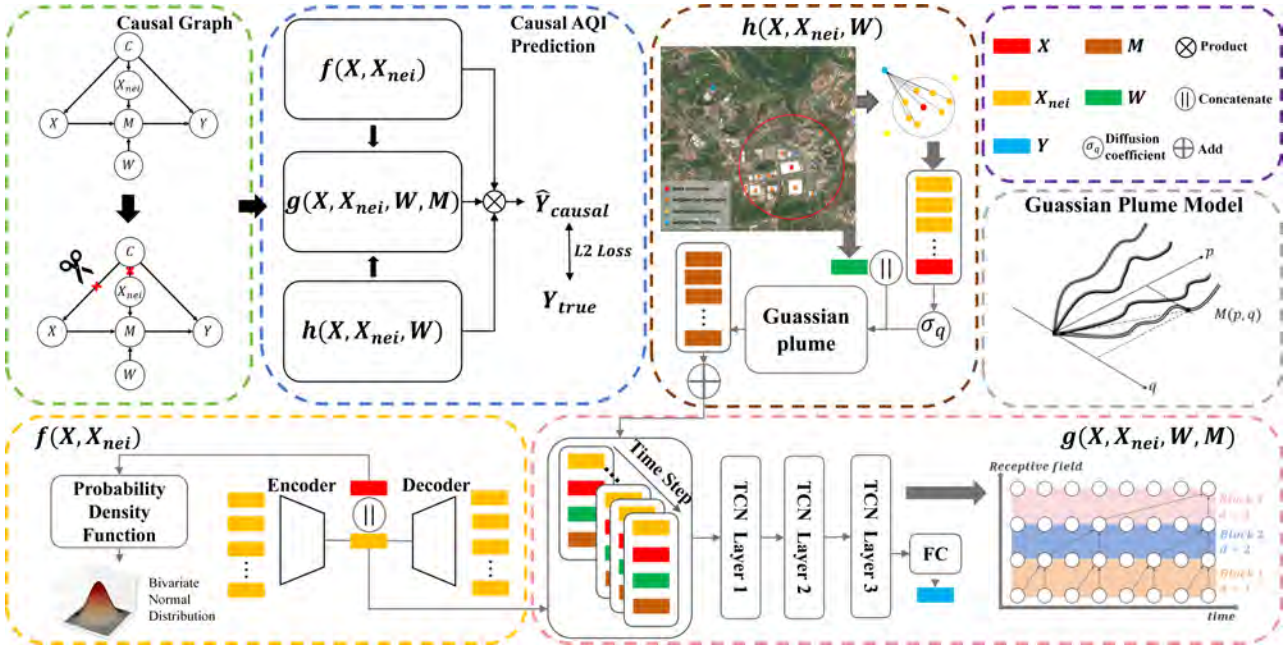



**Figure 2: The architecture of CAP.**

In GPM, emissions downwind disperse laterally from the plume's centerline in a normal distribution pattern. In addition, we omit the effect of elevation, assuming horizontal dispersion only. Therefore, GPM is expressed as:

$$M(p,q) = \frac{X}{W_v}\frac{1}{\sigma_q\sqrt{2\pi}}exp(\frac{-q^2}{2\sigma_q^2}), \tag{4}$$

$$\sigma_q = 0.11p(1+0.0004p)^{-\frac{1}{2}}, \tag{5}$$

where, as depicted in Fig. 2, the $p$-axis direction of the coordinate system is the downwind direction, with $p$ indicating the downwind distance and $q$ representing the crosswind distance. The model-calculated diffusion of industrial emissions is denoted by $M(p,q)$. The atmospheric diffusion coefficient in the $q$-axis direction is represented by $\sigma_q$. Assuming an atmospheric stability class $F$ [2], the diffusion coefficient $\sigma_q$ is detailed in Eq. (5).

As illustrated in Fig. 1(c), both the target enterprise and neighboring enterprises are treated as discrete entities in the pollution dispersion simulation. The aggregate effect of their emissions is captured by the sum $g(X, X_{nei}, W) = \sum M(p,q) \in \mathbb{R}^t$. If the downwind distance $p$ is negative on the $p$-axis, indicating that the enterprise is located downwind of the air quality monitoring station, the $M$-value of the enterprise is set to zero.

*4.2.3 Modeling $p(Y|X, X_{nei}, W, M)$.* This conditional probability function is used to predict the AQI through the industrial emissions $X$, neighboring industrial emissions $X_{nei}$, the meteorological factors $W$, and the mediator variable $M$. We summarize this relationship through a likelihood method, expressing $p(Y|X, X_{nei}, W, M)$ as $g(X, X_{nei}, W, M)$, which can be any backbone models suitable for time series AQI prediction. As shown in Fig. 2, we design the backbone model based on TCN.

TCN is particularly adept at sequence modeling due to their causal convolution strategy, which respects the temporal order of inputs by considering past data for each convolution operation. The input to TCN is a concatenation of data denoted as $(X\|Z\|W\|M)$, where $\|$ symbolizes the concatenation operation. By leveraging dilated convolutions, TCN expand their receptive field, enabling the model to discern both local and global patterns within the data, a critical capability for understanding complex temporal dynamics. Moreover, the TCN architecture includes residual connections, which are instrumental in combating the vanishing gradient problem and allow for the learning of more nuanced and complex representations of temporal data. The full connect layer then synthesizes these features to output $Y$ values that align with the input time step. Specifically expressed as:

$$Y = TCN(X, Z, W, M) \in \mathbb{R}^t. \tag{6}$$

*4.2.4 Estimating $p(Y|do(X, X_{nei}), W)$.* Leveraging our established models, we now proceed to estimate $p(Y|do(X, X_{nei}), W)$. Given the continuous nature of our data, the standard method of summing probability distributions for discrete variables, as outlined in Eq. (1)(b), is inapplicable. Consequently, we undertake a transformation and derivation of Eq. (1)(b) to address this issue, expressed as:

$$\begin{aligned} & p(Y|do(X, X_{nei}), W) \\ & \overset{(a)}{=} \int_m h(X, X_{nei}, W)\, dM \int_x \int_z g(X, Z, M, W) f(X, Z)\, dX dZ \\ & \overset{(b)}{\approx} S \cdot \frac{1}{n}\sum_{i=1}^{n} h(X, X_{nei}, W) f(X, Z) g(X, Z, M, W). \end{aligned} \tag{7}$$

The derivation is explained step-by-step as follows:

**Table 1: Performance comparison with baselines. We bold the best results and underline the best baselines. Note that * means statistical significance with $p$-value $< 0.001$.**

| Dataset | Method | 1-24h | | | | 25-48h | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | RMSE | MAE | SMAPE | MSE | RMSE | MAE | SMAPE |
| $PM_{2.5}$-CQW | MLP | 0.2713 | 0.5102 | 0.3819 | 78.22% | 0.2765 | 0.5160 | 0.3858 | 78.57% |
| | GRU | 0.1120 | 0.3322 | 0.2481 | 54.14% | 0.2314 | 0.4677 | 0.3564 | 61.73% |
| | LSTM | 0.0874 | 0.2929 | 0.2150 | 48.33% | 0.2046 | 0.4040 | 0.3403 | 70.95% |
| | Gaussian-TCN | 0.2207 | 0.4665 | 0.3477 | 67.75% | 0.2152 | 0.4598 | 0.3434 | 68.00% |
| | DAQFF | 0.1519 | 0.3830 | 0.2883 | 64.39% | 0.1705 | 0.4049 | 0.3027 | 68.01% |
| | VMD-EEMD-LSTM | 0.1368 | 0.3644 | 0.2733 | 58.18% | 0.1811 | 0.4075 | 0.3051 | 61.15% |
| | DeepAir | 0.0376 | 0.1928 | 0.1457 | 37.37% | 0.0406 | 0.1982 | 0.1485 | 38.06% |
| | Causal-STGAT | 0.0219 | 0.1657 | 0.1254 | 33.74% | 0.0318 | 0.1768 | 0.1319 | 34.98% |
| | CAP (ours) | **0.0204*** | **0.1388*** | **0.0934*** | **24.43%*** | **0.0310*** | **0.1392*** | **0.0949*** | **25.13%*** |
| $SO_2$-CQE | MLP | 0.2631 | 0.5020 | 0.3761 | 77.89% | 0.2601 | 0.4994 | 0.3737 | 77.86% |
| | GRU | 0.1270 | 0.3543 | 0.2628 | 55.95% | 0.2244 | 0.4615 | 0.3411 | 58.92% |
| | LSTM | 0.0828 | 0.2857 | 0.2099 | 47.35% | 0.2013 | 0.3933 | 0.3179 | 68.13% |
| | Gaussian-TCN | 0.2198 | 0.4656 | 0.3466 | 67.81% | 0.2198 | 0.4661 | 0.3497 | 67.94% |
| | DAQFF | 0.1525 | 0.3819 | 0.2921 | 65.01% | 0.1807 | 0.4011 | 0.3009 | 68.33% |
| | VMD-EEMD-LSTM | 0.1423 | 0.3520 | 0.2647 | 56.14% | 0.1777 | 0.4132 | 0.3180 | 65.15% |
| | DeepAir | 0.0332 | 0.1810 | 0.1364 | 35.70% | 0.0337 | 0.1791 | 0.1441 | 35.82% |
| | Causal-STGAT | 0.0194 | 0.1383 | 0.1050 | 30.06% | 0.0471 | 0.2177 | 0.1761 | 37.94% |
| | CAP (ours) | **0.0164*** | **0.1252*** | **0.0855*** | **22.58%*** | **0.0223*** | **0.1307*** | **0.1022*** | **23.77%*** |

- (a) holds according to the above modeling process and the front-door adjustment.
- (b) is based on the Monte carlo method (MCM), a powerful numerical technique for approximating the value of integrals by random sampling, which can address the challenge of making the integrals in the model difficult to get analytically computed due to the complexity of the model. In Eq. (7)(b), $n$ denotes the number of randomly sampled samples, and $S$ is a constant that denotes the length of the interval, so the existence of $S$ does not affect our final prediction for $Y$, and we can omit it from the computation.

To summarize, we can approximate the probability distribution $p(Y|do(X, X_{nei}), W)$ by Eq. (7)(b), and then for the causal prediction of $Y$ can be expressed as:

$$\hat{Y}_{causal} = h(X, X_{nei}, W) f(X, Z) g(X, Z, M, W), \tag{8}$$

where $\hat{Y}_{causal}$ is the target value for prediction. We use $L2$ loss function to optimize the network model during training to improve the accuracy of causal prediction:

$$\mathcal{L}_{pred} = \frac{1}{n} \sum_{i=1}^{n} (Y_{true} - \hat{Y}_{causal})^2, \tag{9}$$

where $Y_{true}$ denotes the true value of the AQI observed. $\mathcal{L}_{pred}$ denotes the prediction loss of the model, which calculates the difference between the causal predicted value and the true value. Based on all the modules established, our loss function during training can be organized as follows:

$$\mathcal{L} = \mathcal{L}_{recon} + \alpha \mathcal{L}_{pred}, \tag{10}$$

where $\alpha$ is a hyperparameter. Unlike traditional likelihood-based predictions, which are often susceptible to confounders, our causal prediction framework eliminates the influence of confounders, thus gain a unbiased AQI prediction.

## 5 Experiments

In this section, we evaluate the effectiveness of CAP through experiments conducted on real-world datasets.

### 5.1 Experimental Settings

**Datasets:** To our best knowledge, there is, so far, no publicly available dataset in the field of AQI intervention assessment on industrial emissions. We collected two real-world datasets, $PM_{2.5}$-CQW and $SO_2$-CQE, which provide rich industrial emission data, meteorological data and AQI data collected from air quality monitoring stations. The $PM_{2.5}$-CQW dataset covers the $PM_{2.5}$ emission data of more than 300 enterprises in the western region of Chongqing Municipality for one year (from Jan. 1, 2022 to Dec. 31, 2022), while the $SO_2$-CQE dataset includes the $SO_2$ emission data of more than 200 enterprises in the eastern region. The data are collected at an hourly frequency. In our approach, we select the last 70% of the data (from Nov. 1, 2022 to Jun. 30, 2023) as the training set, the first 15% of the data (from Jul. 1, 2022 to Aug. 31, 2022) as the test set, and the remaining 15% of the data (from Sep. 1, 2022 to Oct. 31, 2022) as the validation set. The reason for dividing the dataset in this way is that both $SO_2$ and $PM_{2.5}$ are pollutants with low emissions in summer and high emissions in fall and winter, so the government's control of industrial emissions of these two pollutants will be high in winter and low in summer. That is, the effect of the unmeasured

confounder $C$ on industrial emissions $X$ is larger in the training set and smaller in the test set, corresponding to our causal plots for pre- and post-intervention in Fig. 1(a) and Fig. 1(b). In this way, the prediction effect of our method on the test set will demonstrate that eliminating unmeasured confounding effects is effective for the AQI prediction problem and enables robust AQI prediction.

**Baseline:** To evaluate the effectiveness of our proposed approach, we conduct a comprehensive comparison of CAP with various AQI prediction methods. These methods are categorized into three groups, namely deep Learning Models, Hybrid Models and Causal Inference Models respectively.

**Evaluation Metrics:** We evaluate the performance of methods with Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (SMAPE), and Standard Deviation (SD).

## 5.2 Comparison with Baselines

To ensure fairness, we deploy the same environment, loss function, and dataset for all models. We compare CAP with baseline models for AQI prediction and run each method five times, recording the average metrics. The final results are presented in Table 1.

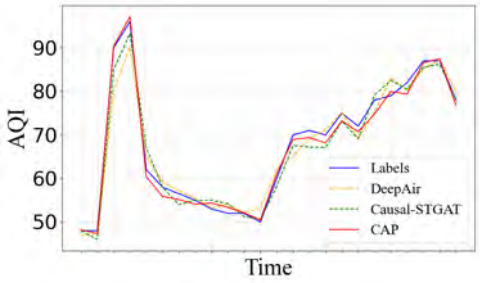



**Figure 3: The performance comparison for 24h prediction.**

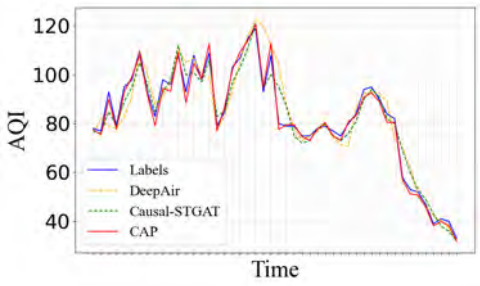



**Figure 4: The performance comparison for 48h prediction.**

The results show that our proposed CAP framework outperforms the baseline model for all metrics on both 24h prediction and 48h AQI prediction tasks. We categorize these methods according to deep learning models, hybrid models, and causal models. By combining multiple deep learning models, the hybrid model has more feature extraction capability, so the hybrid model generally outperforms the deep learning model. Among them, DAQFF [6] leverages multivariate air quality-related time series data, employing a hybrid deep learning model to extract spatio-temporal correlations. DeepAir [36] takes into account the interaction of various factors with the AQI respectively, modeling these to capture their complex effects, and it excels in both long-term and short-term predicting over existing shallow models. However, these methods do not explicitly define the causal relationships between factors, and the input of multiple variables introduces a multitude of confounders, leading to biased predictions.

Causal-STGAT [3] is a causal inference-based model, which is initially evaluated for human trajectory prediction but can be readily applied to air quality prediction, demonstrating good generalization. However, due to the unmeasured confounders in our data, it is necessary to estimate them first, which affects the model's predictive performance and increases computational costs. It could be said that it is not designed to eliminate the effects of unmeasured confounders, only addressing the impact of observed confounders. In fact, most confounders are unobservable and one of the main contributions of our model is the elimination of the effects of unobserved confounders, which is crucial for accurate AQI prediction.

In addition, for the two methods that are most competitive (DeepAir, Causal-STGAT) with our method, we report on comparison with SD in the appendix. And in Fig. 3 and Fig. 4, we visualize the performance of CAP with the next best two models. It can be seen that our method is predicted to track the trend best. That is, our method can more accurately predict the potential AQI after eliminating unobserved confounding, and the model is more resistant to interference, achieving more robust predictions.

## 5.3 Ablation Study

In order to elucidate the performance improvement of our method, we further investigated three variants of CAP named CAP-F, CAP-G, and CAP-C. These three variants disable some of the components of the model. Each of these methods is subjected to ablation experiments on the $PM_{2.5}$-CQW dataset, as described below:

**CAP-F.** It disables the modeling of $f(X, Z)$, formulated as:

$$\hat{Y}_{CAP-F} = h(X, X_{nei}, W)g(X, Z, M, W). \tag{11}$$

**CAP-G.** It disabled the modeling of $h(X, X_{nei}, W)$, which is expressed as:

$$\hat{Y}_{CAP-G} = f(X, Z)g(X, Z, M, W). \tag{12}$$

**CAP-C.** It simultaneously disables the modeling of $f(X, Z)$ and $h(X, X_{nei}, W)$, formulated as:

$$\hat{Y}_{CAP-C} = g(X, Z, M, W). \tag{13}$$

Table 2 and Table 3 summarizes the predictive performance of CAP and its three variants. The results indicate that CAP outperforms in all scenarios. The superior performance of CAP over three variants suggests that using only parts of the models for AQI prediction may not yield satisfactory results. Taking CAP-G as an example, it overlooks the influence of $(X, X_{nei})$ on $M$ and focuses solely on the impact of $M$ on $Y$, *i.e.*, the incomplete explanation of causal effects limits the performance of the three variants. This underscores the importance of interpreting causal relationships between data based on causal inference.

In addition, is it effective to consider the intervention in neighbors? Does considering meteorological factors improve the performance of the model? Comparative studies are conducted on these two questions as well. As shown in Table 2, when the intervention

**Table 2: Comparison with variants of our method on $PM_{2.5}$-CQW for 24h prediction.**

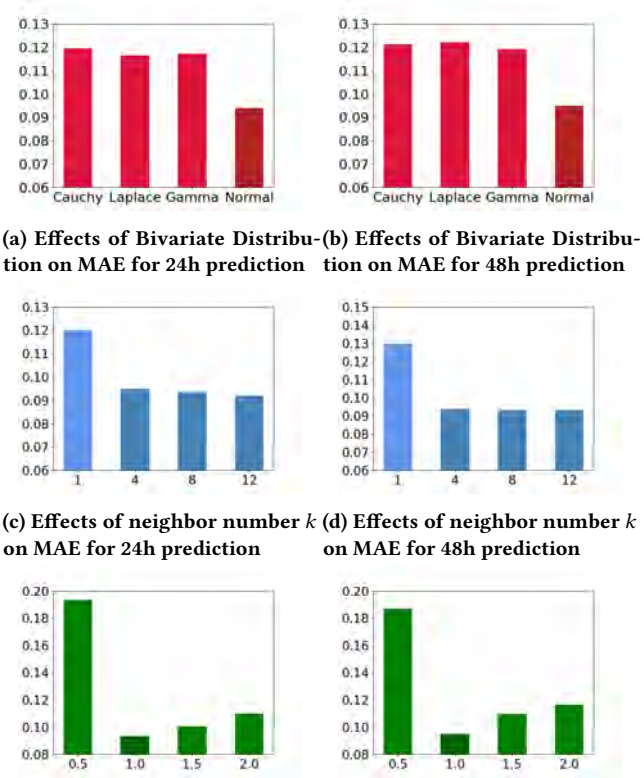
| Method | 1-24h | | | |
|---|---|---|---|---|
| | MSE | RMSE | MAE | SMAPE |
| CAP-F | 0.2645 | 0.5013 | 0.3802 | 78.49% |
| CAP-G | 0.0810 | 0.2692 | 0.1627 | 44.41% |
| CAP-C | 0.4600 | 0.6719 | 0.5071 | 91.15% |
| w/o $X_{nei}$ | 0.0212 | 0.1599 | 0.1298 | 28.64% |
| w/o $W$ | 0.0207 | 0.1399 | 0.0974 | 24.71% |
| CAP | **0.0204** | **0.1388** | **0.0934** | **24.43%** |

**Table 3: Comparison with variants of our method on $PM_{2.5}$-CQW for 48h prediction.**

| Method | 25-48h | | | |
|---|---|---|---|---|
| | MSE | RMSE | MAE | SMAPE |
| CAP-F | 0.3253 | 0.6055 | 0.4291 | 91.71% |
| CAP-G | 0.0599 | 0.2714 | 0.1701 | 46.31% |
| CAP-C | 0.5103 | 0.7099 | 0.5917 | 99.31% |
| w/o $X_{nei}$ | 0.0312 | 0.1662 | 0.1339 | 31.14% |
| w/o $W$ | 0.0313 | 0.1672 | 0.1391 | 32.12% |
| CAP | **0.0310** | **0.1392** | **0.0949** | **25.13%** |

in neighboring enterprises is disabled, the model performance is slightly degraded. The model performance is also affected when we ignore the meteorological factor $W$. Thus, the CAP framework is not only effective but also easy to interpret.

To further demonstrate the effectiveness of the CAP framework, we also do comparative studies on different experimental settings. When constructing the model, we assume that the data follows a bivariate normal distribution. We further explore alternative assumptions, such as the bivariate Cauchy, bivariate gamma, and bivariate Laplace distributions, and conduct comparative analyses, as shown in Fig. 5(a) and Fig. 5(b). The results indicate that, compared to other distributional assumptions, the CAP framework, which assumes a bivariate normal distribution, achieves superior performance, reflecting the validity of our model's assumptions. In Fig. 5(c) and Fig. 5(c), $k = \{1, 4, 8, 12\}$ represents the number of selected neighboring enterprises. When $k = \{4, 8, 12\}$, the predictive performance of the model is comparable, but larger $k$ values lead to increased computational costs and may not be conducive to accurate government control of industrial emissions. Therefore, we selected $k = 4$ as the default setting. Finally, we systematically evaluated the sensitivity of the model performance to the hyperparameter $\alpha$. As shown in Fig. 5(e) and Fig. 5(f), it is obvious that two loss functions are of equal importance. When $\alpha = 0.5$, the model relies more on RE to guide the training process, which may cause the model to fall into a local optimum. When $\alpha > 1$, the model may ignore the important information provided by RE during the optimization process. Therefore, an unbalanced training strategy may impair the model's learning of key features.




**(a) Effects of Bivariate Distribution on MAE for 24h prediction**

**(b) Effects of Bivariate Distribution on MAE for 48h prediction**

**(c) Effects of neighbor number $k$ on MAE for 24h prediction**

**(d) Effects of neighbor number $k$ on MAE for 48h prediction**

**(e) Effects of different $\alpha$ on MAE for 24h prediction**

**(f) Effects of different $\alpha$ on MAE for 48h prediction**

**Figure 5: Comparison with different experimental settings on $PM_{2.5}$-CQW.**

## 6 Conclusion and Future Work

In this paper, we construct a causal graph to describe the causal relationship between industrial emissions and AQI based on SCM. We then propose a CAP framework to predict potential AQI that eliminates unmeasured confounders through the intervention in industrial emissions from the target enterprise and its neighboring enterprises. Our framework is the first to successfully apply SCM to the problem of AQI prediction. In particular, benefiting from our skillful partitioning of the dataset, we are able to demonstrate the validity of robust AQI prediction that eliminates unobserved confounding effects. However, our framework for AQI prediction still relies on the mediator variable $M$ and only considers time slices in explaining causality. Therefore, we will expand the proposed CAP framework in the following directions: 1) addressing scenarios that do not conform to the front-door adjustment, 2) considering the effect of time-lagged variables in the causal graph.

## Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (grant Nos. 62477004, 62377040, 62207007, 623B2002) and Chongqing Natural Science Foundation Innovation and Development Joint Fund (grant. No. CSTB2023NSCQ-LZX0109).

## References


[1] Adel A Abdel-Rahman. 2008. On the atmospheric dispersion and Gaussian plume model. In *Proceedings of the 2nd International Conference on Waste Management, Water Pollution, Air Pollution, Indoor Climate, Corfu, Greece*, Vol. 26.

[2] Gary A Briggs. 1973. *Diffusion estimation for small emissions. Preliminary report.* Technical Report. National Oceanic and Atmospheric Administration, Oak Ridge, Tenn.(USA ....

[3] Guangyi Chen, Junlong Li, Jiwen Lu, and Jie Zhou. 2021. Human trajectory prediction via counterfactual analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9824–9833.

[4] Hu Chen, Guoqu Deng, and Yiwen Liu. 2022. Monitoring the influence of industrialization and urbanization on spatiotemporal variations of AQI and PM2. 5 in three provinces, China. *Atmosphere* 13, 9 (2022), 1377.

[5] Bowen Cui, Minyi Liu, Shanqiang Li, Zhifan Jin, Yu Zeng, and Xiaoying Lin. 2023. Deep learning methods for atmospheric PM2. 5 prediction: A comparative study of transformer and CNN-LSTM-attention. *Atmospheric Pollution Research* 14, 9 (2023), 101833.

[6] Shengdong Du, Tianrui Li, Yan Yang, and Shi-Jinn Horng. 2019. Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering* 33, 6 (2019), 2412–2424.

[7] Yuanfang Du, Shibing You, Weisheng Liu, Tsering-xiao Basang, and Miao Zhang. 2023. Spatiotemporal evolution characteristics and prediction analysis of urban air quality in China. *Scientific Reports* 13, 1 (2023), 8907.

[8] Cees Egmond, R Jonkers, and G Kok. 2006. One size fits all? Policy instruments should fit the segments of target groups. *Energy Policy* 34, 18 (2006), 3464–3474.

[9] Tânia Fontes, Peilin Li, Nelson Barros, and Pengjun Zhao. 2017. Trends of PM2. 5 concentrations in China: A long term approach. *Journal of environmental management* 196 (2017), 719–732.

[10] Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. 2021. Identification and estimation of treatment and interference effects in observational studies on networks. *J. Amer. Statist. Assoc.* 116, 534 (2021), 901–918.

[11] Wei Huang, Tianrui Li, Jia Liu, Peng Xie, Shengdong Du, and Fei Teng. 2021. An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability. *Information Fusion* 75 (2021), 28–40.

[12] Yuan Huang, Yuxing Xiang, Ruixiao Zhao, and Zhe Cheng. 2020. Air quality prediction using improved PSO-BP neural network. *Ieee Access* 8 (2020), 99346–99353.

[13] Lei Jiang and Ling Bai. 2018. Spatio-temporal characteristics of urban air pollutions and their causal relationships: Evidence from Beijing and its neighboring cities. *Scientific reports* 8, 1 (2018), 1279.

[14] Zhou Kang and Zhiyi Qu. 2017. Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou. In *2017 2nd IEEE international conference on computational intelligence and applications (ICCIA)*. IEEE, 155–160.

[15] Ujjwal Kumar and VK Jain. 2010. ARIMA forecasting of ambient air pollutants (O 3, NO, NO 2 and CO). *Stochastic Environmental Research and Risk Assessment* 24 (2010), 751–760.

[16] Baohong Li, Haoxuan Li, Anpeng Wu, Minqin Zhu, Qingyu Cao, Kun Kuang, et al. [n. d.]. A Generative Approach for Treatment Effect Estimation under Collider Bias: From an Out-of-Distribution Perspective. In *Forty-first International Conference on Machine Learning*.

[17] Baohong Li, Haoxuan Li, Ruoxuan Xiong, Anpeng Wu, Fei Wu, and Kun Kuang. 2024. Learning Shadow Variable Representation for Treatment Effect Estimation under Collider Bias. In *Forty-first International Conference on Machine Learning*. https://openreview.net/forum?id=ycXo4tQIpN

[18] Ying Li, Yung-ho Chiu, and Liang Chun Lu. 2018. Energy and AQI performance of 31 cities in China. *Energy Policy* 122 (2018), 194–202.

[19] Yuan-Chien Lin, Chun-Yeh Lai, and Chun-Ping Chu. 2021. Air pollution diffusion simulation and seasonal spatial risk analysis for industrial areas. *Environmental Research* 194 (2021), 110693.

[20] Jun Luo and Yaping Gong. 2023. Air pollutant prediction based on ARIMA-WOA-LSTM model. *Atmospheric Pollution Research* 14, 6 (2023), 101761.

[21] Sen Ni, Pengfei Jia, Yang Xu, Liwen Zeng, Xiaoyu Li, and Min Xu. 2023. Prediction of CO concentration in different conditions based on Gaussian-TCN. *Sensors and Actuators B: Chemical* 376 (2023), 133010.

[22] Judea Pearl. 2009. *Causality*. Cambridge university press.

[23] Judea Pearl and Dana Mackenzie. 2018. *The book of why: the new science of cause and effect*. Basic books.

[24] Soumik Ray, Achal Lama, Pradeep Mishra, Tufleuddin Biswas, Soumitra Sankar Das, and Bishal Gurung. 2023. An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. *Applied Soft Computing* 149 (2023), 110939.

[25] Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* 100, 469 (2005), 322–331.

[26] Kunwar P Singh, Shikha Gupta, Atulesh Kumar, and Sheo Prasad Shukla. 2012. Linear and nonlinear modeling approaches for urban air quality prediction. *Science of the Total Environment* 426 (2012), 244–255.

[27] Mauricio Tec, James G Scott, and Corwin M Zigler. 2023. Weather2vec: Representation learning for causal inference with non-local confounding in air pollution and climate studies. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 14504–14513.

[28] Fan Wang, Chaochao Chen, Weiming Liu, Tianhao Fan, Xinting Liao, Yanchao Tan, Lianyong Qi, and Xiaolin Zheng. 2024. CE-RCFR: Robust counterfactual regression for consensus-enabled treatment effect estimation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3013–3023.

[29] Hao Wang, Jiajun Fan, Zhichao Chen, Haoxuan Li, Weiming Liu, Tianqiao Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, and Ruiming Tang. 2024. Optimal transport for treatment effect estimation. In *Advances in Neural Information Processing Systems*.

[30] Haotian Wang, Kun Kuang, Haoang Chi, Longqi Yang, Mingyang Geng, Wanrong Huang, and Wenjing Yang. 2023. Treatment effect estimation with adjustment feature selection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2290–2301.

[31] Haotian Wang, Kun Kuang, Long Lan, Zige Wang, Wanrong Huang, Fei Wu, and Wenjing Yang. 2023. Out-of-distribution generalization with causal feature separation. *IEEE Transactions on Knowledge and Data Engineering* (2023).

[32] Haotian Wang, Wenjing Yang, Longqi Yang, Anpeng Wu, Liyang Xu, Jing Ren, Fei Wu, and Kun Kuang. 2022. Estimating individualized causal effect with confounded instruments. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1857–1867.

[33] Hao Wang, zhengnan Li, Haoxuan Li, Xu Chen, Mingming Gong, BinChen, and Zhichao Chen. 2025. Optimal Transport for Time Series Imputation. In *International Conference on Learning Representations*.

[34] Zhenghua Wang and Zhihui Tian. 2017. Prediction of air quality index based on improved neural network. In *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*. IEEE, 200–204.

[35] LJ Xu, JX Zhou, Y Guo, TM Wu, TT Chen, QJ Zhong, D Yuan, PY Chen, and CQ Ou. 2017. Spatiotemporal pattern of air quality index and its associated factors in 31 Chinese provincial capital cities. *Air Quality, Atmosphere & Health* 10 (2017), 601–609.

[36] Xiuwen Yi, Junbo Zhang, Zhaoyuan Wang, Tianrui Li, and Yu Zheng. 2018. Deep distributed fusion network for air quality prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 965–973.

[37] Yunjiang Yu, Chun Dai, Yigang Wei, Huiming Ren, and Jiawen Zhou. 2022. Air pollution prevention and control action plan substantially reduced PM2. 5 concentration in China. *Energy Economics* 113 (2022), 106206.

[38] Peijiang Zhao and Koji Zettsu. 2020. Mastgn: Multi-attention spatio-temporal graph networks for air pollution prediction. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 1442–1448.

[39] Zixi Zhao, Jinran Wu, Fengjing Cai, Shaotong Zhang, and You-Gan Wang. 2022. A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting. *Ecological Indicators* 144 (2022), 109416.

[40] Chuan Zhou, Yaxuan Li, Chunyuan Zheng, Haiteng Zhang, Min Zhang, Haoxuan Li, and Mingming Gong. 2025. A Two-Stage Pretraining-Finetuning Framework for Treatment Effect Estimation with Unmeasured Confounding. *arXiv preprint arXiv:2501.08888* (2025).

[41] Julie Yixuan Zhu, Chenxi Sun, and Victor OK Li. 2017. An extended spatio-temporal granger causality model for air quality estimation with heterogeneous urban big data. *IEEE Transactions on Big Data* 3, 3 (2017), 307–319.

[42] Hao Zou, Haotian Wang, Renzhe Xu, Bo Li, Jian Pei, Ye Jun Jian, and Peng Cui. 2023. Factual observation based heterogeneity learning for counterfactual prediction. In *Conference on Causal Learning and Reasoning*. PMLR, 350–370.


## A Supplementary Experiments

In this section, we report standard deviation results for model performance. We compare with the two most competitive baselines, as shown in Table 4 and Table 5. From the tables, our method achieves the best performance with less volatility and better stability in both the 24h and 48h AQI prediction tasks.

**Table 4: Performance comparison with SD for 24h prediction.**

| Dataset | Method | 1-24h | | | |
|---|---|---|---|---|---|
| | | MSE | RMSE | MAE | SMAPE |
| $PM_{2.5}$-CQW | DeepAir | 0.0376 (+/-0.0004) | 0.1928 (+/-0.0004) | 0.1457 (+/-0.0009) | 37.37% (+/-0.16%) |
| | Causal-STGAT | 0.0219 (+/-0.0001) | 0.1657 (+/-0.0003) | 0.1254 (+/-0.0009) | 33.74% (+/-0.15%) |
| | CAP | **0.0204 (+/-0.0002)** | **0.1388 (+/-0.0002)** | **0.0934 (+/-0.0003)** | **24.43% (+/-0.09%)** |
| $SO_2$-CQE | DeepAir | 0.0332 (+/-0.0004) | 0.1810 (+/-0.0004) | 0.1364 (+/-0.0008) | 35.70% (+/-0.14%) |
| | Causal-STGAT | 0.0194 (+/-0.0004) | 0.1383 (+/-0.0005) | 0.1050 (+/-0.0009) | 30.06% (+/-0.13%) |
| | CAP | **0.0164 (+/-0.0003)** | **0.1252 (+/-0.0002)** | **0.0855 (+/-0.0004)** | **22.58% (+/-0.10%)** |

**Table 5: Performance comparison with SD for 48h prediction.**

| Dataset | Method | 24-48h | | | |
|---|---|---|---|---|---|
| | | MSE | RMSE | MAE | SMAPE |
| $PM_{2.5}$-CQW | DeepAir | 0.0406 (+/-0.0006) | 0.1982 (+/-0.0004) | 0.1485 (+/-0.0008) | 38.06% (+/-0.20%) |
| | Causal-STGAT | 0.0318 (+/-0.0003) | 0.1768 (+/-0.0003) | 0.1319 (+/-0.0007) | 34.98% (+/-0.11%) |
| | CAP | **0.0310 (+/-0.0002)** | **0.1392 (+/-0.0002)** | **0.0949 (+/-0.0004)** | **25.13% (+/-0.08%)** |
| $SO_2$-CQE | DeepAir | 0.0337 (+/-0.0005) | 0.1791 (+/-0.0003) | 0.1441 (+/-0.0008) | 35.82% (+/-0.20%) |
| | Causal-STGAT | 0.0471 (+/-0.0007) | 0.2177 (+/-0.0003) | 0.1761 (+/-0.0009) | 37.94% (+/-0.14%) |
| | CAP | **0.0223 (+/-0.0004)** | **0.1307 (+/-0.0002)** | **0.1022 (+/-0.0005)** | **23.77% (+/-0.11%)** |