Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/71/20241663

# Empowering air quality prediction through "AI + big data technology" in the perspective of environmental law—A quasi-natural experiment analysis based on the "national big data comprehensive experimental zone" pilot program

**Mengping Dai[1,2,*], Xinrui Zu[1], Lingling Gong[1], Ziyi Fan[1]**

[1]Hebei University of Economics and Business, Xuefu Road No. 47, Xinhua District, Shijiazhuang, Hebei Province, China

[2]2187374857@qq.com
*corresponding author

**Abstract.** The current environmental legal system primarily focuses on environmental pollution and ecological damage, yet it does not adequately address climate change as an environmental risk issue. Consequently, environmental law struggles to directly respond to air pollution and climate change. Due to the complexity of air pollution regulation, there is a need for legal frameworks that extend beyond conventional environmental pollution controls. The advent of "AI + Big Data" has provided new momentum and opportunities for predicting and improving air quality. This paper, viewed through the lens of environmental law and based on the policy of the National Big Data Comprehensive Experimental Zone, combined with the Difference-in-Differences (DID) model, outlines the impact mechanisms of Big Data and Artificial Intelligence in empowering air quality prediction and improvement. It also explores the path choices for achieving the objectives of accuracy and effectiveness in air quality prediction from the perspective of environmental law.

**Keywords:** Environmental Law, Air Quality, Big Data and Artificial Intelligence, Digital Empowerment, Multi-Period DID Model.

## 1. Introduction

### *1.1. Research Background on Climate Regulation under Environmental Law*

The main legislation that explicitly addresses the control of greenhouse gases within the current environmental law framework is Article 2 of the "Air Pollution Prevention and Control Law." The legal provisions on the coordinated control of greenhouse gases and air pollutants reflect China's current understanding of climate change and are the result of various interest negotiations. Therefore, instead of being entangled in the binary choice of whether carbon dioxide is a pollutant, it is more productive to seek possible synergistic paths based on clarifying the differences between the two. In fact, both traditional air pollution control and climate change regulation require a diversified regulatory approach that includes command-and-control measures and market mechanisms[1]. Denying carbon dioxide as a pollutant does not mean that pollution control experiences and methods cannot be applied to carbon dioxide. Viewing carbon dioxide as an air pollutant does not imply fully replicating traditional air pollution control methods in the field of greenhouse gases; it mainly signifies that in the absence of specialized climate change legislation, climate change regulation can selectively apply the norms or systems already established in the air pollution domain.

Among them, the Environmental Impact Assessment (EIA) system strives to achieve a balance between the scientific and democratic aspects of environmental decision-making by granting the ecological environment departments the authority to approve EIA reports, stipulating the EIA obligations of enterprises, and empowering public participation[2]. According to the "Environmental Impact Assessment Law," the relevant departments of the State Council, the people's governments at or above the city level with districts, and their respective departments should conduct EIAs on their comprehensive plans and special plans. For construction projects that impact the environment, regardless of the extent of the impact, the corresponding EIA procedures must be performed according to classification management standards. Planning EIAs mainly focus on the systemic, long-term, and cumulative impacts that the implementation of plans may have on the ecological environment. The industry scope and scale of operations involved are far larger than individual construction projects, and the resulting greenhouse gas emissions are also considerable. Therefore, incorporating greenhouse gas considerations into planning EIAs can not only enhance the government's emission reduction capacity from the source but also serve as a crucial means for the government to fulfill its obligations to address climate change.

### *1.2. Air Quality Improvement and Prediction*

This paper aims to deeply explore the impact of digital intelligence technology, particularly the digital empowerment effect based on the pilot policy of the "National Big Data Comprehensive Experimental Zone," on urban air quality, as well as the underlying mechanisms and spatial effects. It examines how digital intelligence technology can contribute to optimizing industrial structures (including greening residents' lifestyles and promoting the intensive transformation of industrial enterprises) and enhancing urban green innovation. Through environmental monitoring, integrating atmospheric environmental data from various sources such as weather forecasts, satellite remote sensing, and ground observation stations, and using the DID model analysis, this research aids governments and relevant departments in more effectively implementing emission reduction measures and accurately targeting pollution sources, thereby achieving the goal of scientifically managing and improving the quality of the ecological environment.

## 2. Literature Review

### *2.1. Review of Literature*

Existing studies mainly focus on the regulatory path of environmental law under climate change, exploring institutional paths for coordinated control within the current environmental law framework and utilizing legal experience from environmental pollution regulation to respond to climate legislation. Yumin Jiang and Shaoqing Guo, among others, emphasize the value of digital intelligence technology in building environmental data platforms and facilitating cross-departmental information sharing, arguing that this is conducive to the scientific and intelligent decision-making in government environmental governance. Hongli Wang and colleagues propose constructing a theoretical framework that includes digital intelligence technology and carbon emission performance, incorporating both into a unified analytical system.

### *2.2. Technological Innovation in Environmental Monitoring with Big Data*

The technological innovation in environmental monitoring brought by big data has significantly changed the monitoring of pollution sources. Traditional monitoring methods mainly rely on manual sampling and laboratory analysis, which are time-consuming, labor-intensive, and complex to operate. With the development of sensor technology, remote sensing technology, drone technology, and others, modern

DOI: 10.54254/2755-2721/71/20241663

environmental monitoring equipment has become increasingly intelligent, automated, and precise[3]. For instance, the establishment of sensor networks enables real-time monitoring of environmental parameters, allowing for the rapid and accurate acquisition of various pollutant concentrations and distributions. At the same time, remote sensing technology combined with Geographic Information Systems (GIS) can conduct remote monitoring over large areas, providing real-time insights into environmental changes[4]. The application of drone technology broadens the monitoring range, enabling more comprehensive coverage and high-altitude, low-altitude cruising and sampling in complex environments, greatly improving monitoring efficiency and accuracy, and providing strong support for environmental management and decision-making.

## 3. Empirical Analysis of Air Quality Improvement Effects Before and After Policy Implementation in Pilot Cities

### *3.1. Model Setup*

To explore whether big data has an impact on air quality improvement, this paper uses the "National Big Data Comprehensive Pilot Zone" initiative as an exogenous policy shock and employs a difference-in-differences (DID) model for testing. A multi-period DID model is constructed to assess the effects of the progressively implemented policy. The model compares the air quality improvement effects between the pilot cities of the "National Big Data Comprehensive Pilot Zone" policy and non-pilot cities, as well as the differences before and after policy implementation in the pilot cities. The specific model is as follows:

$$aqieit = \alpha 0 + \alpha 1 didit + \alpha xit + \mu i + Tt + \varepsilon it$$

aqieit:The dependent variable, representing the air quality improvement effect in city i during year t.;didit:The "National Big Data Comprehensive Pilot Zone" pilot policy;α1: The implementation effect of the "National Big Data Comprehensive Pilot Zone" pilot policy.;xit:Control variables.;μi:Individual fixed effects.;Tt:Time fixed effects.;εit:Random error term.

### *3.2. Explanation of Variables*

Dependent Variable: The air quality improvement effect primarily refers to the achievements in improving air quality. Based on the core requirements of the "General Plan for the Construction of Ecological and Environmental Big Data," this paper selects the reciprocal of PM10 concentration as the measurement indicator. The larger the value, the better the air quality improvement effect; conversely, the smaller the value, the worse the effect.

Core Explanatory Variable: The core explanatory variable in this paper is the DID term of the "National Big Data Comprehensive Pilot Zone" pilot (didit), used to measure the level of digitalization in the region. If a city is a pilot city of the "National Big Data Comprehensive Pilot Zone" and the observation time is after the pilot year, then didit=1; otherwise, policyit=0.

Control Variables: This paper introduces government investment (gi) and population size (ps) as control variables to account for the influence of other factors on the air quality improvement effect.

Other Variables: The paper also explores energy consumption transformation and the scale of industrial enterprises. Electricity consumption serves as a proxy variable for green transformation of energy consumption, while indicators for large and medium-sized industrial enterprises act as a proxy for the scale of industrial enterprises.

### *3.3. Data Sources*

The data used in this paper are sourced from the "China City Statistical Yearbook" and the "China Statistical Yearbook" published by the National Bureau of Statistics. Panel data from 108 prefecture-level cities from 2012 to 2022 were selected as samples to evaluate the role of the "National Big Data Comprehensive Pilot Zone" pilot in promoting air quality improvement. Additionally, considering the potential for heteroscedasticity in the data, the indicators were log-transformed.

**Table 1.** Descriptive Statistics Table

| Variable | Definition | Measurement Indicator | Observations | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|---|
| *aqie* | *Air Quality Improvement Effect* | *1/PM10* | 11<br>88 | 0.01 | 45.35 | 54.00 | 305.00 |
| *gi* | *Government Investment* | *Fiscal Expenditure on Energy Conservation and Environmental Protection* | 11<br>88 | 107.37 | 116.18 | 10.93 | 458.44 |
| *ps* | *Population Size* | *Population Density* | 11<br>88 | 1384.92 | 604.92 | 450.50 | 2195.00 |
| *ecgt* | *Green Transformation of Energy Consumption* | *Electricity Consumption* | 11<br>88 | 778.62 | 242.26 | 218.16 | 1281.00 |
| *e* | *Industrial Enterprise Scale* | *Large and Medium-sized Industrial Enterprises* | 11 | 477 | 29 | 79 | 101 |
| *s* | *Degree* | *Indicator* | 88 | .30 | 8.04 | .00 | 3.00 |

## 4. Empirical Analysis of Results

### *4.1. Basic Regression Analysis*

This study employs a multi-period DID model to examine the role of the "National Big Data Comprehensive Pilot Zone" policy in promoting improvements in air quality. Since control variables may influence the regression results of the core explanatory variables, a stepwise regression method is utilized to test the effect of big data on urban air quality improvement, with estimates made using clustered robust standard errors at the city level.

**Table 2.** Basic Regression Results

| VARIABLES | (1)<br>FE_xtreg1<br>ln_aqie | (2)<br>FE_xtreg2<br>ln_aqie | (3)<br>FE_xtreg3<br>ln_aqie |
|---|---|---|---|
| didit | 0.469*** | 0.450*** | 0.267** |
| ln_gi | (0.108) | (0.112)<br>0.153*<br>(0.0924) | (0.135)<br>0.187*<br>(0.100) |
| ln_ps | | | -0.326<br>(0.212) |
| City Fixed Effects | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes |
| Observations | 1188 | 1188 | 1188 |

Note: (1) The values in parentheses are t-statistics. (2) ***, **, * respectively indicate significance levels of 1%, 5%, and 10%. (3) Data source is Stata 17. The same applies to the following tables and will not be repeated.

Table 2, column (1) presents the results without adding control variables, indicating that at the 1% significance level, the estimated coefficient of didit is significant, that is, compared to non-pilot cities, the implementation of the "National Big Data Comprehensive Pilot Zone" pilot policy has enhanced the air quality improvement effect in pilot cities.

DOI: 10.54254/2755-2721/71/20241663

In Table 2, columns (2) and (3) show the results after adding two control variables sequentially, and the results remain significantly positive. Regarding control variables, government investment (lngi) has a positive impact on the improvement of air quality, and its estimated coefficient passes the 1% significance test; the impact of population size (lnps) on the improvement of air quality is negative, and its regression coefficient does not pass the significance test.

*4.2. Robustness Test*

**Parallel Trend Test.** A dynamic model is constructed for hypothesis testing, as follows:

$$aq\iota e_{it} \quad \beta_0 + \sum_{\delta=-1}^{\delta=-5} \beta_\delta pre_{it} + \beta cur_{it} + \sum_{\tau=}^{\tau=4} \beta_\tau post_{it} + \beta_1 X_{it} + \mu_i + \eta_t + \varepsilon_{it}$$

preit and postit are counterfactual dummy variables; preit: before the "National Big Data Comprehensive Pilot Zone" pilot year, t takes values of -4, -3, -2, -1; postit: after the "National Big Data Comprehensive Pilot Zone" pilot year, t takes values of 1, 2, 3, 4, 5, 6; currit: the year of policy implementation, t takes the value of 0. Other variables are set the same as in the basic regression analysis.

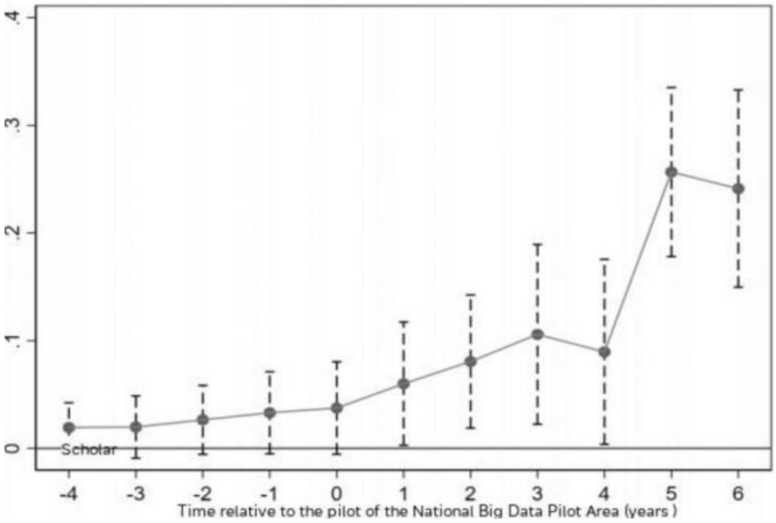


**Figure 2.** Parallel Trend Test Chart

There is no significant difference in the quality improvement effect; after the policy implementation, a significant difference in the air quality improvement effect between pilot and non-pilot cities is observed. Additionally, since the pilot policy was implemented at the end of the year, there is a certain degree of latency in the manifestation of its effects. In summary, the parallel trends assumption is met, making the use of the difference-in-differences method appropriate.

**Placebo Test.** Although a series of characteristic factors were controlled in the baseline regression model, there may still be unobservable factors that could influence the estimation results, potentially introducing bias. Accordingly, this paper conducts a placebo test, randomly generating "National Big Data Comprehensive Pilot Zone" pilot cities and repeating the regression simulation 500 times based on the benchmark model.

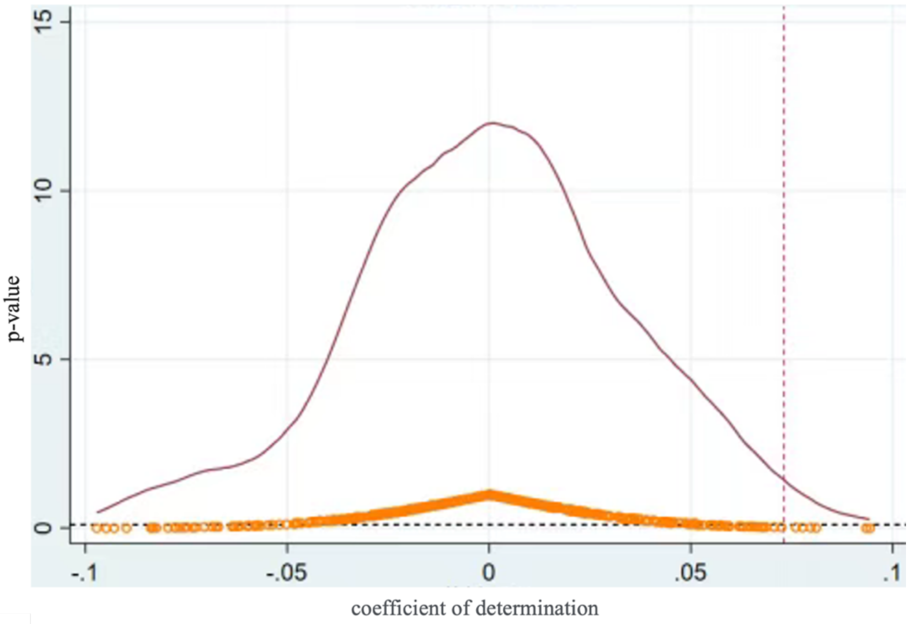


**Figure 3.** Placebo Test Chart

From Figure 3, it can be seen that the P-value distribution of the 500 simulated regressions and the kernel density distribution of the regression coefficients both basically conform to the normal distribution, indicating that the regression results have a certain degree of robustness.

**Test for the Impact of Non-Random Selection.** Due to the possible non-random selection of "National Big Data Comprehensive Pilot Zone" pilot cities, this paper introduces a series of interaction terms to optimize the basic regression model:

$$aqieit = \alpha 0 + \alpha 1 didit + \alpha xit + ci \times didit + \mu i + Tt + \varepsilon it$$

ci:Includes two city factors such as municipalities directly under the central government and desert cities. After introducing urban factors, the "National Big Data Comprehensive Pilot Zone" pilot policy has a positive impact on the improvement of urban air quality.

**Table 3.** Results of the Test for Non-Random Factors

| | Non-random factors |
|---|---|
| didit | 0.450***<br>(0.112) |
| Municipality Directly under the Central Government | 0.207*<br>(0.0887) |
| Desert City | 0.153*<br>(0.0924) |
| Urban Fixed Effect | Yes |
| Year Fixed Effect | Yes |
| Observations | 1188 |

PSM. This paper employs the Propensity Score Matching (PSM) method to mitigate the issue of selection bias. As shown in Table 4, the results indicate that at the 1% significance level, the estimated coefficient of didit is significantly positive, suggesting that after propensity score matching, the positive impact of the "National Big Data Comprehensive Pilot Zone" pilot policy on the improvement of air

DOI: 10.54254/2755-2721/71/20241663

quality remains significant. This also indicates that the conclusions of this paper possess a certain degree of robustness.

Other Policy Effect Tests. During the period from 2012 to 2022, pilot cities may have implemented other policies related to the "National Big Data Comprehensive Pilot Zone" pilot, which could potentially affect the evaluation results. Accordingly, this paper has controlled for the impact of representative policies such as green energy pilot, environmental tax pilot, and exhaust emission control pilot on the improvement of air quality.

**Table 4.** Impact of Other Policies

| VARIABLES | a1<br>ln_aqie | a2<br>ln_aqie | a3<br>ln_aqie |
|---|---|---|---|
| didit | 0.45*** | 0.38*** | 0.42*** |
| | (0.112) | (0.107) | (0.110) |
| Green Energy Pilot Program | 0.0175<br>(0.0424) | | |
| Environmental Tax Pilot Program | | 0.0357<br>(0.0568) | |
| Exhaust Emission Control Pilot Program | | | 0.0420<br>(0.0732) |
| Urban Fixed Effect | yes | yes | yes |
| Year Fixed Effect | yes | yes | yes |
| Observations | 1188 | 1188 | 1188 |

Table 4 indicates that after controlling for the impact of three policies, the pilot policy still exhibits robustness in promoting the improvement of urban air quality.

**Endogeneity Test.** To eliminate the endogeneity issue caused by reverse causality, a suitable instrumental variable is employed, and the two-stage least squares method (IV-2SLS) is utilized to address the bias arising from such endogeneity. The dependent variable lagged by one period is used for the endogeneity test. The instrumental variable passes the test, and simultaneously, at the 1% and 5% significance levels, the regression coefficient of didit is significantly positive, and it is greater than the coefficient in the baseline regression. This suggests that while using the instrumental variable to mitigate the endogeneity issue, the effect of big data on improving air quality is even stronger.

## 5. Analysis of Impact Mechanism

The aforementioned results demonstrate that big data can effectively promote the improvement of urban air quality. This paper further explores the impact mechanism of big data on air quality improvement from two aspects: energy consumption structure and the scale structure of industrial enterprises. The model is established as follows:

$$kit = \theta 0 + \theta 1 policyit + \theta xit + \mu i + Tt + \varepsilon it$$

In the model, kit represents the two structures mentioned. The energy consumption structure is tested from the perspective of green transformation of energy consumption, while the scale structure of the industrial sector is tested from the degree of industrial enterprise scale. The logarithm of electricity consumption (lnecgt) is used as a proxy variable for the green transformation of energy consumption, and the logarithm of the indicator of large and medium-sized industrial enterprises (lnes) is used as a proxy variable for the scale degree of industrial enterprises. The definitions of the remaining variables are the same as in the previous text.

DOI: 10.54254/2755-2721/71/20241663

**Table 6.** Results of Impact Mechanism Analysis

| | Green Transition of Energy Consumption *lnecgt* | Scale Degree of Industrial Enterprises *lnes* |
|---|---|---|
| *did* | 0.330*** (0.114) | 0.225** (0.0535) |
| *Control Variables* | yes | yes |
| *Urban Fixed Effect* | yes | yes |
| *Year Fixed Effect* | yes | yes |
| Observations | 1188 | 1188 |

**Energy Consumption Structure:**

As can be seen from Table 5, in terms of the green transition of energy consumption, the difference-in-differences term did is significant at the 1% significance level. This indicates that the "National Big Data Comprehensive Pilot Zone" pilot policy can promote the green transformation of energy consumption, thereby further enhancing the improvement of air quality through reduced energy usage.

**Industrial Enterprise Scale Structure:**

Table 5 also reveals that from the perspective of the scale degree of industrial enterprises, the "National Big Data Comprehensive Pilot Zone" pilot policy can foster the scaling of industrial enterprises, which in turn promotes the improvement of air quality.

In Summary, big data and artificial intelligence can drive the green transition of energy consumption and the scaling of industrial enterprises, thereby facilitating adjustments in the structure of industrial energy consumption and the scale of industrial enterprises, and consequently enhancing the improvement of air quality.

## 6. Conclusions and Recommendations

**Rooted in environmental legal norms, it is essential to return to the perspective of rights and responsibilities allocation, administrative procedures, and supervisory mechanisms.** Establishing a legal mindset for government environmental obligations and creating a public law intervention in environmental legal systems can provide a rich legal foundation for climate change regulation. At the same time, climate change regulation should also establish mechanisms connected with pollution control, identify and explore specific paths for responding to climate change with environmental legal systems. Of course, environmental law itself also needs continuous transformation, rethinking legislative purposes, expanding the basis of rationality, and enhancing the institutional capacity of legal procedures to respond to uncertainty and complexity [5].

**Promote the Green Transition of Energy Consumption and Achieve the Scaling of Industrial Enterprises**

Utilize big data technology to analyze massive environmental data, compensating for the deficiencies of traditional relational database technology and data analysis tools. Promote the green transition of energy consumption and the scaling of industrial enterprises, thereby improving air quality.

**Utilize the Internet of Things for Information Transmission to Achieve Intelligent Monitoring and Management**

In the Internet of Things, objects can be connected to the internet through information sensing devices, and based on the multi-directional transmission and exchange of object information, achieve intelligent identification and management. Surveys of the current state of environmental monitoring show that IoT technology has been widely applied in air quality monitoring, hydrological environment monitoring, and geological environment monitoring. Taking air quality monitoring as an example, traditional monitoring systems have fixed monitoring station equipment locations. Although this monitoring method can achieve long-term monitoring of air quality in a certain area, it has the disadvantage of a

DOI: 10.54254/2755-2721/71/20241663

limited monitoring range. With the use of IoT technology, not only is the connection of various sensors achieved, expanding the scope of information sharing, but the monitoring entities are also broadened.

**Leverage the Advantages of Artificial Intelligence to Stimulate Environmental Optimization Effects**

Fully leverage the technological innovation promotion, increased investment in emission reduction equipment, and the substitution mechanism for low-skilled labor brought about by artificial intelligence technology, continuously stimulating the environmental optimization effects of artificial intelligence. From the perspective of technological innovation promotion, artificial intelligence should be used to improve and enhance R&D foundations and innovation capabilities, continuously achieving innovative breakthroughs and promoting innovation as the core driving force for pollution reduction. From the perspective of increased investment in emission reduction, the intelligent industrial equipment should be fully utilized to create an industrial internet system, continuously enhancing management efficiency and operational capabilities through big data collection and intelligent algorithm processing, thereby optimizing resource allocation, reducing unnecessary production inputs, and supporting increased investment in emission reduction equipment. From the perspective of substituting low-skilled labor, enterprises should be supported in optimizing the existing labor factor structure with intelligent equipment, improving overall production efficiency and green production capabilities, and promoting the emission reduction effect.

## References

[1] He Xiangbai, The Environmental Law Approach to Climate Change Regulation: Legal Demands and System Selection, [A], Journal of Zhejiang Gongshang University, 2024-07-23

[2] Zhang Chenhe, Wang Dongdong, Li Xiaolan, et al. Verification Analysis of Air Quality Forecast in Liaoning Area Based on Numerical Models and Objective Methods [J]. Environmental Protection Science, 2024, 50(1): 110-119.

[3] Ye Lei, Geng Jinghua, Fang Wen, et al. Research on Pollutant Concentration Prediction at Discharge Outlets Based on Deep Learning [J/OL]. Environmental Science, 1-11 [2024-02-27].

[4] Zhang Wanting, Wan Bin, Wang Pu, et al. Discussion on the Comprehensive Efficiency Improvement of the Grid Environmental Air Quality Monitoring System with "Big Data+" [J]. Journal of Anhui Agricultural Science, 2024, 30(03): 87-90.

[5] Gao Xin, Zhu Shuangqing. The Path of Using Automatic Monitoring Data of Pollution Sources from the Perspective of Administrative Law Enforcement [J/OL]. Journal of Hebei Environmental Engineering College: 1-6 [2024-02-27].