



## Correlation Analysis of Main Pollutant Concentration-A Case Study of Zhengzhou

Kai-guang ZHANG<sup>1</sup>, Hong-ling MENG<sup>2</sup>, Ming-ting BA<sup>3</sup>, Yan-min SUN<sup>4</sup>

<sup>1</sup> Zhengzhou Normal University, Zhengzhou 450044, China, [zzgis@zznu.edu.cn](mailto:zzgis@zznu.edu.cn)

<sup>2</sup> Zhengzhou Normal University, Zhengzhou 450044, China, [hnmhl@126.com](mailto:hnmhl@126.com)

<sup>3</sup> Zhengzhou Normal University, Zhengzhou 450044, China, [bmt1234@126.com](mailto:bmt1234@126.com)

<sup>4</sup> Zhengzhou Normal University, Zhengzhou 450044, China, [zzgis@126.com](mailto:zzgis@126.com)

### Abstract

Air pollution is one of the main problems to be solved in the sustainable development of China's economy, its main pollution components include PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub>, the pollution component governance is an effective means of atmospheric environmental management. Based on the monitor data of six main pollutant concentrations in Zhengzhou from 2015 to 2018, this paper analyzes the correlation characteristics between their concentrations by using correlation analysis, the multiple correlation characteristics of the one pollutant concentration with the other five pollutant concentrations by using multiple correlation analysis, the independent linear interpretation ratio between the six pollutant concentrations by using partial correlation analysis, at last, a pollutant independent emission index is defined to describe the independent emission level of one pollutant, then utilize the index to study the distribution characteristics of six main pollution concentrations in the study period in Zhengzhou. The results show that there is a significant correlation between the six pollutant concentrations. PM<sub>2.5</sub>, O<sub>3</sub> and PM<sub>10</sub> are the primary pollutants in Zhengzhou, the PM<sub>2.5</sub> concentration is controlled by PM<sub>10</sub> concentration and CO concentration, similarly, the PM<sub>10</sub> concentration is controlled by PM<sub>2.5</sub> concentration. In the polluted weather, O<sub>3</sub> has the highest level of independent emissions. The main task of Zhengzhou in pollutant composition governance is to control the emission of PM<sub>2.5</sub> and O<sub>3</sub>.

**Keywords:** Air quality; Pollutant concentration; Correlation analysis; Multiple correlation analysis; Partial correlation analysis; Pollutant independent emission index.

### 1. Introduction

With the continuous development of China's economy, the advancement of urbanization and industrialization, the consumption of energy and natural resources have been continuing to increase. A large amount of harmful substances discharged into the atmosphere have been causing many serious ecological problems, especially atmospheric problem in many urbanized regions. Air pollution seriously has been affecting people's physical and mental health, and largely inhibiting the healthy development of the regional economy[1-5].

The statistical results show that atmospheric pollutants, mainly include Total Suspended Particulate Matter, Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Oxide (NO<sub>x</sub>), Volatile Organic Compounds (VOCs), Photochemical Oxide, Greenhouse[6,7]. In order to analyze and forecast the short-term air quality status and development trend in a region, the Ministry of Environmental Protection of China formulated "Environmental Air Quality Standards" (GB3095-2012) and "Technical Regulations for Environmental Air Quality Index" (hereafter this text will be abbreviated as Technical Regulations) (HJ633-2012) in 2012, Which, based on the impact degree of atmospheric pollutants as PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub> on human health, ecology and environment, equivalently divides their concentrations into 7 levels, and simplifying them into a single conceptual index value, that is, Air Quality Index(AQI) to classify the air pollution level and the short-time air quality status.

In recent years, based on the air quality index, many researchers have analyzed the variation characteristics of air quality in some typical regions at different scales, studied change patterns of air quality by natural environment element impacts, and the temporal and spatial distribution characteristics of air quality, have achieved many valuable results[1-9].

The Technical Regulations firstly calculates the Individual Air Quality Indices(IAQI) of the 6 main pollutants based on the equivalent divided pollutant concentration, then uses their maximum as the value of air quality index, takes the corresponding pollutant as the primary pollutant. Although the single maximum value could reflect the real-time air quality in a region to a certain extent, largely ignores the impact of other pollutants on air quality. For this reason, many scholars have tried to find a mathematical indicator that could comprehensively describe the air quality status and its development trend, by using neural network model[10,11], spatial econometric model[12,13], regression analysis model[14-15], principal component analysis model, the time series model, the analytic hierarchy model and the optimal hybrid model describe and predict[16-19], have obtained many valuable scientific research results.

This paper uses the correlation analysis method to define a pollutant independent emission index, to study the emission level of air pollutants in a region, in order to provide a scientific reference for controlling air pollution from the source. Firstly, introduce the AQI calculation model, secondly, based on the monitor data of 6 pollutant concentrations in Zhengzhou from 2015 to 2018, analyzes the correlation characteristics between the 6 pollutant concentrations, and then defines a pollutant independent emission index by eliminating the correlation between the pollutant concentrations, at last, utilizes this index to study the distribution characteristics of the 6 pollution concentrations in the study period in Zhengzhou.

## 2. The definition of air quality index

Air Quality Index (AQI) is the maximum of the six Individual Air Quality Indices (IAQI), and takes the corresponding pollutant as the primary pollutant.

The calculation steps are as follows:

(1) According to the equivalent impact degrees of the six air pollutants as PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub> on human health, ecology and environment, dividing the pollutant daily average concentrations of PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub>, CO, and 8-hour moving average concentration of O<sub>3</sub> (μg / m<sup>3</sup>) into 7 levels as Table 1.

(2) Calculates Individual Air Quality Index (IAQI) of pollutant  $P$  using

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}} (C_p - BP_{Lo}) + IAQI_{Lo} \quad (1)$$

where  $C_p$  is the concentration of pollutant  $P$ ,  $BP_{Hi}$  and  $BP_{Lo}$  respectively are the extremes high-value and low-value  $C_p$  belongs to,  $IAQI_{Hi}$  and  $IAQI_{Lo}$  respectively are the individual air quality indices of  $BP_{Hi}$  and  $BP_{Lo}$  (see Table 1). The  $IAQI_p$  of pollutant  $P$  describes the influence degree of pollutant  $P$  on the short-term air quality status and its change trends[3-8].

**Table 1.** Individual Air Quality Index and corresponding pollutant concentration extreme value

IAQI	Pollutant Concentration Interval					
	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
0	0	0	0	0	0	0
50	35	50	50	40	2	100
100	75	150	150	80	4	160
150	115	250	470	180	14	210
200	150	350	800	280	24	265
300	250	420	1600	565	36	800
400	350	500	2100	750	48	1000
500	500	600	2620	940	60	1200

(3) Calculates AQI using

$$AQI = \max \{ IAQI_{PM2.5}, IAQI_{PM10}, IAQI_{SO_2}, IAQI_{NO_2}, IAQI_{CO}, IAQI_{O_3} \} \quad (2)$$

and takes the corresponding pollutant as the primary pollutant.

### 3. The correlation analysis of pollutant concentration

#### 3.1 Correlation analysis

The correlation coefficient is a statistical indicator used to describe the relationship between two variables. Assume  $(x_i, y_i) \quad i = 1, 2, 3 \dots n$  are the observation values of random variable  $(X, Y)$ , the correlation coefficient of  $X$  and  $Y$  is defined as

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{D(X)D(Y)}}, \quad (3)$$

where  $\bar{x}$  and  $\bar{y}$ ,  $D(X)$  and  $D(Y)$  are respectively their means and variances.

$r_{XY}$  belongs to  $[-1, 1]$ , usually, the statistic  $t = r\sqrt{(n-2)/(1-r^2)} \sim t(n-2)$  ( $n$  is number of observations) is used as the test function of significant level. Under the significant level of 0.05, the greater  $|r|$  means the stronger correlation between  $X$  and  $Y$ ,  $|r| = 1$  means  $X$  and  $Y$  is linearly dependent, and  $r = 0$  means  $X$  and  $Y$  is mutual independent[20-24].

In order to study the correlation between the 6 pollutant concentrations, taking Zhengzhou as an example, use Formula (3) calculate the correlation coefficient between them in the study period. The results are shown in Table 2.

**Table 2.** The correlation coefficients of the pollutant concentrations (Zhengzhou, 2015-2018)

	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM2.5	1.000	0.883	0.517	0.637	0.839	-0.366
PM10	0.883	1.000	0.621	0.708	0.748	-0.249
SO <sub>2</sub>	0.517	0.621	1.000	0.599	0.554	-0.374
NO <sub>2</sub>	0.637	0.708	0.599	1.000	0.601	-0.339
CO	0.839	0.748	0.554	0.601	1.000	-0.423
O <sub>3</sub>	-0.366	-0.249	-0.374	-0.339	-0.423	1.000

Under the significant level of (both sides) 0.000, there are significant correlations between the 6 pollutant concentrations. The positive correlations appear between the 5 pollutant concentrations as PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub> and CO, their correlation coefficients are all greater than 0.500, indicating that the increase of one of the 5 pollutant concentrations in the region might lead to the increase of the other 4 pollutant concentrations, controlling the emission of one pollutant would reduce the air pollution by the remaining 4 pollutants. The negative correlations appear between O<sub>3</sub> and PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub>, CO, showing the increases of the concentration of PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub>, CO might cause the decrease of the O<sub>3</sub> concentration, vice versa.

#### 3.2 Multiple correlation analysis

There are significant correlations between the 6 pollutant concentrations, the change of one pollutant concentration would cause the changes of other 5 pollutant concentrations, in order to quantitatively analyze the degree of linear influence between them, the multiple correlation coefficient is introduced.

Suppose the concentration of pollutant  $i$  could be linearly interpreted by the concentrations of other 5 pollutants

$$x_i = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \dots + \alpha_6 x_6 + \varepsilon_i \quad (k \neq i). \quad (4)$$

Using the least squares calculate the value  $\hat{\alpha}_j$  of  $\alpha_j$ , the regression value of  $x_i$  could be represented as

$$\hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_k x_k + \dots + \hat{\alpha}_6 x_6 \quad (k \neq j). \quad (5)$$

The total variance of  $i$  pollutant concentration is

$$SST_i = \frac{1}{n} \sum_1^n (x_i - \bar{x}_i)^2 = \frac{1}{n} \sum_1^n (\hat{x}_i - \bar{x}_i)^2 + \frac{1}{n} \sum_1^n \varepsilon_i^2 = SSR_i + SSD_i, \quad (6)$$

where  $SSR_i$  is the linear interpretation part in the total variance of  $i$  pollutant concentration by the other 5 pollutant concentrations, and  $SSD_i$  is part of its own interpretation.

The multiple correlation coefficient of  $i$  is defined as the arithmetic square root of the ratio of the linear interpretation part to the total variance

$$Mr_i = \sqrt{SSR_i / SST_i} , \tag{7}$$

then

$$Mr_i^2 = SSR_i / SST \tag{8}$$

is the ratio of the linear interpretation part in the total variance.

$Mr_i$  belongs to  $[0,1]$ , usually, the statistic  $F_i = (n - 6)SSR_i / 5SSE_i \sim F(5, n - 6)$  ( $n$  is number of observations) is used as the test function of significant level. Under the significant level of 0.05, the larger  $Mr_i$  means the more closely the concentration of the pollutant  $i$  is linear correlation with the concentration of the other 5 pollutants.  $Mr_i = 1$  means the concentration of the pollutant  $i$  is linearly dependent with the concentration of the other 5 pollutants, and  $Mr_i = 0$  means they are linearly independent[20-24].

Taking Zhengzhou as an example, use Formula (4-8) calculate  $Mr_i$ ,  $Mr_i^2 \times 100$ ,  $SSR_i$ ,  $SSD_i$  and  $F_i$  in the study period. The results are shown in Table 3.

Under the significant level of 0.000,  $F \gg F_{0.005}(5,1210) = 12.140$  shows that one pollutant concentration is significant correlation with the other 5 pollutant concentrations. The linear interpretation degrees of PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub> and CO by the other 5 pollutants are all more than 50%, especially PM2.5 and PM10, the degrees reach up to 80%. The concentration change of O<sub>3</sub>, interpreted by the other 5 pollutants, is the minimum about 30%.

**Table 3.** The multiple correlation analysis of the pollutant concentrations (Zhengzhou,2015-2018)

	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
$Mr_i$	0.931	0.922	0.711	0.746	0.857	0.536
$Mr_i^2 \times 100$	86.587	84.855	50.557	55.703	73.462	28.755
$SSR_i$	3492711.497	5720348.768	195577.666	271718.056	301.465	1182117.838
$SSD_i$	541027.858	1016053.442	191268.567	216081.362	108.906	2928866.359
$F_i$	1562.278	1362.452	247.452	304.310	669.886	97.673

### 3.3 Partial correlation analysis

The multiple correlation coefficient reflects the dependent degree of the concentration of one pollutant with the concentrations of the other 5 pollutants as a whole. In order to quantify the degree of independent linear correlation between the concentrations of the two pollutants, analyze true relationship between two pollutants, it is necessary to eliminate the impact of the concentrations of the other 4 pollutants, In other words, under the condition of controlling the concentration of other 4 pollutants, to analyze the linear correlation of the concentrations between two pollutants, that is, partial correlation coefficient.

Assume  $x_i$  and  $x_j$  respectively are the concentrations of pollutants  $i$  and  $j$ , the concentration of pollutant  $i$  could be linear expression by the other 4 pollutant concentrations(except for  $j$ )

$$x_i = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \dots + \alpha_6 x_6 + \varepsilon_i (k \neq i, j) . \tag{9}$$

Due to the change of the independent variable, the linear interpretation part of Formula (9)  $Mr_{i(0 < k \leq 6; k \neq i, j)}^2$  is different with that of Formula (4), this difference

$$D_{Mr^2}(i, j) = Mr_{i(0 < k \leq 6; k \neq i)}^2 - Mr_{i(0 < k \leq 6; k \neq i, j)}^2 \tag{10}$$

is the proportion of the pollutant  $j$  independent linear interpretation to the concentration of the pollutant  $i$ .

The partial correlation coefficient of  $i$  and  $j$  is defined as

$$Pr_{ij}^2 = (Mr_{i(0 < k \leq 6; k \neq i)}^2 - Mr_{i(0 < k \leq 6; k \neq i, j)}^2) / (1 - Mr_{i(0 < k \leq 6; k \neq i, j)}^2) \tag{11}$$

$Pr_{ij}$  belongs to  $[-1,1]$ , usually, the statistic  $t = (Pr_{ij} \sqrt{(n-6)}) / \sqrt{1-Pr_{ij}^2} \sim t(n-6)$  ( $n$  is number of observations) is used as the test function of significant level. Under the significant level of 0.05, the larger  $|Pr_{ij}|$  means the larger impact degree of the pollutant  $j$  concentration on the pollutant  $i$  concentration [20-24].

Taking Zhengzhou as an example, use Formula (4), (9), (10) and (11) calculate  $Pr_{ij}$  and  $D_{Mr^2}(i, j)$  in the study period. The results are shown in Table 4 and Table 5 .

Under the significant level of 0.000, there is significant partial correlation between two pollutant concentrations in the 6 main pollutants. The independent linear interpretation ratio is proportional to the partial correlation coefficient between them, but the independent linear interpretation ratio of  $j$  to  $i$  is different with that of  $i$  to  $j$  .

The partial correlation coefficients between PM2.5and PM10, PM2.5and CO is greater than 0.500, but with the rest of 3 pollutants, the coefficients is negative value. For PM10, the positive correlation pollutants respectively are PM2.5, SO<sub>2</sub>, NO<sub>2</sub> and O<sub>3</sub> with correlation degree from strong to weak, its partial correlation coefficient with CO is negative value. For SO<sub>2</sub>, the positive correlation pollutants respectively are PM10, NO<sub>2</sub> and CO with correlation degree from strong to weak, its partial correlation coefficients with O3 and PM2.5 are negative value.

**Table 4.** The partial correlation analysis of the pollutant concentrations (Zhengzhou,2015-2018)

$Pr_{ii}^2$	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM2.5	1.000	0.706	-0.260	-0.030	0.544	-0.213
PM10		1.000	0.362	0.315	-0.053	0.315
SO <sub>2</sub>			1.000	0.217	0.189	-0.248
NO <sub>2</sub>				1.000	0.063	-0.145
CO					1.000	-0.152
O <sub>3</sub>						1.000

**Table 5.** The independent linear interpretation part between the pollutant concentrations (Zhengzhou,2015-2018)

$D_{Mr^2}(i, j)$	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM2.5	0.000	13.349	0.974	0.012	5.638	0.638
PM10	15.011	0.000	2.275	1.658	0.042	1.662
SO <sub>2</sub>	3.591	7.459	0.000	2.434	1.826	3.233
NO <sub>2</sub>	0.040	4.868	2.181	0.000	0.175	0.945
CO	11.155	0.075	0.980	0.105	0.000	0.625
O <sub>3</sub>	3.390	7.850	4.660	1.520	1.677	0.000

The pollutant  $i$  concentration  $C_i$  contains two parts, one part is  $C_{Ri}$  the linear interpretation part by the other 5 pollutant concentrations, the other is  $C_{Di}$  the part of its own interpretation

$$C_i = C_{Ri} + C_{Di} = 1 - Mr_i^2 + C_{Di} \tag{12}$$

The linear interpretation part by only  $j$  (called the independent linear interpretation ratio of  $i$  by  $j$ ) could be represented as

$$C_{R(i,j)} = Mr_i^2 \times \frac{D_{Mr^2}(i, j)}{\sum_{j \neq i} D_{Mr^2}(i, j)} \tag{13}$$

The linear interpretation part by all the 5 pollutants could be represented as

$$C_{Ri} = Mr_i^2 = \frac{Mr_i^2}{\sum_{j \neq i} D_{Mr^2}(i, j)} \sum_{j \neq i} D_{Mr^2}(i, j) \tag{14}$$

**Table 6.** The independent linear interpretation ratio between the pollutant concentrations (Zhengzhou,2015-2018)

	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM2.5	13.413	56.079	4.091 3	0.050	23.685	2.680
PM10	61.689	15.145	9.349	6.8137	0.173	6.830
SO <sub>2</sub>	9.791	20.337	49.443	6.6367	4.979	8.815
NO <sub>2</sub>	0.271	33.032	14.800	44.297	1.187	6.412
CO	63.328	0.426	5.564	0.5960	26.538	3.548
O <sub>3_8h</sub>	5.104	11.820	7.017	2.289	2.5259	71.245

Taking Zhengzhou as an example, use Formula (13) calculate the independent linear interpretation ratio in the study period. The results are shown in Table 6 .

The independent linear interpretation ratios of PM2.5 by PM10 and CO are greater than that of its own interpretation part. The independent linear interpretation ratios of PM10 and CO by PM2.5 are greater than that of its own interpretation part.. These pollutants are their upstream pollutants, their concentrations are controlled with the upstream pollutant concentrations, but there is a certain difference in the degree of control, the impact of PM2.5 on PM10 and PM2.5 on CO are significantly higher than that of the latter. The governance of their concentration must control the upstream pollutant concentration and its own pollutant concentration.

## 4 Pollutant independent emission index

### 4.1 Pollutant independent emission index

AQI describes the overall air quality status in a region, and classifies the air pollution level. In order to analyze the total pollution sources, study the emission levels of pollutants, and control the air pollution from the source, an index needs to be defined to describe the independent emission level of pollutant.

The basic idea of the definition of a Pollutant independent emission index is: For the concentration of a single pollutant is made of the linear interpretation part by the other 5 pollutant concentrations and the part of its own interpretation, it do not represent its emission level, its true emission level should be the part of its own interpretation. In the same reason, the Individual Air Quality Index (IAQI) also contains 2 part, as the linear interpretation part by the other 5 pollutant concentrations and the part of its own interpretation.

The Individual Pollutant Independent Emission Index (IPIEI)of pollutant  $i$  is defined as

$$IPIEI_i = IAQI_i \times (1 - Mr_i^2) . \tag{15}$$

The pollutant independent emission index (PIEI) is defined as

$$PIEI = \max \{ IPIEI_{PM2.5}, IPIEI_{PM10}, IPIEI_{SO_2}, IPIEI_{NO_2}, IPIEI_{CO}, IPIEI_{O_3} \} , \tag{16}$$

the corresponding pollutant called the primary independent emission pollutant.

### 4.2 Pollutant independent emission classification

In order to classify and forecast the independent emission levels, first calculate the maximum value of pollutant independent emission indices in the region from historical statistics, refer to the Grading Standard Proportion in the Technical Regulations, classify the pollutant emission degree into 6 levels as good, moderate, lightly emitted, moderately emitted, heavily emitted and severely emitted.

### 4.3 Pollutant independent emission analysis in Zhengzhou

Carrying out the correlation analysis, the partial correlation analysis and the multiple correlation analysis on the monitor data of the 6 pollutant concentrations in Zhengzhou from 2015 to 2018, calculating  $r_{ij}$ ,  $Mr_i$ ,  $Pr_{ij}^2$  and  $D_{Mr^2}(i, j)$ , the results are showed in Table 2-6. Based the Formula 15, the independent emission index of the 6 pollutants are

$$\begin{aligned} IPIEI_{2.5} &= 0.134 IAQI_{2.5} & IPIEI_{10} &= 0.151 IAQI_{10} \\ IPIEI_{SO_2} &= 0.494 IAQI_{SO_2} & IPIEI_{NO_2} &= 0.443 IAQI_{NO_2} \\ IPIEI_{CO} &= 0.265 IAQI_{CO} & IPIEI_{O_3} &= 0.712 IAQI_{O_3} \end{aligned} \tag{17}$$

Use Formula (17) calculate the individual independent emission indices and the pollutant independent emission indices. The maximum pollutant independent emission index is 386, then the minimum values of the 7 levels respectively are 0, 39, 77, 116, 154, and 232.

Due to the comprehensive consideration of the correlation between the 6 pollutant concentrations, there is a certain difference in describing the air status between AQI and PIEI. Table 7 is AQI and PIEI description comparison table, there are 44.00% of the days, the two indexes describe the same result, the remaining 56.00% of the air pollution levels determined by AQI is lower the Pollutant independent emission levels determined by PIEI.

Table 8 is the primary independent emission pollutants and primary pollutant transfer matrix, the primary pollutants are mainly PM2.5 (37.00%), O<sub>3</sub> (29.11%), PM10 (24.92%) and NO<sub>2</sub> (8.88%). CO is only one day, SO<sub>2</sub> has never been a primary pollutant. The primary independent emission pollutants are mainly O<sub>3</sub> (53.95%), CO (22.68%) and NO<sub>2</sub> (22.04%), PM10 is only 4 days, and PM2.5 has never been primary independent emission pollutant.

Among the days with PM2.5 as the primary pollutant, about 62.7% days CO is the primary independent emission pollutant, followed by 24.44% of O<sub>3</sub> and 12.44% of NO<sub>2</sub>. When the O<sub>3</sub> is the primary pollutant, it also is the primary independent emission pollutant. Among the days with PM10 as the primary pollutant, about 55.45% days O<sub>3</sub> is the primary independent emission pollutant, followed by 42.24% of NO<sub>2</sub>. Among the days with NO<sub>2</sub> as the primary pollutant, about 77.78% days the primary independent emission pollutant is itself, followed by 22.22% of NO<sub>2</sub>.

**Table 7.** The AQI and PIEI description transform matrix (Zhengzhou, 2015-2018)

	Good	Moderate	Lightly Polluted	Moderately Polluted	Heavily Polluted	Severely Polluted	Total
Good	7						7
Moderate	20	201					221
Lightly Emitted		314	154				468
Moderately Emitted		14	231	64			309
Heavily Emitted			21	73	82		176
Severely Emitted					8	27	35
Total	27	529	406	137	90	27	1216

**Table 8.** The primary independent emission pollutant and the primary pollutant transform matrix (Zhengzhou,2015-2018)

	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM2.5						
PM10	2	2				
SO <sub>2</sub>						
NO <sub>2</sub>	56	128		84		
CO	282	5			1	
O <sub>3</sub>	110	168		24		354

## 5 Conclusions and discussion

Based on the monitor data of 6 main pollutant concentrations in Zhengzhou from 2015 to 2018, this paper analyzes the correlation characteristics between them by using correlation analysis. The results show that: There are significant correlations between the 5 main pollutant concentrations, as PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub> and CO, which are significant negative correlation with the main pollutant concentration of O<sub>3</sub>. The linear interpretation ratio of the pollutant O<sub>3</sub> concentration by other 5 pollutant concentrations is lowest about 28.76%, this ratios of PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub> and CO by other 5 pollutant concentration are more than 50%. There is significant partial correlation between two pollutant concentrations, the independent linear interpretation ratio is asymmetric. The independent linear interpretation ratio of PM2.5 by PM10 and CO are greater than its own interpretation part, this feature also appears in CO by PM2.5 and PM10 by PM2.5.

The pollutant independent emission index describes the independent emission level of the pollutant, by eliminating the interaction of the pollutant concentrations. In Zhengzhou, the primary independent emission pollutants mainly are O<sub>3</sub>, CO, NO<sub>2</sub> and PM<sub>2.5</sub>. When PM<sub>2.5</sub> is the primary pollutant, CO and O<sub>3</sub> are the primary independent emission pollutants, O<sub>3</sub> and NO<sub>2</sub> are the primary independent emission pollutants in the days with PM<sub>10</sub> as the primary pollutant. As NO<sub>2</sub> is the primary pollutant, it is also the primary independent emission pollutant.

Air pollution is one of the main problems to be solved in the sustainable development of China's economy, the governance of pollution components is an effective means of atmospheric environmental management. PM<sub>2.5</sub>, O<sub>3</sub> and PM<sub>10</sub> are the primary pollutants in Zhengzhou, the concentration of PM<sub>2.5</sub> is controlled by PM<sub>10</sub> and CO, similarly, the concentration of PM<sub>10</sub> is controlled by PM<sub>2.5</sub>. In the polluted weather, O<sub>3</sub> has the highest level of independent emission. The main task of Zhengzhou in pollutant composition governance is to control the emission of PM<sub>2.5</sub> and O<sub>3</sub>.

## References

- [1] Lin Xueqin, WANG Dai. Spatio-temporal variations and socio-economic driving forces of air quality in Chinese cities. *Acta Geographica Sinica*, 2016, 71(8): 1357-1371.
- [2] CHEN Liding, ZHOU Weiqi, HAN Lijian, et al. Developing key technologies for establishing ecological security patterns at the Beijing-Tianjin-Hebei urban megaregion. *Acta Ecologica Sinica*, 2016, 36 (22) :7125-7129.
- [3] BIE Tong, HAN Lijian, TIAN Shufang, et al. Method for quantifying the contribution of urbanization on population exposure to air pollution. *Acta Ecologica Sinica*, 2018, 38 (13):4570-4583.
- [4] Ma Z, Hu X, Sayer A, et al.. 2016. Satellite-based spatiotemporal trends in PM<sub>2.5</sub> concentrations:China, 2004-2013. *Environmental Health Perspectives*, 2016, 124:184-192.
- [5] Liu H M, Fang C L, Zhang X L, et al. The effect of natural and anthropogenic factors on haze pollution in Chinese cities:A spatial econometrics approach. *Journal of Cleaner Production*, 2017, 165:323-333.
- [6] WANG Guanlan, XUE Jianjun, ZHANG Jianzhong. Analysis of Spatial-temporal Distribution Characteristics and Main Cause of Air Pollution in Beijing-Tianjin-Hebei Region in 2014. *Meteorological and Environmental Sciences*, 2016, 39( 1) : 34 -42.
- [7] DENG Xiajun, CAI Zhenqun, XIANG Xiaomei, et al. Air Pollutant Timing Characteristics and Its Correlation with Meteorological Conditions in Lishui. *Meteorological and Environmental Sciences*, 2015, 38 (2):60-65.
- [8] LU Lingyue, Li Hongyuan. Air Quality Evaluation of Beijing-Tianjin-Hebei Region of China Based on the Fuzzy Comprehensive Evaluation Method. *Acta Scientiarum Naturalium Universitatis Nankaiensis*, 2016, 49 (1):62-68.
- [9] HAO Kaiyue, CHEN Xiangyu, LI Yuanwei, et al. Correlative Analysis of Ultraviolet Radiation and Air Quality in Linzhi. *Environmental Science & Technology*, 2018, 41(07):103-106.
- [10] ZHAO Qilin, QIU Fei, YANG Jian. Application of NARX Neural Network Model in Environmental Air Quality Prediction in Kunming. *Environmental Monitoring in China*, 2019, 35(03):42-48.
- [11] QIU Chen, LUO Jing, ZHAO Chao-wen, CUI Kai-hui. Classification and Prediction of Air Quality Model Based on BP Neural Network. *Computer Engineering & Software*, 2019, 40(02):129-132.
- [12] ZHANG Xuexin, Zhou Yongcen. Prediction of Air Quality Wuhan and Analysis of the Influence on its Air Pollution Exerted by Those of Surrounding Cities. *Journal of Quantitative Economics*, 2019, 36(02): 42-48.
- [13] Jiang Lei, Zhou Haifeng, Bai Ling. Impact of Urbanization on Urban Air Quality Based on Spatial Econometric Models. *Tropical Geography*, 2019, 3:461-471.
- [14] LEI Lujin, LU Jun, HANG Ji. Analysis of Vehicle Restriction Policy on Air Pollution Control: A Case Study of Beijing. *Ecological Economy*, 2019, 35(07):203-208.
- [15] ZHANG Quan, TIAN Yong, YE Bojia. Research on Influence Factors of Airport Air Quality Based on Multiple Linear Regression. *Environmental Protection Science*, 2019, 45(1):35-43.
- [16] ZHANG Jie-qiong, WANG Ya-qian, GAO Shuang. Study on the relationship between meteorological elements and air pollution at different time scales based on KZ filtering. *China Environmental Science*, 2018, 38(10):3662-3672
- [17] Huang Shujun Li Kunquan Zhu Guangyi. Evaluation and analysis on the indoor air quality of colleges based on fuzzy analytic hierarchy process. *Environmental Engineering*, 2014, 5:90-94.
- [18] Yang Fan, Yang Shilin. Improvement and application of ahp in air quality assessment. *Environmental engineering*, 2016, 34(S1):887-889.
- [19] Zhang Hongri, Yin Xiaolin. Evaluation of Atmospheric Environmental Quality of Qingdao Based on Analytic Hierarchy Process. *Environmental Science and Management*, 2015, 40(07):180-184.



- [20] LIAO Zhiheng, SUN Jiaren, FAN Shao-jia, et al. Variation characteristics and influencing factors of air pollution in Pearl River Delta area from 2006 to 2012. *China Environmental Science*, 2015, 35(02): 329-336.a
- [21] Li L, Qian J, Ou C Q, et al. Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001-2011. *Environmental Pollution*, 2014, 190:75-81.
- [22] Chen Yanguang. *Mathematical Methods for Geography*. Beijing: Science Press, 2017.
- [23] Zhang Kaiguang, Ba Mingting, Meng Hongling. Spatial-temporal characteristics of Henan Highway traffics network accessibility and its evolution pattern. *Science of Surveying and Mapping*. 2017, 42(6):87-92.
- [24] ZHU Changlin, MENG Shuangshuang, ZHANG Rongguo. Correlation analysis and spatial and temporal distribution characteristics of main atmospheric pollutants in xi'an. *Environmental Engineering*, 2017, 35(12):86-91.