A new method for evaluating air quality using an ideal grey close function cluster correlation analysis method

Xiaoling Ren¹, Zhenfu Luo², Shuyu Qin³, Xinqian Shu¹ & Yuanyuan Zhang³

To scientifically and reasonably evaluate air quality with a large amount of monitored data, this paper proposes a new evaluation method called ideal grey close function cluster correlation analysis (IGCFCCA). Taking the air quality in Ningxia Province, China, as an example, according to China’s air quality standard, SO₂, NO₂, PM₁₀, PM₂.₅, and O₃ are selected as evaluation indexes to perform the evaluation. The results show that the air quality in this region in 2018 can be divided into three classifications, among which the relatively poor air quality in March, April, and May is the first classification, the better air quality in August and September is the third classification, and the air quality in other months falls under the second classification. Correlation analysis is used to qualitatively determine that these three classifications correspond to first-level air quality in China’s air quality standard, and the correlation degree, which is the distance between the three classifications and the first-level air quality, is quantitatively determined. Specifically, the correlation degrees of the first-classification, second-classification, and third-classification of air quality are 0.674, 0.697, and 0.71, respectively. The research results indicate potential directions and objectives for air quality management to achieve scientific management.

The air environment is a dynamic and complex system. The air quality is influenced by some pollutants, such as SO₂, NO₂, PM₁₀, and O₃. The concentrations of these pollutants are changing constantly. However, the monitored data used in analyses are usually collected in a certain period, and examples include one-hour average, few-hour average, one-month average and one-year average data. Instantaneous data collected every minute or second are difficult to collect and analyse. Therefore, this collection approach is considered a grey system. In a grey system, some information is known, and some information is unknown1–7.

At present, China’s air quality standard (GB3095-2012) divides air quality into two levels and stipulates the concentrations of pollutants in first-level and second-level air8,9. The concentrations of pollutants are comparatively lower in first-level air, and they are higher in second-level air. The major pollutants include SO₂, NO₂, PM₁₀, O₃, and others. However, when people evaluate air quality according to GB3095-2012, there may be some problems. First, according to the national standard, the common evaluation methods can only determine which level the current air is associated with. However, there is no analysis of how much the current air belongs to the level, and it is not clear how far the current air is from the standard level. The space for improving the current air quality is also very vague. It is necessary to develop a method to quantitatively calculate the correlation degree, which is the distance between the current air and the two levels of air standards. Second, to determine the air quality in a certain area in a period of time, the concentrations of pollutants are usually monitored every day. However, the amount of monitored data is very large. Obviously, if people compare and analyse each recorded value, the workload will be very large, and tasks will be almost impossible to complete. Therefore, people usually calculate the average value of the data first and then analyse the average. However, among so many monitored data, which data should be taken as a group for average calculation is a problem. In other words, determining how to scientifically classify data is the key. Data with similar characteristics can be classified into one group. These different classifications can be analysed and evaluated. Therefore, the results of the analysis can be scientific.

¹China University of Mining and Technology-Beijing, Beijing 100083, China. ²School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China. ³Yinchuan College, China University of Mining and Technology, Yinchuan 750411, Ningxia, China. ⁴email: 378814620@qq.com; 1334213027@qq.com
At present, there are many methods for comprehensively evaluating atmospheric environmental quality, including the air pollution index (API) method, ambient air quality index (AQI) method, single factor index method, green air pollution comprehensive index method, analytic hierarchy process, artificial neural network models, and fuzzy comprehensive evaluation method. Due to the different evaluation principles of various evaluation methods, each method has unique advantages and disadvantages. Among them, the API and AQI methods are simple, intuitive and convenient to use but only applicable for evaluating the short-term air quality in cities. The single factor index method is clear and easy to implement, but it is only applicable to areas where coal pollution is the main pollution type. The analytical hierarchy process (AHP) is simple, practical and systematic, but quantitative results are limited; additionally, when there are many indicators, the statistics will be complex, and weights will be difficult to determine. The artificial neural network evaluation method has the advantages of a fast operation speed, self-adaptation and strong fault tolerance, but the disadvantage is that when the data are poorly correlated, the evaluation results will exhibit homogenization phenomena. IGCFCCA is a kind of fuzzy comprehensive evaluation method based on fuzzy mathematics, the fuzzy principle and the grey close function. The method can solve the common incomplete data problem and mainly assesses the relationships between uncertainty and incomplete information analysis, model building and forecasting. The method only needs a small amount of data and can achieve good prediction results.

In this paper, the IGCFCCA method is used to evaluate the air quality in Ningxia Province. The method can not only scientifically classify a large amount of data but also calculate the correlation degree between each classification and the relevant standard. This approach can provide an important basis for comprehensive environmental management. Moreover, this new method provides a scientific reference and an important basis for the establishment and optimization of other industry standards in the future.

**Basic principle and methods**

A sample, which comes from the monitored data reports of some environmental management departments, is first classified by ideal grey close function cluster analysis. Then, the level of the sample is determined by grey correlation analysis, and comprehensive evaluation conclusions are established according to the correlation degree between the classification of the sample and the levels specified in GB3095-2012.

**The classification of the sample to be evaluated.** Establishing the evaluation index sequence matrix for the selected sample. Let \( S = \{s_1, s_2, \ldots, s_m\} \) be a sequence of air-influencing variables, i.e., \( X = \{x_1, x_2, \ldots, x_n\}\); \( x_{ik} \) is the original monitoring data for \( s_i \) \((i = 1, 2, \ldots, n)\) and \( x_{ik} \) \((k = 1, 2, \ldots, n)\); \( i \) and \( m \) represent the number of objects considered in clustering; \( k \) and \( n \) are the number of the influencing indexes which are the pollutants mentioned above. Accordingly, the following matrix can be established (Eq. 1).

\[
S = \begin{bmatrix}
  s_1 & x_{11} & x_{12} & \cdots & x_{1n} \\
  s_2 & x_{21} & x_{22} & \cdots & x_{2n} \\
  \vdots & \vdots & \vdots & \cdots & \vdots \\
  s_m & x_{m1} & x_{m2} & \cdots & x_{mn}
\end{bmatrix}
\]

(Eq. 1)

Establishing the matrix of ideal-value grey close function clusters. Let \( X_k = \{x_{01}, x_{02}, \ldots, x_{0n}\} \) be the ideal-value sequence corresponding to each influential index. The principle for determining the ideal value is as follows (Eqs. 2, 3, 4).

The first situation: The larger the influencing index (\( x_{ik} \)) is, the better the air quality is; in this case, the ideal value

\[
x_{0k} = \max \{x_{ik}; i = 1, 2, \ldots, m\}, k = 1, 2, \ldots, n.
\]

(Eq. 2)

The second situation: The smaller the influencing index (\( x_{ik} \)) is, the better the air quality is; in this case, the ideal value

\[
x_{0k} = \min \{x_{ik}; i = 1, 2, \ldots, m\}, k = 1, 2, \ldots, n.
\]

(Eq. 3)

Third, the air quality is best when the influencing index (\( x_{ik} \)) displays a moderate value, and the ideal value is

\[
x_{0k} = M.
\]

(Eq. 4)

According to the ideal value \( x_{0k} \) (Eqs. 2, 3 or Eq. 4) and the original monitored data (\( x_{ik} \)), the grey close function value \( y_{ik} \) is calculated by using (Eq. 5).

\[
y_{ik} = \frac{x_{0k}}{x_{ik}} (i = 1, 2, \ldots, m; k = 1, 2, \ldots, n)
\]

(Eq. 5)

where \( x_{ik} \) is the original monitored data and \( x_{0k} \) is the ideal value corresponding to the k-th influential index. Moreover, the function value \( y_{ik} \) is dimensionless, and \( y_{ik} \in [0,1] \). \( y_{ik} \) denotes the correlation degree of \( s_i \) and \( s_k \) for the k-th index. Specifically, the larger \( y_{ik} \) is, the closer \( s_i \) is to the ideal value \( s_{0k} \) and the smaller \( y_{ik} \) is, the farther \( s_i \) is from \( s_{0k} \).

Thus, the following grey close matrix \( Y \) can be established (Eq. 6).
\[
Y = \begin{bmatrix}
    y_{11} & y_{12} & \ldots & y_{1n} \\
    y_{21} & y_{22} & \ldots & y_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    y_{m1} & y_{m2} & \ldots & y_{mn}
\end{bmatrix}
\]  

(6)

In this case, Y is the grey close function value. Moreover, \( (y_{01}, y_{02}, \ldots, y_{0n}) = (1,1,\ldots,1)_{1\times n} \) is the ideal sequence, and the bigger \( y_{0k} \) is, the better \( s \) is; the biggest \( y_{0k} \) is equal to 1.

The classification of the sample to be evaluated. Because the influence of each influencing index is different, the weight of each influencing index needs to be considered. Let \( P_i \) be the comprehensive analysis value of \( s \). \( P_i \) can be expressed as follows (Eq. 7)

\[
P_i = \sum_{k=1}^{n} W_{ik} y_{ik} (i = 1, 2, \ldots, m)
\]

(7)

where \( W \) is the weight of each influencing index, and since the number of indexes is \( k \), the number of \( W \) values is also \( k \) (\( W_1, W_2, \ldots, W_k \)). Corresponding, the following equation can be established (Eq. 8).

\[
W_k = \sum_{j=1}^{m} \sum_{i=1}^{n} X_{ik} (k = 1, 2, \ldots, n)
\]

(8)

Based on the actual comprehensive analysis value \( P_i, P_j = (P_{i1}, P_{i2}, \ldots, P_{im})^T \). The following equation (Eq. 9) can be used to calculate the grey close value \( P_y \) of \( P_i \) in relation to \( P_j \).

\[
P_y = \frac{\min(p_i, p_j)}{\max(p_i, p_j)} (i, j = 1, 2, \ldots, m)
\]

(9)

Then,

\[
P = (P_y)_{m \times m}.
\]

(10)

If \( P \) (Eq. 10) satisfies the following three conditions: (1) reflexivity, where \( P_{ij} = 1 \) \((i = j)\); (2) symmetry, where \( P_{ij} = P_{ji} \); and (3) normativity, where \( P_{ij} \subseteq [0,1] \), we can select the appropriate threshold value from the \( P \) matrix, intercept the branches with weight values less than \( \lambda \), which is the similarity coefficient \(^4^,^5 \), and establish the classification \( S_t \) \((t = 1, 2, \ldots, c) \) when \( \lambda \) level meets the relevant requirement. \( S_t \) represents each classification of the air in a given region. The following equations (Eqs. 11, 12) can be established.

\[
S_t' = (S_{t1}', S_{t2}', \ldots, S_{tn}')^T
\]

(11)

\[
S_{ik} = (s_{i1}', s_{i2}', \ldots, s_{in}')
\]

(12)

where \( S_t \) is the \( t \)-th classification, \( S_{ik} \) is the \( k \)-th index of the \( t \)-th classification, \( t \) is the number of classifications, and \( k \) is the number of influencing indexes.

\( S_{ik} \) can be expressed in the following matrix form (Eq. 13).

\[
S_{ik} = \begin{bmatrix}
s_{i1} & s_{i2} & \ldots & s_{in} \\
s_{i2} & s_{i2} & \ldots & s_{in} \\
\vdots & \vdots & \ddots & \vdots \\
s_{i1} & s_{i2} & \ldots & s_{in}
\end{bmatrix}
\]

(13)

Correlation degree analysis of the sample to be evaluated. Let \( S_t \) be the sample to be evaluated, and let \( X = (x_1, x_2, \ldots, x_m) \), which is the influencing index set mentioned above and is the evaluation index used for \( S_t \). Let \( S_0 \) be the stated air quality classification in the GB3095-2012. Then, the equation for the correlation coefficient is as follows (Eq. 14)\(^4^,^5\):

\[
\zeta_t(k) = \frac{\min_{t \in \mathcal{E}, k \in \mathcal{N}} |S_t(k) - S_0(k)| + \epsilon \max_{t \in \mathcal{E}, k \in \mathcal{N}} |S_t(k) - S_0(k)|}{|S_t(k) - S_0(k)| + \epsilon \max_{t \in \mathcal{E}, k \in \mathcal{N}} |S_t(k) - S_0(k)|}
\]

(14)

where \( \zeta_t(k) \) is the correlation coefficient and \( \epsilon \) is the resolution coefficient, with a general value of 0.5\(^4^,^5\).

Moreover, the correlation degree (\( R_t \)) equation is as follows (Eq. 15).

\[
R_t = \frac{1}{n} \sum_{k=1}^{n} \zeta_t(k)
\]

(15)
The value of $R_t$ is calculated by using (Eq. 15). The maximum value of $R_t$ indicates that the sample to be evaluated has the highest correlation degree with the considered air quality level. Therefore, the sample is classified correspondingly.

### Air quality assessment—taking Ningxia Province in China as an example

The classification of the samples to be evaluated. Monthly reports of the air quality in Ningxia Province in 2018 were provided by the Department of Ecology and Environment of Ningxia Province. The monthly report data were used to establish the cluster of samples $S_i$ (Table 1) (Eq. 1). Each sample included five kinds of pollutants. Moreover, the concentrations of SO$_2$, NO$_2$, PM$_{10}$, and PM$_{2.5}$ were based on monthly averages calculated from 24-h averages, and the concentration of O$_3$ was the monthly average calculated from the 8-h average values.

- $x_1$ is the SO$_2$ concentration; $x_2$ is the NO$_2$ concentration; $x_3$ is the PM$_{10}$ concentration; $x_4$ is the PM$_{2.5}$ concentration; and $x_5$ is the O$_3$ concentration. For these pollutants, the lower the concentration is, the better the air quality is.

As shown in Table 1, because the management department only provided some monitored data and the data in January are incomplete, only the data that are listed in the table from February to December can be effectively analysed. However, the focus of this study is on the new analysis and evaluation method (IGCFCCA), and almost all of the data can be analysed by this method.

According to (Eq. 3), the five ideal values are as follows: $x_{01}$ is 9, $x_{02}$ is 17, $x_{03}$ is 56, $x_{04}$ is 25, and $x_{05}$ is 76. Based on the sample data in Table 1, the ideal-value grey close matrix (Eq. 6) can be obtained from (Eq. 5); according to (Eq. 8), the weights of $x_1$, $x_2$, $x_3$, $x_4$, and $x_5$ are $w_1 = 0.06$, $w_2 = 0.09$, $w_3 = 0.34$, $w_4 = 0.12$, and $w_5 = 0.39$, respectively. Consequently, the comprehensive analysis value $P_i$ ($i = 1, 2, ..., 11$) (Table 2) of $S_i$ is calculated with (Eq. 7). The grey close function value $y_{ik}$ (Eq. 5) and the comprehensive analysis value $P_i$ are shown in Table 2.

With $P_i$ ($P_1$, $P_2$, ..., and $P_{11}$) as known numbers, $P_j$ ($j = 1, 2, ..., 11$) can be calculated from (Eq. 9). The corresponding elements of the grey similar matrix (Eq. 10) are shown in Table 3.

### Table 1. Air quality in Ningxia Province in 2018.

| Index | SO$_2$ ($x_1$) | NO$_2$ ($x_2$) | PM$_{10}$ ($x_3$) | PM$_{2.5}$ ($x_4$) | O$_3$ ($x_5$) |
|-------|----------------|----------------|-------------------|--------------------|---------------|
| January | 93 | 46 | | | |
| February | 40 | 27 | 86 | 40 | 104 |
| March | 30 | 32 | 167 | 55 | 129 |
| April | 18 | 27 | 159 | 47 | 141 |
| May | 14 | 23 | 150 | 45 | 162 |
| June | 14 | 24 | 74 | 26 | 178 |
| July | 9 | 17 | 81 | 29 | 160 |
| August | 10 | 20 | 56 | 25 | 150 |
| September | 13 | 27 | 65 | 26 | 129 |
| October | 19 | 37 | 87 | 39 | 112 |
| November | 27 | 43 | 155 | 57 | 83 |
| December | 32 | 37 | 141 | 50 | 76 |

### Table 2. Grey close function value and the comprehensive analysis value.

| Index | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Comprehensive analysis value ($P_i$) |
|-------|-------|-------|-------|-------|-------|-------------------------------------|
| $S_1$ | 0.225 | 0.630 | 0.651 | 0.625 | 0.731 | 0.651 |
| $S_2$ | 0.300 | 0.531 | 0.335 | 0.455 | 0.589 | 0.464 |
| $S_3$ | 0.500 | 0.630 | 0.352 | 0.532 | 0.539 | 0.481 |
| $S_4$ | 0.643 | 0.739 | 0.373 | 0.556 | 0.469 | 0.482 |
| $S_5$ | 0.643 | 0.708 | 0.757 | 0.962 | 0.427 | 0.641 |
| $S_6$ | 1.000 | 1.000 | 0.691 | 0.862 | 0.475 | 0.673 |
| $S_7$ | 0.900 | 0.850 | 1.000 | 1.000 | 0.507 | 0.787 |
| $S_8$ | 0.692 | 0.630 | 0.862 | 0.962 | 0.589 | 0.736 |
| $S_9$ | 0.474 | 0.459 | 0.644 | 0.641 | 0.679 | 0.631 |
| $S_{10}$ | 0.333 | 0.395 | 0.361 | 0.439 | 0.916 | 0.590 |
| $S_{11}$ | 0.281 | 0.459 | 0.397 | 0.500 | 1.000 | 0.645 |
The following information can be obtained from Table 3. If $\lambda = 0.945$, $S_2$, $S_3$ and $S_4$ correspond to the first classification $S'_1$; $S_7$ and $S_8$ correspond to the third classification $S'_3$; and the other $S$ values correspond to the second classification $S'_2$. $S_2$, $S_3$ and $S_4$ are the samples for March, April and May, respectively, and $S_7$ and $S_8$ are the samples for August and September, respectively. Cluster $S'_{tk}$ (Eq. 13) (Table 4) includes $S'_1$, $S'_2$ and $S'_3$.

The samples (Table 1) can be divided into three classifications, and the class-based approach provides two main advantages. First, if the data in each month are compared and analysed with the air standards, the workload will be large, and errors will easily accumulate. In contrast, only analysing the three classifications can greatly improve the work efficiency. Second, this classification method can be used to establish national or local standards. For example, actual statistical data over many years can be classified by this method, and the classification results can be used as new comparison standards, which would be beneficial to the analysis and evaluation of statistical data in the future.

**Sample evaluation and correlation degree analysis.** In the former parts of the paper, the samples from each month in 2018 are divided into three classifications ($S'_1$, $S'_2$ and $S'_3$). The concentrations of these pollutants in the air quality standard (GB3095-2012) are used for comparison, and the comparison of the data is shown in Fig. 1.

As shown in Fig. 1, compared with that in the first-level air standard, the $SO_2$ concentration in the third-classification air standard is lower, and the $NO_2$ concentrations in the three air classes are all lower than the...
concentration in the first-level air standard. In other words, the concentration of NO₂ in the region meets the first-level air standard throughout the year, and the concentration of SO₂ in August and September meets the first-level air standard. Therefore, according to the first-level air standard, the region should strengthen the management of PM₁₀, PM₂.₅ and O₃ emissions throughout the year, and the management of SO₂ emissions in months other than August and September should be strengthened.

Compared with the second-level air, the concentrations of SO₂, NO₂, O₃ in the three air classes are all lower than that in the second-level standard, the concentrations of PM₁₀ and PM₂.₅ in the third classification of air are lower those in the second-level standard. In other words, the concentrations of NO₂, SO₂ and O₃ in the region meet the second-level air standard throughout the year. Moreover, the concentrations of PM₁₀ and PM₂.₅ in August and September meet the second-level air standard. Therefore, according to the second-level air standard, the region should strengthen the management of PM₁₀ and PM₂.₅ emissions.

According to grey theory, the cluster data and the data (S′₀₁ and S′₀₂ from air quality standard) used for comparison must be initialized⁴,⁵, and the initial values are shown in Table 5.

According to Eqs. 14 and 15, the correlation degree R and the correlation coefficient ζ of the first-level standard are shown in Table 6, and the correlation degree and correlation coefficient of the second-level standard are shown in Table 7.

According to Tables 6 and 7, all three classifications have the highest correlation with the first-level air standard. Therefore, the air quality in Ningxia Province in 2018 was associated with the first-level standard. More importantly, this result quantitatively indicates a correlation between the three classifications and the first-level air standard. The correlation degrees of the first classification, second classification and third classification with the first-level air standard are 0.674, 0.697 and 0.71, respectively. Therefore, it is clear that the gaps between the three classifications and the compared air standard are 0.326, 0.303 and 0.29. Moreover, the reason why the correlation degree cannot reach 1 is that some pollutant concentrations in the monitored data for these classifications are lower than the first-level air standard, and the remaining pollutant values are higher. Therefore, there is still room to continue to improve the air quality in the region. The region should continue to reduce the concentrations of pollutants and further improve the correlation degrees of all classifications of air with the first-level air standards.

Conclusions

(1) A new method of air quality assessment, IGFCCA, is proposed. The advantage of the method is that it can quantitatively characterize the correlation degree between the current air quality and the corresponding standard level. Specifically, the results of this method indicated that the air quality in Ningxia Province in 2018 was correlated with first-level air in China’s air quality standard. The correlation degrees of the first classification, second classification and third classification of air quality with the first-level air standard are 0.674, 0.697 and 0.71, respectively. Therefore, the region should continue to reduce the concentrations of

| Index | x₁ | x₂ | x₃ | x₄ | x₅ |
|-------|----|----|----|----|----|
| S₀₁   | 1.000 | 1.322 | 7.676 | 2.371 | 6.967 |
| S₀₂   | 1.000 | 1.312 | 4.426 | 1.709 | 5.057 |
| S₀₃   | 1.000 | 2.043 | 5.261 | 2.217 | 12.130 |
| S₀₄   | 1.000 | 2.000 | 2.000 | 0.750 | 5.000 |
| S₀₅   | 1.000 | 0.667 | 1.167 | 0.583 | 2.667 |

Table 5. Data initialization results.

| Correlation coefficient and correlation degree | ζ₁ | ζ₂ | ζ₃ | ζ₄ | ζ₅ | R₁ |
|-----------------------------------------------|----|----|----|----|----|----|
| S₁₀   | 1.000 | 0.807 | 0.333 | 0.637 | 0.591 | 0.674 |
| S₁₁   | 1.000 | 0.638 | 0.333 | 0.558 | 0.955 | 0.697 |
| S₁₂   | 1.000 | 0.988 | 0.522 | 0.708 | 0.333 | 0.710 |

Table 6. Correlation with the first-level air standard.

| Correlation coefficient and correlation degree | ζ₁ | ζ₂ | ζ₃ | ζ₄ | ζ₅ | R₁ |
|-----------------------------------------------|----|----|----|----|----|----|
| S₂₀   | 1.000 | 0.832 | 0.333 | 0.646 | 0.431 | 0.648 |
| S₂₁   | 1.000 | 0.716 | 0.333 | 0.591 | 0.405 | 0.609 |
| S₂₂   | 1.000 | 0.775 | 0.536 | 0.743 | 0.333 | 0.677 |

Table 7. Correlation with the second-level air standard.
pollutants, especially PM_{10}, PM_{2.5} and O_3, and further improve the correlation degrees of all classifications with the first-level air standards. Notably, this method can be used in other industries.

(2) The air quality in Ningxia Province in 2018 was classified into three classifications by ideal grey close function cluster analysis. Specifically, the relatively poor air quality in March, April and May and the comparatively better air quality in August and September correspond to the third classification, and the air quality in the remaining months corresponds to the second classification. In addition, the classification method can be used as a reference when establishing other classification standards, such as national standards, regional standards, and industry standards.

Received: 31 March 2021; Accepted: 19 November 2021
Published online: 02 December 2021

References

1. Alfaro-Saiz J.J., Bas M.C., Giner-Bosch V., Rodríguez-Rodríguez R., Verdecho M.J. An evaluation of the environmental factors for supply chain strategy decisions using grey systems and composite indicators. J. Appl. Math. Model. 79, 490–505 (2020).
2. Sarraf, F. & Nejad, S.H. Improving performance evaluation based on balanced scorecard with grey relational analysis and data envelopment analysis approaches: case study in water and wastewater companies. J. Eval. Program. 79, (2020).
3. Xue, J., Van Gelder, P. H. A. J. M., Reniers, G., Papadimitriou, E. & Wu, C. Z. Multi-attribute decision-making method for prioritizing maritime traffic safety influencing factors of autonomous ship maneuvering decisions using grey and fuzzy theories. J. Saf. Sci. 120, 325 (2019).
4. Deng, J. L. Basic Method of Grey System 106–143 (Huazhong University of Science and Technology Press, 2005).
5. Deng, J. L. Grey Control System 90–180 (Huazhong University of Science and Technology Press, 1993).
6. Yu, J. M., Zhang, X. N. & Xiong, C. L. A methodology for evaluating micro-surfacing treatment on asphalt pavement based on grey system models and grey rational degree theory. J. Constr. Build. Mater. 150, 215 (2017).
7. Zhu, C. H. & Li, N. P. Study on grey clustering model of indoor air quality indicators. J. Procedia Eng. 205, 2816 (2017).
8. Zheng, J. Fuzzy mathematics-based comprehensive evaluation of atmospheric environmental quality of Urumqi from 2001 to 2011. J. Environ. Pollut. Control 36, 28–33 (2014).
9. Lü, L. Y. & Li, H. Y. Air quality evaluation of Beijing-Tianjin-Hebei region of China based on the fuzzy comprehensive evaluation method. J. Acta Scientiarum Natural. Universitatis Nankaiensis 49, 62–66 (2016).
10. Pai, T. Y., Hanaki, K. & Chiu, R. J. Forecasting hourly roadside particulate matter in Taipei County of Taiwan based on first-order and one-variable grey model. J. Clean-Soil Air Water 41, 737–742 (2013).
11. Li, X. L. et al. Influence of social-economic activities on air pollutants in Beijing. J. Open Geosci. 9, 314–321 (2017).
12. Wu, X. J., Cheng, C. X., Zurita-Milla, R. & Song, C. Q. An overview of clustering methods for geo-referenced time series: From one-way clustering to co- and tri-clustering. J. Int. J. Geogr. Inf. Sci. 3, 1–27 (2020).
13. Govender, P. & Sivakumar, V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). J. Atmos. Pollut. Res. 11, 1–50 (2019).
14. Ren, S. J. et al. Thermal properties of coal during low temperature oxidation using a grey correlation method. J. Fuel 260, 7 (2020).

Author contributions

All authors have read and agreed to the published version of the manuscript. The division of labor among the authors is as follows: writing—original draft preparation, X.R.; editing, Z.L.; inspecting the calculation process, Y.Z.; revising the format, X.S.; checking the professional terms, X.R.; inspecting the professional terms, X.R.; checking the calculation process, Y.Z.; revising the format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021