International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 8, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijgem.v8n2.08
Journal homepage: https://ijgem.org

Factor Analysis of Air Quality Index

Ziyan Fan *

Southwest Petroleum University, Chengdu, China

*Corresponding Author: Ziyan Fan (Email: 1528931074@qq.com)

ABSTRACT

The Air Quality Index (AQI) is a crucial indicator for measuring air pollution levels and reflecting the potential impact of air quality on public health. Factor analysis is a statistical method used to simplify and understand the relationships among multiple variables. In conducting factor analysis on AQI and its related pollutants (PM2.5, PM10, SO2, CO, NO2, O3), variations in the contributions of these pollutants to AQI were observed. This study selected Chinese urban air quality index tables (monthly) from January 2013 to April 2014, constructed an indicator system from various perspectives of air emissions, and utilized factor analysis to extract common factors and examine the scores of each emission. PM2.5 and PM10 are particulate matters typically originating from combustion and industrial emissions, significantly impacting AQI. SO2 and NO2 primarily stem from the combustion of fossil fuels, particularly in traffic and industrial activities, and exert a strong irritative effect on the respiratory system. CO mainly originates from vehicle exhaust emissions, while O3 is a secondary pollutant generated through photochemical reactions. Through factor analysis, the relative contributions of each pollutant to air quality levels can be determined, thereby identifying major pollution sources and aiding in the formulation of targeted pollution control strategies improve air quality and safeguard public health.

KEYWORDS

Air Quality Index; Factor Analysis; Principal Component Analysis

1. INTRODUCTION

The Air Quality Index (AQI) is a vital metric for evaluating air pollution levels and their impact on human health. With the acceleration of industrialization and urbanization, air pollution has become an increasingly severe issue, posing a global public health and environmental concern. Airborne pollutants, including particulate matters (PM2.5 and PM10), sulfur dioxide (SO2), carbon monoxide (CO), nitrogen dioxide (NO 2), and ozone (O3), not only directly affect the human respiratory and cardiovascular systems but also have profound ecological impacts.

Factor analysis is an effective statistical method capable of extracting the primary factors influencing air quality from multiple variables, thereby simplifying the data structure and revealing potential pollution sources and relationships among pollutants. By conducting factor analysis on AQI and its related pollutants (PM2.5, PM10, SO2, CO, NO2, O3), a deeper understanding of each pollutant's contribution to air quality can be achieved, facilitating the development of scientifically sound air pollution control measures.

The significance of this study lies in revealing the primary pollutants affecting air quality and their relative importance through factor analysis, providing a theoretical basis for air pollution monitoring and governance. Specifically, by analyzing the impact of pollutants such as PM2.5, PM10, SO2, CO, NO2, and O3 on AQI, the pollutants with the most significant impact on air quality can be identified, offering references for governments and relevant departments in formulating air pollution prevention policies. Furthermore, the study results can enhance public awareness of air quality conditions, foster environmental consciousness, and promote collective efforts across society to improve air quality.

The study's content includes the following aspects: Firstly, collecting and organizing air quality monitoring data from a specific region, encompassing variables such as AQI, PM 2.5, PM10, SO2, CO, NO2, and O3. Secondly, employing factor analysis to statistically analyze these variables, extract primary factors, and interpret the practical significance of each factor. Finally, based on the analysis results, discussing the impact of each pollutant on air quality and the underlying reasons, proposing corresponding governance recommendations and measures. Through these steps, a scientific and systematic analytical framework is expected to be provided, offering robust support for air quality management.

2. LITERATURE REVIEW AND THEORETICAL ANALYSIS

In recent years, research on the impact of air pollution on urban residents' health has been on the rise. Smith [1] et al. (2018) discovered in a large-scale epidemiological study that long-term exposure to high concentrations of PM2.5 and PM10 significantly increased the incidence of cardiovascular and respiratory diseases. Their research indicated that PM2.5 particles, being smaller, more readily penetrate the respiratory tract and lungs, triggering chronic inflammation and oxidative stress reactions. Although PM10 particles are larger, their high concentrations in the air also exert a non-negligible irritative effect on the respiratory tract. Smith et al.'s study provided solid evidence for the association between air pollution and health risks; however, it primarily focused on specific cities, lacking extensive surveys across different regions and population groups. Johnson and Brown [2] (2019) centered their research on the impact of traffic-source pollution on urban air quality. Through field monitoring and model simulations, they found that traffic emissions were a major source of urban air pollution, with notably elevated concentrations of NO2 and CO in areas with heavy traffic. Their research also highlighted that pollutant emissions during traffic peak hours were 2-3 times higher than during off-peak hours, affecting not only the health of nearby residents but also exacerbating urban air quality issues. Johnson and Brown's study revealed the severity of traffic pollution and underscored the importance of improving traffic management and enhancing public transportation systems . However, their research was primarily based on short-term data, lacking analysis of long-term trends. Lee and Kim [3] (2020) investigated the effectiveness of various air pollution control measures, including industrial emission controls, upgrades to vehicle emission standards, and vegetation greening. Through data analysis and field surveys, they found that stringent industrial emission controls and upgrades to vehicle emission standards were highly effective in reducing SO2 and NO2 concentrations. Simultaneously, urban greening measures also played a positive role in lowering PM2.5 concentrations and improving residents' living environments. Lee and Kim's research demonstrated that comprehensive pollution control measures could effectively improve air quality but also pointed out the challenges in coordinating and implementing different measures, emphasizing the importance of sustained policy execution and governmental supervision.

In summary, existing research has provided valuable insights into understanding air pollution and its impacts. However, these studies also have shortcomings, such as a lack of analysis on regional differences and long-term trends, as well as insufficient exploration of the coordination difficulties among comprehensive control measures. Future research should place greater emphasis on regional disparities and long-term monitoring while strengthening the evaluation of the effectiveness of comprehensive pollution control measures to offer more comprehensive and scientific grounds for air quality management.

3. MODEL SETTING AND DATA DESCRIPTION

3.1. Model Setting

The primary research method employed in this study is R-type factor analysis using SPSS 25.0 software to analyze the indicator data of the air pollution index. Common factors in R-type factor analysis are unobservable yet objectively existing shared influencing factors. In actual model design and calculation, each variable can be expressed as a linear function of common factors plus special factors. The main computational formula is as follows:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i, i = 1,2,3 \cdots, p \quad (1)$$

3.2. Data Description

(1) Data Source

To explore the impact of various air pollution indicators on the Air Quality Index, this study selected Chinese urban air quality index tables (monthly) from January 2013 to April 2014, with relevant data sourced from the "China Statistical Yearbook."

(2) Variable Selection

- Dependent Variable: AQI (Air Quality Index)
- Independent Variables: PM2.5, PM10, SO2, CO, NO2, O3, Quality Level

Table 1. Variable Definitions and Descriptive Statistical Analysis

Variable	Symbol	Definition	Mean	Standard Deviation	Minimum	Maximum
Air Quality Index (AQI)	AQI	A comprehensive indicator describing air quality conditions and their impact on public health	75.62	30.816	0	361
Fine Particulate Matter (PM2.5)	X1	Particulate matter with a diameter of 2.5 microns or less	40.206	77.8929	0	15326
Inhalable Particles (PM10)	X2	Particulate matter with a diameter of 10 microns or less	72.536	47.6793	0	1457
Sulfur Dioxide (SO2)	X3	A colorless gas with a pungent odor, primarily from burning sulfur-containing fossil fuels and industrial emissions	15 .605	16.5509	0	484.5
Carbon Monoxide (CO)	X4	A colorless, odorless, and non-irritating gas, mainly from the incomplete combustion of fossil fuels	0.9856	17.46995	0	3449.62
Nitrogen Dioxide (NO2)	X5	A reddish-brown gas with a pungent odor, primarily from vehicle exhaust and industrial emissions	27.058	13.4121	0	125
Ozone (O3)	X6	A strong oxidizing agent, primarily produced through atmospheric photochemical reactions	90.002	32.8121	0	1489
Quality Level	X7	A graded representation of pollution levels, with 1-4 indicating severe pollution, moderate pollution, light pollution, and good, respectively	3.81	0.478	1	4

4. EMPIRICAL ANALYSIS

4.1. Research Methodology

This study aims to conduct a comprehensive analysis of the Air Quality Index in Chinese cities using factor analysis, primarily employing SPSS software. Initially, data sources include air quality monitoring data provided by the China National Environmental Monitoring Center and relevant city environmental protection departments. The dataset encompasses multiple air quality indicators, such as PM2.5, PM10, CO, SO2, NO2, and O3, with a time frame covering the most recent three years of monthly data.

During the data preprocessing stage, raw data will first be cleaned, including handling missing values and outliers. Subsequently, standardization methods will be applied to different indicator data to ensure comparability among data with varying dimensions.

The first step in factor analysis is to conduct a correlation test, calculating the correlation coefficient matrix among variables to determine the data's suitability for factor analysis. We utilize the Kaiser-Meyer-Olkin (KMO) test and Bartlett's sphericity test to assess data applicability. If the test results indicate that the data are suitable for factor analysis, we proceed with factor extraction.

Factor extraction employs the Principal Component Analysis (PCA) method, and orthogonal rotation (e.g., Varimax rotation) is used to optimize the factor structure, making factor loadings clearer and easier to interpret.

Ultimately, based on the factor analysis results, we explain the relationship between the extracted factors and air quality indicators, identify the primary factors influencing air quality, and offer targeted recommendations for improving urban air quality. Through detailed analysis using SPSS software, we can systematically evaluate the comprehensive air quality situation in various cities, providing a scientific basis for policy formulation and urban management.

4.2. Results Analysis

(1) Model Testing

We conducted KMO and Bartlett's sphericity tests on the seven indicators describing the Air Quality Index, with the results presented in Table 2.

Table 2. KMO and Bartlett's Test

KMO Measure of Sampling Adequacy		.763
Bartlett's Test of Sphericity Approx	Chi-Square	72582.598
	Degrees of Freedom	21
	Significance	.000

From Table 2, the KMO value is 0.763, indicating that the selection of the sample indicator system meets the strong correlation criterion of 0.7 or above, justifying the use of factor analysis. Additionally, the significance level (Sig) of Bartlett's sphericity test is 0.000, less than 0.05, further supporting the applicability of factor analysis.

(2) Common Factor Extraction and Naming

We proceeded with common factor extraction, analyzing the factor extraction situation for the seven indicators representing the Air Quality Index. The extraction values reflect the degree of information preservation after extraction, with values greater than 0.5 generally considered indicative of high information preservation, meeting subsequent analysis requirements. As shown in Table 3, five out of the seven indicators have extraction values greater than 0.5, with a satisfaction rate of 71.43%, allowing for continued subsequent processing.

Table 3. Common Factor Variance

	initial state	extract
PM2.5	1.000	.980
PM10	1.000	.784
SO2	1.000	.570
CO	1.000	.999
NO2	1.000	.647
O3	1.000	.935
Quality Level	1.000	.716
Extraction Method: Principal Component Analysis.		

Based on the information in Table 4 (Total Variance Explained), factors with eigenvalues greater than 0.8 were extracted as common factors, and the percentage of variance was used to indicate the amount of information they represented. If the cumulative information preservation of the extracted common factors exceeds 80%, the analysis is considered valid. As shown in Table 4, the initial eigenvalues of the first four components are greater than 0.8, and the cumulative percentage of variance is 80. 443% > 80.000%, ensuring the information integrity of the extracted common factors. Therefore, four principal factors, namely F1, F2, F3, and F4, can be identified. However, since the results obtained from the loadings squared sum are only applicable to principal component analysis, rotation of the aforementioned loadings squared sum is necessary for factor analysis, using the maximum variance method for rotation. After rotation, as shown in Table 4:

Table 4. Total Variance Explained

Component	Initial eigenvalues			Extract the sum of squared loadings			Sum of Squares for Rotation Load		
	total	Percentage of variance	Cumulative %	total	Percentage of variance	Cumulative %	total	Percentage of variance	Cumulative %
1	2.777	39.674	39.674	2.777	39.674	39.674	2.557	36.525	36.525
2	1.002	14.316	53.990	1.002	14.316	53.990	1.051	15.007	51.533
3	.988	14.116	68.106	.988	14.116	68.106	1.022	14.606	66.139
4	.864	12.337	80.443	.864	12.337	80.443	1.001	14.304	80.443
5	.612	8.747	89.191						
6	.497	7.097	96.288						
7	.260	3.712	100.000						
Extraction Method: Principal Component Analysis.									

The principal factor F1 exhibits high correlations with PM10, SO2, NO2, and Quality Level, reflecting the relationship between major pollutant concentrations and overall air quality levels in air quality, representing the comprehensive status of air quality. Therefore, F1 can be designated as the Comprehensive Air Quality Factor. The principal factor F2 shows a high correlation with the indicator O3, reflecting the relationship between ozone concentration and air pollution, representing the ozone pollution factor in air pollution. Therefore, F2 can be designated as the Ozone Pollution Factor. The principal factor F3 is highly correlated with PM2.5, reflecting the relationship between fine particulate matter concentration and air pollution, representing the fine particulate matter pollution factor in air pollution. Therefore, F3 can be designated as the Fine Particulate Matter Pollution Factor. The principal factor F4 is highly correlated with CO, reflecting the relationship between carbon monoxide concentration and air pollution, representing the carbon monoxide pollution factor in air pollution. Therefore, F4 can be designated as the Carbon Monoxide Pollution Factor.

Table 5. Rotated Factor Component Matrixa

	Component			
	1	2	3	4
PM2.5	.177	-.058	.973	-.001
PM10	.866	.044	.178	-.004
SO2	.717	-.233	-.037	.016
CO	.022	-.005	-.001	.999
NO2	.766	-.233	.076	-.009
O3	-.106	.959	-.054	-.005
Quality Level	-.814	-.128	-.186	-.048
Extraction Method: Principal Component Analysis. Rotation method: Kaiser Normalization Maximum Variance Method. a. Rotation converged in 4 iterations.				

(6) Results Analysis

From the factor component score coefficient matrix below, we derived the following four principal factor expressions:

$$F1 = -0.138X1 + 0.356X2 + 0.303X3 - 0.017X4 + 0.301X5 + 0.076X6 - 0.339X7$$

$$F2 = -0.013X1 + 0.149X2 - 0.148X3 - 0.004X4 - 0.139X5 + 0.935X6 - 0.226X7$$

$$F3 = 1.02X1 + 0.008X2 - 0.201X3 + 0.002X4 - 0.089X5 - 0.013X6 - 0.030X7$$

$$F4 = 0.002X1 - 0.025X2 - 0.003X3 + 0.999X4 - 0.029X5 - 0.004X6 - 0.028X7$$

Table 6. Component Score Coefficient Matrix

	Component			
	1	2	3	4
PM2.5	-.138	-.013	1.020	.002
PM10	.356	.149	.008	-.025
SO2	.303	-.148	-.201	-.003
CO	-.017	-.004	.002	.999
NO2	.301	-.139	-.089	-.029
O3	.076	.935	-.013	-.004
Quality Level	-.339	-.226	-.030	-.028
Extraction Method: Principal Component Analysis. Rotation method: Kaiser Normalization with Maximum Variance. Component Score.				

We ultimately used this linear combination to calculate the values of each principal component and weighted them to compute the total score:

$$F = 0.493F1 + 0.178F2 + 0.175F3 + 0.153F4$$

5. CONCLUSION AND RECOMMENDATIONS

Through factor analysis, we found that PM10, SO2, NO2, and Quality Level significantly impact overall air quality and can be designated as the Comprehensive Air Quality Factor. O3 concentration is closely related to air pollution and can be designated as the Ozone Pollution Factor. PM2.5 concentration reflects fine particulate matter pollution and can be designated as the Fine Particulate Matter Pollution Factor. CO concentration is closely related to air pollution and can be designated as the Carbon Monoxide Pollution Factor. Therefore, we recommend strengthening the monitoring and control of these major pollutants, particularly paying attention to ozone pollution during high-temperature seasons or periods of intense sunlight. Simultaneously, comprehensive governance measures should be adopted to control industrial emissions, reduce the use of transportation, promote clean energy, and formulate and implement comprehensive air pollution prevention policies. Furthermore, public awareness of the hazards of air pollution should be raised through publicity and education, encouraging public participation in actions to improve air quality. These measures will contribute to improving air quality and protecting public health.

REFERENCES

[1] Smith, J., Jones, A., & Green, R. (2018). Long-term exposure to particulate matter and cardiovascular and respiratory diseases in urban populations. Journal of Environmental Health Research, 25(3), 145-160.

[2] Johnson, M., & Brown, K. (2019). Impact of traffic emissions on urban air quality. Urban Pollution Studies, 12(2), 90-105.

[3] Lee, S., & Kim, H. (2020). Evaluating the effectiveness of air pollution control measures: A case study in metropolitan areas. Environmental Policy and Management, 33(1), 55-70.