



OPEN ACCESS

EDITED BY

Sayali Sandbhor,
Symbiosis International University, India

REVIEWED BY

Komal Shukla,
Environment and Climate Change Canada
(ECCC), Canada
Roseline Oluwaseun Ogundokun,
Landmark University, Nigeria

*CORRESPONDENCE

Bian Chao,
✉ 15526387@qq.com

RECEIVED 21 June 2023

ACCEPTED 14 February 2024

PUBLISHED 02 April 2024

CITATION

Chao B and Guang Qiu H (2024), Air pollution concentration fuzzy evaluation based on evidence theory and the K-nearest neighbor algorithm.
Front. Environ. Sci. 12:1243962.
doi: 10.3389/fenvs.2024.1243962

COPYRIGHT

© 2024 Chao and Guang Qiu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Air pollution concentration fuzzy evaluation based on evidence theory and the K-nearest neighbor algorithm

Bian Chao^{1,2*} and Huang Guang Qiu¹

¹Management College Xi'an University of Architecture and Technology, Xi'an, China, ²Yinchuan Institute of Science and Technology, Yinchuan, China

Background: Air pollution, characterized by complex spatiotemporal dynamics and inherent uncertainty, poses significant challenges in accurate air quality prediction, and current methodologies often fail to adequately address these complexities.

Objective: This study presents a novel fuzzy modeling approach for estimating air pollution concentrations.

Methods: This fuzzy evaluation method integrates an improved evidence theory with comprehensive weighting and the K-nearest neighbor (KNN) interval distance within the framework of the matter-element extension model. This involves generating the basic probability assignment (BPA) based on interval similarity, performing sequential fusion using the Dempster–Shafer evidence theory, enhancing the fusion results via comprehensive weighting, and conducting fuzzy evaluation of air pollution concentrations using the matter-element extension KNN interval distance.

Results: Our method achieved significant improvements in monitoring air pollution concentrations, incorporating spatiotemporal factors and pollutant concentrations more effectively than existing methods. Implementing sequential fusion and subjective–objective weighting reduced the error rate by 38% relative to alternative methods.

Discussion: Fusion of multi-source air pollution data via this method effectively mitigates inherent uncertainty and enhances the accuracy of the KNN method. It produces more comprehensive air pollution concentration fusion results, improving accuracy by considering spatiotemporal correlation, toxicity, and pollution levels. Compared to traditional air-quality indices, our approach achieves greater accuracy and better interpretability, making it possible to develop more effective air quality management strategies. Future research should focus on expanding the dataset to include more diverse geographical and meteorological conditions, further refining the model to integrate external factors like meteorological data and regional industrial activity, and improving computational efficiency for real-time applications.

KEYWORDS

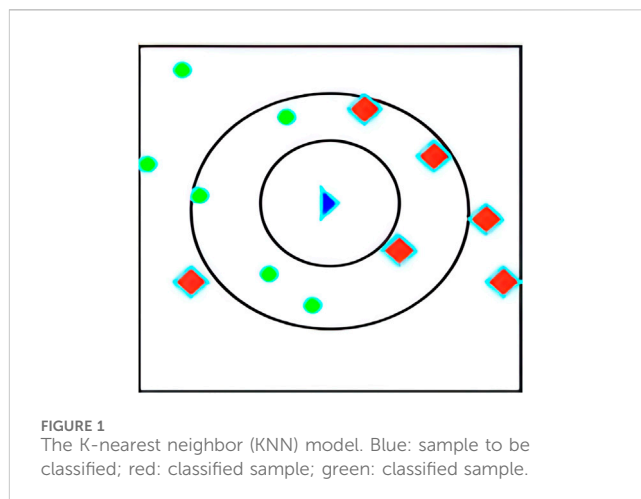
evidence theory, K-nearest neighbor, air pollution concentration fusion, fuzzy evaluation method, basic probability assignment

Abbreviations: AQI, air quality index; BPA, basic probability assignment.

1 Introduction

Atmospheric pollution is an urgent issue globally, especially in regions undergoing rapid industrialization and urbanization. The detrimental effects of pollution on ecological and climatic stability and human health are increasingly evident (Murena, 2004; Chen and Zhu, 2014; Chen et al., 2022). With the increasing global focus on environmental protection and low-carbon development, research is focusing on how technological innovation, green finance, and policy design can reduce environmental pollution and foster sustainable development. For example, Li et al. (2023) explored how digital finance could facilitate green technological innovations in polluting industries by easing funding constraints and augmenting research and development investments. Meanwhile, Feng et al. (2023) focused on reducing NO_x emissions via advanced catalytic technologies, thus improving air quality. The intensification of global climate change has amplified the importance of research into urban heat-island effects (Shang et al., 2023), biodiversity loss (Wang et al., 2022), and the global carbon cycle (Zhang et al., 2021; Xiong et al., 2022).

Within this context, Wu et al. (2023) investigated the impact of clustered institutional investors on low-carbon innovation in family businesses, discussing how green finance and family governance can promote sustainable development in a changing economic environment. Kong et al. (2023) proposed a lifecycle-oriented low-carbon product design method to meet the challenges of climate change. Using cloud computing technology, Shang et al.



(2021) explored factors influencing urban carbon footprints in China, and proposed key strategies for optimizing carbon emission predictions and low-carbon economic development (Luo et al., 2024). These studies provide technological and financial solutions while highlighting the significance of policy design in advancing environmental protection and sustainable development, which require accurate assessment of atmospheric pollution.

The air quality index (AQI) is a widely accepted assessment tool for air pollution; however, it presents several limitations. For instance, by aggregating the concentration levels of multiple

TABLE 1 Literature review of related studies.

References	Method	Advantage	Disadvantage
Murena (2004)	Involved developing an air quality index (AQI) through principal component analysis and linear algebra to rank states by pollution levels based on data from five pollutants in the transportation sector	Furnished an initial understanding of the impact of air pollution in industrialized and urbanized areas	Possibly lacked granular data and in-depth analysis of complex pollutant interactions
Chen and Zhu (2014)	Delved deeper into the air pollution issues of industrialized and urbanized areas	Focused on high-risk areas, delivering more specific data on atmospheric pollution	Centered on particular regions, limiting the broader applicability of the research findings
Pope III and Dockery (2006)	Leveraged more precise data and measurement methodologies to study atmospheric pollution	Provided high-quality data, advancing the accuracy of air pollution research	Used many resources and technology to acquire and analyze data
Vaidya and Kumar (2006)	Evaluated and critiqued the limitations and deficiencies of the analytic hierarchy process (AHP)	Exposed the constraints of AHP, promoting the development of new methodologies and techniques	Primarily represented a critical study, offering no explicit solutions
Li et al. (2010)	Investigated the performance of the grey relational analysis when dealing with intricate, non-linear environmental variables	Disclosed the method's applicability in certain contexts	Faced challenges in data integrity and precision when handling complex, non-linear environmental elements
Carslaw and Rhys-Tyler (2013)	Employed deep learning technologies to predict concentrations of NO ₂ and PM ₁₀	Enhanced prediction accuracy and adaptability, capable of processing intricate datasets	Required substantial data and computational resources
Cui et al. (2022)	Criticized the AQI for its shortcomings in handling complex atmospheric pollutants	Pointed out the limitations of AQI, offering a direction for the improvement and development of new assessment standards	Presented a critical study that failed to provide specific solutions or measures for improvement
Sun et al. (2022b)	Adopted the improved D-S evidence theory for a comprehensive assessment of air quality; determined the weights of evidence through the entropy weight method and introduced decision credibility by calculating the dispersion of different evidence decisions	Validated the efficacy of the DCre-Weight model, which showed higher accuracy and consistency than other enhanced methods of the evidence theory and recent fuzzy synthetic evaluation methods; yielded improved credibility of fusion results and well-articulated uncertainty	Presented limitations in complexity, applicability, and comparison with other latest methods

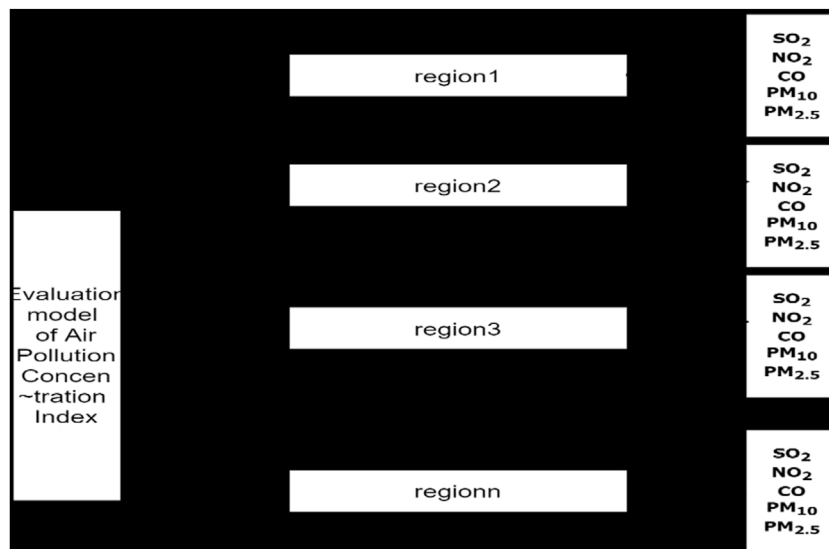


FIGURE 2
Pollution concentration evaluation factor set.

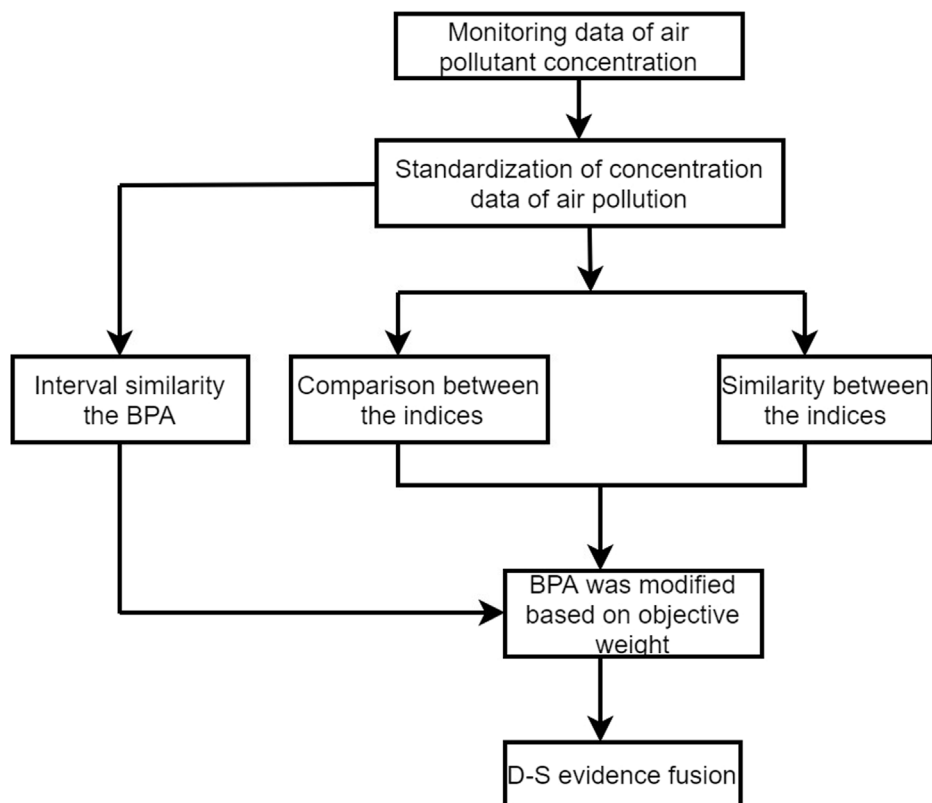


FIGURE 3
Air pollution concentration evaluation model.

pollutants into a single metric, the AQI may obscure the risks of high concentrations of specific pollutants (Pope et al., 2022). Further complications arise from the fact that not all pollutants

are considered in AQI computations (Priti and Kumar, 2022), and the AQI calculations and standards differ between countries, leading to interpretative challenges (Karavas et al.,

TABLE 2 Air pollution concentration standard.

Pollutant	Air pollution concentration levels ($\mu\text{g}/\text{m}^3$)						
	I	II	III	IV	V	VI	VII
PM _{2.5}	35	75	115	150	250	350	500
PM ₁₀	50	150	250	350	420	500	600
SO ₂	50	150	475	800	1,600	2,100	2,620
NO ₂	40	80	180	280	565	750	940
CO	2	4	14	24	36	48	60
O ₃ ₈	100	160	215	265	800	1,000	1,200

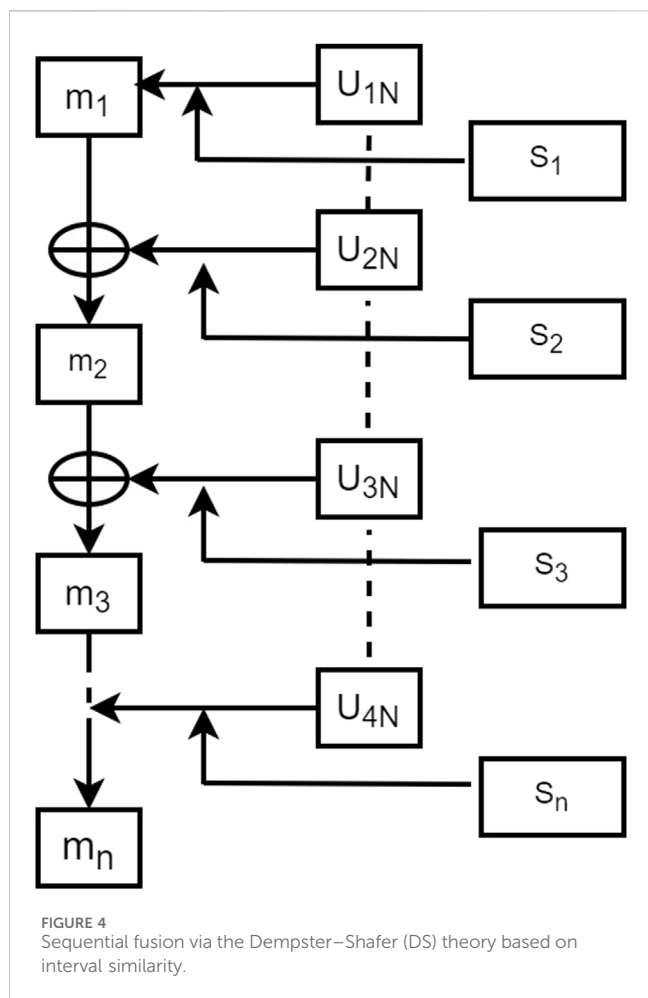


FIGURE 4 Sequential fusion via the Dempster–Shafer (DS) theory based on interval similarity.

2021). Moreover, the AQI is inadequate for capturing and processing the complexities of atmospheric pollutants and their interactions (Cui et al., 2022). For instance, Zhang and Cao (2015) examined the heterogeneity and complexity of PM_{2.5} levels, highlighting the shortcomings of current assessment strategies. Further, accurate data and measurement techniques are indispensable for controlling air pollution (Pope III and Dockery, 2006).

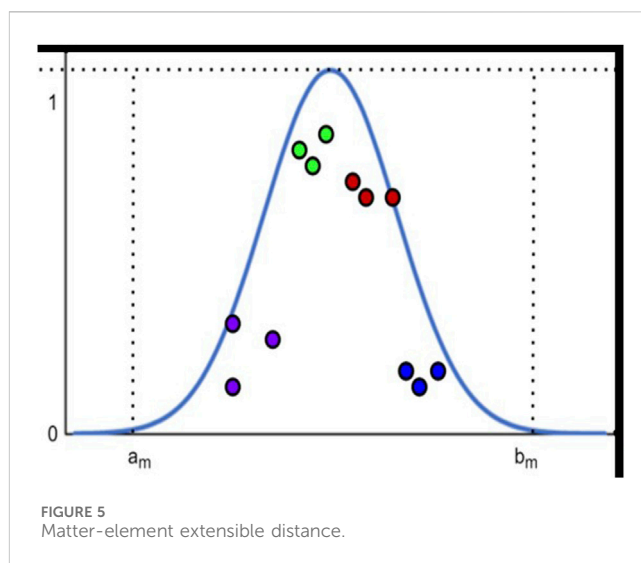


FIGURE 5 Matter-element extensible distance.

TABLE 3 Nondimensionalization of the air pollution concentration standards.

Pollutant	Air pollution concentration levels						
	I	II	III	IV	V	VI	VII
PM _{2.5}	0.17	0.36	0.55	0.71	1.19	1.66	2.37
PM ₁₀	0.15	0.45	0.75	1.06	1.27	1.51	1.81
SO ₂	0.04	0.13	0.43	0.72	1.44	1.89	2.35
NO ₂	0.10	0.20	0.44	0.69	1.40	1.85	2.32
CO	0.07	0.15	0.52	0.89	1.34	1.79	2.23
O ₃ ₈	0.28	0.34	0.52	0.69	1.38	1.72	2.07

Vaidya and Kumar (2006) expressed concerns about the accuracy and consistency of other traditional assessment methods, such as the analytic hierarchy process, under complex, dynamic, and variable environmental conditions. While the analytic hierarchy process, a multi-criteria decision-making tool, is used to determine the significance of various pollutants (Saaty, 2008), its reliance on subjective expert judgments may introduce bias (Dyer and Forman, 1992). Grey relational analysis may be suited to specific scenarios, but it presents challenges in model selection and membership function when dealing with complex, non-linear environmental factors (Wang and Klir, 2009; Li et al., 2010). Quantification of uncertain information was proposed by Zadeh (1965), and the fuzzy comprehensive evaluation method (Mo et al., 2020) offered a new approach to handle ambiguous and uncertain data. However, challenges persist regarding the scientific rigor and reproducibility of data generated using the analytic hierarchy process.

Considering these issues, attention has shifted towards more advanced and innovative assessment techniques. The potential of evidence theory and K-nearest neighbor (KNN) algorithms to handle complex and dynamic air quality data has been widely

TABLE 4 Concentrations of different air pollutants in different regions.xi'an.

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
Xingqing District	11	70	1.3	21	99	64
Xiaozhai	12	76	1	26	96	63
Municipal People's Stadium	8	80	1.1	22	121	76
Kwong Wan Tam	11	66	1.2	20	118	70
Bureau of Culture and Sports	10	51	1.2	48	118	97
Radio monitoring center	20	59	1	27	132	98
Qujiang	8	62	1	24	114	60
Textile City	13	45	1.3	18	123	74
Economic Development Zone	12	74	1	26	110	83
Grass Beach (control point)	10	55	0.9	27	168	99
Chang'an District	9	52	0.9	44	95	62
High-voltage switch factory	11	66	0.9	27	120	84
High-tech Western District	11	71	1	25	99	77
Lintong District	15	67	1.2	34	97	67
Xingqing District	18	84	2	16	186	126
Xiaozhai	20	94	1.5	11	174	120
Municipal People's Stadium	18	91	1.7	21	199	128
Kwong Wan Tam	21	71	2	25	191	142
Bureau of Culture and Sports	16	70	1.5	28	192	152
Radio monitoring center	28	57	1.4	46	146	117
Qujiang	12	81	1.5	8	194	112
Textile City	19	56	2	12	194	138
Economic Development Zone	26	82	1.7	30	169	133
Grass Beach (control point)	19	66	1.6	27	203	129
Chang'an District	14	75	1.5	13	192	124
High-voltage switch factory	23	82	1.5	21	201	135
High-tech Western District	22	88	1.6	14	173	127
Lintong District	24	72	2.2	48	167	120
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Xingqing District	10	75	1.1	33	98	57
Xiaozhai	13	80	1.4	35	150	60
Municipal People's Stadium	11	82	1.4	41	122	60
Kwong Wan Tam	10	76	0.8	19	109	65
Bureau of Culture and Sports	11	55	0.9	63	106	61
Radio monitoring center	13	58	1.1	27	196	118

(Continued on following page)

TABLE 4 (Continued) Concentrations of different air pollutants in different regions.xi'an.

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
Qujiang	10	71	1.1	23	192	75
Textile City	8	71	1	22	113	62
Economic Development Zone	13	84	1.1	39	167	74
Grass Beach (control point)	8	68	1	29	149	85
Chang'an District	11	55	0.8	53	129	54
High-voltage switch factory	14	80	1.2	54	156	80
High-tech Western District	13	77	1.3	55	124	73
Lintong District	14	73	1	31	127	76

TABLE 5 Concentrations of different air pollutants in different regions.beijing.

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
Yanqing Summer Capital	9	31	0.8	49	69	31
Miyun New City	8	21	0.5	50	43	19
Pinggu New City	4	30	0.8	29	65	34
Miyun Town	7	38	0.7	36	54	23
Fengtai Xiaotun	8	64	1	21	104	39
Huairou New City	6	23	0.5	40	42	19
Yanqing Shiheyang	9	32	0.9	46	62	32
Daxing Old Palace	4	61	0.7	16	80	33
Fangshan and Yanshan	6	36	0.8	48	56	31
Tongzhou Dongguan	5	57	0.6	15	85	36
Fengtai Yungang	7	46	0.6	22	66	29
Three stores in Mentougou	8	43	0.9	32	70	29
ancient city	6	47	0.8	20	95	38
Olympic Sports Center	7	54	0.6	25	61	27
Changping Town	8	34	0.7	26	52	25
Huairou Town	5	30	0.6	38	39	20
Shunyi New City	7	42	0.8	28	60	29
Haidian Wanliu	10	53	0.9	22	76	30
Official Garden	7	59	0.8	21	62	29
Agricultural Exhibition Hall	5	63	0.6	18	74	30
Temple of Heaven	3	58	0.9	15	58	32
Dongsi	4	52	0.6	18	63	30
Dingling (control point)	7	30	0.5	28	46	28
Wanshou West Palace	5	63	0.9	17	67	32

(Continued on following page)

TABLE 5 (Continued) Concentrations of different air pollutants in different regions.beijing.

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Yanqing Summer Capital	5	39	0.6	36	47	20
Miyun New City	2	68	0.8	25	68	26
Pinggu New City	3	19	0.3	46	25	15
Miyun Town	3	36	0.6	38	39	19
Fengtai Xiaotun	4	59	0.8	23	51	24
Huairou New City	5	37	0.5	37	35	20
Yanqing Shiheying	5	65	0.7	26	68	28
Daxing Old Palace	3	52	0.5	30	48	21
Fangshan and Yanshan	3	45	1.1	27	65	24
Tongzhou Dongguan	2	26	0.4	42	43	17
Fengtai Yungang	3	29	0.6	42	41	19
Three stores in Mentougou	4	37	0.6	37	40	21
ancient city	2	57	0.7	24	59	23
Olympic Sports Center	5	67	0.7	23	51	22
Changping Town	2	37	0.4	33	37	16
Huairou Town	3	25	0.4	41	30	15
Shunyi New City	3	47	0.7	35	42	20
Haidian Wanliu	4	65	0.7	19	57	21
Official Garden	6	65	0.7	22	46	21
Agricultural Exhibition Hall	5	67	0.8	20	61	29
Temple of Heaven	3	56	0.7	28	36	21
Dongsi	3	57	0.8	24	50	23
Dingling (control point)	2	21	0.4	42	29	14
Wanshou West Palace	6	71	0.9	23	49	23

explored (Cover and Hart, 1967; Dempster, 1967; Shafer, 1976; Xiao et al., 2013; Dai et al., 2018; Wang et al., 2021; Sun et al., 2022a; Franklin et al., 2023). However, their efficacy under various geographical and meteorological conditions remains questionable. The research of Carslaw and Rhys-Tyler (2013), and the application of deep learning in predicting the concentrations of pollutants such as NO₂ and PM₁₀ (Kukkonen et al., 2003), signal a shift towards more accurate and flexible assessment techniques. As a significant improvement, Sun et al. (2022b) presented an ambient air quality evaluation model based on an advanced evidence theory. Important research developments in this field are presented in Table 1. While advanced air quality assessment techniques provide significant insights, they are often unable to distinguish between sources of pollution and may not account for the latest scientific understanding

of pollutant interactions. Moreover, the evolving nature of industrial emissions and urban development requires more adaptable and sophisticated analytical methods. Thus, recognizing these shortcomings, the development and deployment of more precise and efficient air quality assessment techniques have become paramount for both scientific research and policy formulation.

In this study, we present a fuzzy evaluation method that innovatively integrates evidence theory and the KNN algorithm, aiming to enhance atmospheric pollution concentration assessment precision and efficiency. This approach strives to overcome the limitations of conventional evaluation methods by synthesizing multifaceted, uncertain, and ambiguous environmental data, thus improving the precision and reliability. Via comparison with established evaluation techniques, our study reveals the proposed

TABLE 6 Concentrations of different air pollutants in different regions.tianjin

Region	Pollutant concentration					
	SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	CO (mg/m ³)	O ₃ (μg/m ³)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
BinshuiEastRoad	22	57	1.6	20	97	50
JiansheRoad	18	67	1.5	23	79	43
NorthRingRoad	10	36	1.2	45	56	29
Xisido	18	59	1.2	20	85	50
ZhongshanNorthRoad	14	64	1.4	14	95	49
DaliRoad	17	65	1.3	18	81	52
BinshuiWestRoad	15	57	1.2	14	84	52
JinguRoad	18	67	1.4	16	92	59
HexiYijingRoad	21	57	1.5	23	78	45
DiweiRoad	19	59	1.7	24	93	67
XinlaoRoad	20	74	1.9	19	102	66
YongyangWestRoad	15	61	1.5	20	90	47
Tuanpowa	10	64	1.4	18	93	59
HanbeiRoad	20	52	1.7	19	78	44
YongmingRoad	23	72	1.6	19	88	58
FourthStreet	20	64	1.9	14	77	53
LeapForwardRoad	14	68	1.5	20	99	57
HuaiheRoad	16	55	1	17	81	47
Forwardlane	19	66	1.4	20	94	56
DazhiguNo8Road	16	67	1.5	16	70	53
Thepathofdiligenceandfrugality	20	68	1.3	15	84	44
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Yanqing Summer Capital	18	46	0.8	41	54	22
Miyun New City	14	54	0.9	35	63	31
Pinggu New City	11	28	0.9	47	40	18
Miyun Town	12	60	0.8	32	52	25
Fengtai Xiaotun	14	70	1.1	35	82	34
Huairou New City	18	60	0.9	34	90	48
Yanqing Shiheyang	12	63	0.9	37	73	33
Daxing Old Palace	29	67	0.8	31	72	34
Fangshan and Yanshan	15	54	1.1	38	59	28
Tongzhou Dongguan	9	59	1.1	38	57	27
Fengtai Yungang	14	57	1.1	40	57	25
Three stores in Mentougou	11	46	0.7	50	39	18
ancient city	12	65	0.6	37	71	31

(Continued on following page)

TABLE 6 (Continued) Concentrations of different air pollutants in different regions.tianjin

Region	Pollutant concentration					
	SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	CO (mg/m ³)	O ₃ (μg/m ³)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
Olympic Sports Center	9	47	0.5	48	38	16
Changping Town	14	58	0.8	40	62	30
Huairou Town	19	70	0.7	29	57	25
Shunyi New City	12	62	0.9	37	75	33
Haidian Wanliu	15	71	1.3	34	78	33
Official Garden	14	63	1.1	40	64	30
Agricultural Exhibition Hall	13	65	0.8	36	65	30
Temple of Heaven	16	67	1.3	40	70	31

method’s advantages and contributions, which can be summarized as follows:

- Holistic evaluation of atmospheric pollution concentrations: The approach introduces an assessment technique that overcomes the constraints of traditional AQI. It integrates multi-source data and physical property analyses, comprehensively evaluates atmospheric pollution patterns, and accurately quantifies key pollutant concentrations.
- Innovative synthesis of evidence theory and KNN: This synthesis enhances the precision of data fusion and relevance of the results.
- Focus on data quality and accuracy: This hybrid approach emphasizes data integrity by applying interval similarity and subjective-objective weighting and considering variability in pollutant concentrations and toxicological characteristics.

To validate the robustness of this approach and its applicability in diverse environmental contexts, the analysis includes data from representative days across three distinct urban settings: Xi’an, Beijing, and Tianjin, rather than focusing on a single urban area. These cities were chosen for their varied geographic and meteorological profiles, thus enhancing the robustness of the findings. The findings confirm the method’s efficacy across varied urban environments and address complex regional differences in air pollution. This multifaceted approach substantially enhances methodological rigor in air quality assessment.

2 Materials and methods

2.1 Theoretical background

2.1.1 KNN method

The KNN method, a commonly used machine learning algorithm for classification and regression, stores all available cases and classifies new data or cases based on a similarity (distance) measure. It functions by finding the K-nearest neighbors to the unknown sample within the known samples (Altman, 1992; Huihui and Yanming, 2013; Wang et al., 2021). Next, the class of the unknown sample is determined based on the

class of the nearest neighbors, typically via “majority rule.” The three elements of the KNN model are the choice of k , a distance measurement, and a classification-decision rule.

Regarding the choice of k , the blue triangle in Figure 1 represents the sample to be classified. When $k = 7$, of the seven samples nearest the sample to be classified, three belong to the red class and four to the green class. Therefore, the sample to be classified is predicted to belong to the green class. However, when $k = 9$, the sample to be classified belongs to the red class. Therefore, when the samples are unbalanced, the choice of k substantially impacts the results.

Next, distance is measured as follows:

Let the feature space χ be an n -dimensional real vector space R^n , x_i and $x_j \in \chi$, $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$, $x_j = (x_{j1}, x_{j2}, \dots, x_{jn})^T$, with the distance d_p between x_i and x_j defined as

$$d_p(x_i, x_j) = \left(\sum_{l=1}^n |x_{il} - x_{jl}|^p \right)^{\frac{1}{p}} \quad (1)$$

When $p = 2$, this is the Euclidean distance:

$$d_2(x_i, x_j) = \left(\sum_{l=1}^n |x_{il} - x_{jl}|^2 \right)^{\frac{1}{2}} \quad (2)$$

when $p = 1$, it is the Manhattan distance:

$$d_1(x_i, x_j) = \left(\sum_{l=1}^n |x_{il} - x_{jl}| \right) \quad (3)$$

and when $p = \infty$, it is the maximum distance between the coordinates, i.e.,

$$d_\infty(x_i, x_j) = \max_l \left(|x_{il} - x_{jl}| \right) \quad (4)$$

2.1.2 Evidence theory

The evidence theory, proposed by Dempster (2008), has been further developed by Shafer (Ai et al., 2022; He et al., 2022; Liu et al., 2022; Ren et al., 2022). Evidence theory applies fuzzy logic to handle uncertainty, via the following steps.

- Establishment of a discernment framework

The discernment framework is a set of all objects or entities under consideration. Subsets in the discernment framework are defined as follows:

TABLE 7 Standardized air pollutant concentrations across different regions.xi'an.

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Xingqing District	0.25	0.71	1.00	0.10	0.05	0.10
Xiaozhai	0.33	0.89	0.25	0.27	0.01	0.08
Municipal People's Stadium	0.00	1.00	0.50	0.13	0.36	0.41
Kwong Wan Tam	0.25	0.60	0.75	0.07	0.32	0.26
Bureau of Culture and Sports	0.17	0.17	0.75	1.00	0.32	0.95
Radio monitoring center	1.00	0.40	0.25	0.30	0.51	0.97
Qujiang	0.00	0.49	0.25	0.20	0.26	0.00
Textile City	0.42	0.00	1.00	0.00	0.38	0.36
Economic Development Zone	0.33	0.83	0.25	0.27	0.21	0.59
Grass Beach (control point)	0.17	0.29	0.00	0.30	1.00	1.00
Chang'an District	0.08	0.20	0.00	0.87	0.00	0.05
High-voltage switch factory	0.25	0.60	0.00	0.30	0.34	0.62
High-tech Western District	0.25	0.74	0.25	0.23	0.05	0.44
Lintong District	0.58	0.63	0.75	0.53	0.03	0.18
Xingqing District	0.38	0.74	0.75	0.20	0.70	0.35
Xiaozhai	0.50	1.00	0.13	0.08	0.49	0.20
Municipal People's Stadium	0.38	0.92	0.38	0.33	0.93	0.40
Kwong Wan Tam	0.56	0.39	0.75	0.43	0.79	0.75
Bureau of Culture and Sports	0.25	0.37	0.13	0.50	0.81	1.00
Radio monitoring center	1.00	0.03	0.00	0.95	0.00	0.13
Qujiang	0.00	0.66	0.13	0.00	0.84	0.00
Textile City	0.44	0.00	0.75	0.10	0.84	0.65
Economic Development Zone	0.88	0.68	0.38	0.55	0.40	0.53
Grass Beach (control point)	0.44	0.26	0.25	0.48	1.00	0.43
Chang'an District	0.13	0.50	0.13	0.13	0.81	0.30
High-voltage switch factory	0.69	0.68	0.13	0.33	0.96	0.58
High-tech Western District	0.63	0.84	0.25	0.15	0.47	0.38
Lintong District	0.75	0.42	1.00	1.00	0.37	0.20
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Xingqing District	0.33	0.69	0.50	0.32	0.00	0.05
Xiaozhai	0.83	0.86	1.00	0.36	0.53	0.09
Municipal People's Stadium	0.50	0.93	1.00	0.50	0.24	0.09
Kwong Wan Tam	0.33	0.72	0.00	0.00	0.11	0.17
Bureau of Culture and Sports	0.50	0.00	0.17	1.00	0.08	0.11
Radio monitoring center	0.83	0.10	0.50	0.18	1.00	1.00

(Continued on following page)

TABLE 7 (Continued) Standardized air pollutant concentrations across different regions.xi'an.

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Qujiang	0.33	0.55	0.50	0.09	0.96	0.33
Textile City	0.00	0.55	0.33	0.07	0.15	0.13
Economic Development Zone	0.83	1.00	0.50	0.45	0.70	0.31
Grass Beach (control point)	0.00	0.45	0.33	0.23	149	85
Chang'an District	0.50	0.00	0.00	0.77	129	54
High-voltage switch factory	1.00	0.86	0.67	0.80	156	80
High-tech Western District	0.83	0.76	0.83	0.82	124	73
Lintong District	1.00	0.62	0.33	0.27	127	76

TABLE 8 Standardized air pollutant concentrations across different regions.beijing

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
Yanqing Summer Capital	0.86	0.23	0.60	0.97	0.46	0.60
Miyun New City	0.71	0.00	0.00	1.00	0.06	0.00
Pinggu New City	0.14	0.21	0.60	0.40	0.40	0.75
Miyun Town	0.57	0.40	0.40	0.60	0.23	0.20
Fengtai Xiaotun	0.71	1.00	1.00	0.17	1.00	1.00
Huairou New City	0.43	0.05	0.00	0.71	0.05	0.00
Yanqing Shiheyang	0.86	0.26	0.80	0.89	0.35	0.65
Daxing Old Palace	0.14	0.93	0.40	0.03	0.63	0.70
Fangshan and Yanshan	0.43	0.35	0.60	0.94	0.26	0.60
Tongzhou Dongguan	0.29	0.84	0.20	0.00	0.71	0.85
Fengtai Yungang	0.57	0.58	0.20	0.20	0.42	0.50
Three stores in Mentougou	0.71	0.51	0.80	0.49	0.48	0.50
ancient city	0.43	0.60	0.60	0.14	0.86	0.95
Olympic Sports Center	0.57	0.77	0.20	0.29	0.34	0.40
Changping Town	0.71	0.30	0.40	0.31	0.20	0.30
Huairou Town	0.29	0.21	0.20	0.66	0.00	0.05
Shunyi New City	0.57	0.49	0.60	0.37	0.32	0.50
Haidian Wanliu	1.00	0.74	0.80	0.20	0.57	0.55
Official Garden	0.57	0.88	0.60	0.17	0.35	0.50
Agricultural Exhibition Hall	0.29	0.98	0.20	0.09	0.54	0.55
Temple of Heaven	0.00	0.86	0.80	0.00	0.29	0.65
Dongsi	0.14	0.72	0.20	0.09	0.37	0.55
Dingling (control point)	0.57	0.21	0.00	0.37	0.11	0.45
Wanshou West Palace	0.29	0.98	0.80	0.06	0.43	0.65
:	:	:	:	:	:	:

(Continued on following page)

TABLE 8 (Continued) Standardized air pollutant concentrations across different regions.beijing

Region	Pollutant concentration					
	SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	CO (mg/m ³)	O ₃ (μg/m ³)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
:	:	:	:	:	:	:
Yanqing Summer Capital	0.75	0.38	0.38	0.63	0.51	0.40
Miyun New City	0.00	0.94	0.63	0.22	1.00	0.80
Pinggu New City	0.25	0.00	0.00	1.00	0.00	0.07
Miyun Town	0.25	0.33	0.38	0.70	0.33	0.33
Fengtai Xiaotun	0.50	0.77	0.63	0.15	0.60	0.67
Huairou New City	0.75	0.35	0.25	0.67	0.23	0.40
Yanqing Shiheyang	0.75	0.88	0.50	0.26	1.00	0.93
Daxing Old Palace	0.25	0.63	0.25	0.41	0.53	0.47
Fangshan and Yanshan	0.25	0.50	1.00	0.30	0.93	0.67
Tongzhou Dongguan	0.00	0.13	0.13	0.85	0.42	0.20
Fengtai Yungang	0.25	0.19	0.38	0.85	0.37	0.33
Three stores in Mentougou	0.50	0.35	0.38	0.67	0.35	0.47
ancient city	0.00	0.73	0.50	0.19	0.79	0.60
Olympic Sports Center	0.75	0.92	0.50	0.15	0.60	0.53
Changping Town	0.00	0.35	0.13	0.52	0.28	0.13
Huairou Town	0.25	0.12	0.13	0.81	0.12	0.07
Shunyi New City	0.25	0.54	0.50	0.59	0.40	0.40
Haidian Wanliu	0.50	0.88	0.50	0.00	0.74	0.47
Official Garden	1.00	0.88	0.50	0.11	0.49	0.47
Agricultural Exhibition Hall	0.75	0.92	0.63	0.04	0.84	1.00
Temple of Heaven	0.25	0.71	0.50	0.33	0.26	0.47
Dongsi	0.25	0.73	0.63	0.19	0.58	0.60
Dingling (control point)	0.00	0.04	0.13	0.85	0.09	0.00
Wanshou West Palace	1.00	1.00	0.75	0.15	0.56	0.60

$$\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_N\} \quad (5)$$

If Θ is a finite complete set of N mutually exclusive elements, it is called a discernment framework. The set of $2N$ elements formed by the power set 2^Θ of Θ is:

$$2^\Theta = \{\varphi, \theta_1, \theta_2, \theta_3, \dots, \theta_N, \theta_1 \cup \theta_2 \cup \theta_3, \dots, \Theta\} \quad (6)$$

b) Determination of the basic probability assignment (BPA)

Initially, the evidence-based confidence (support) in each combination is established by the evidence processor. The basic probability is assigned as follows:

Let Θ be a discernment framework. The power set 2^Θ of Θ forms the set of propositions $2^\Theta, \forall A \in \Theta$.

If the function $m: 2^\Theta \rightarrow [0, 1]$ satisfies the conditions

$$m(\varphi) = 0 \text{ and } \sum_{A \in \Theta} m(A) = 1 \quad (7)$$

then m is referred to as the BPA, and $m(A)$ is the basic probability assigned to proposition A (i.e., the confidence accurately assigned to A).

There are two ways to calculate the BPA: via expert opinion or by constructing the corresponding mathematical model. The approach of using expert experience requires values provided by different experts, and thus tends to be subjective. Therefore, this method calculates BPA by constructing the corresponding mathematical model.

c) Evidence-combination rule

TABLE 9 Standardized air pollutant concentrations across different regions.tianjing

Region	Pollutant concentration					
	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (mg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	PM _{2.5} (µg/m ³)
BinshuiEastRoad	0.92	0.55	0.67	0.19	0.89	0.55
JiansheRoad	0.62	0.82	0.56	0.29	0.50	0.37
NorthRingRoad	0.00	0.00	0.22	1.00	0.00	0.00
Xisido	0.62	0.61	0.22	0.19	0.63	0.55
ZhongshanNorthRoad	0.31	0.74	0.44	0.00	0.85	0.53
DaliRoad	0.54	0.76	0.33	0.13	0.54	0.61
BinshuiWestRoad	0.38	0.55	0.22	0.00	0.61	0.61
JinguRoad	0.62	0.82	0.44	0.06	0.78	0.79
HexiYijingRoad	0.85	0.55	0.56	0.29	0.48	0.42
DiweiRoad	0.69	0.61	0.78	0.32	0.80	1.00
XinlaoRoad	0.77	1.00	1.00	0.16	1.00	0.97
YongyangWestRoad	0.38	0.66	0.56	0.19	0.74	0.47
Tuanpowa	0.00	0.74	0.44	0.13	0.80	0.79
HanbeiRoad	0.77	0.42	0.78	0.16	0.48	0.39
YongmingRoad	1.00	0.95	0.67	0.16	0.70	0.76
FourthStreet	0.77	0.74	1.00	0.00	0.46	0.63
LeapForwardRoad	0.31	0.84	0.56	0.19	0.93	0.74
HuaiheRoad	0.46	0.50	0.00	0.10	0.54	0.47
Forwardlane	0.69	0.79	0.44	0.19	0.83	0.71
DazhiguNo8Road	0.46	0.82	0.56	0.06	0.30	0.63
Thepathofdiligenceandfrugality	0.77	0.84	0.33	0.03	0.61	0.39
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Yanqing Summer Capital	0.45	0.42	0.38	0.57	0.31	0.19
Miyun New City	0.25	0.60	0.50	0.29	0.48	0.47
Pinggu New City	0.10	0.00	0.50	0.86	0.04	0.06
Miyun Town	0.15	0.74	0.38	0.14	0.27	0.28
Fengtai Xiaotun	0.25	0.98	0.75	0.29	0.85	0.56
Huairou New City	0.45	0.74	0.50	0.24	1.00	1.00
Yanqing Shiheyang	0.15	0.81	0.50	0.38	0.67	0.53
Daxing Old Palace	1.00	0.91	0.38	0.10	0.65	0.56
Fangshan and Yanshan	0.30	0.60	0.75	0.43	0.40	0.38
Tongzhou Dongguan	0.00	0.72	0.75	0.43	0.37	0.34
Fengtai Yungang	0.25	0.67	0.75	0.52	0.37	0.28
Three stores in Mentougou	0.10	0.42	0.25	1.00	0.02	0.06
ancient city	0.15	0.86	0.13	0.38	0.63	0.47

(Continued on following page)

TABLE 9 (Continued) Standardized air pollutant concentrations across different regions.tianjing

Region	Pollutant concentration					
	SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	CO (mg/m ³)	O ₃ (μg/m ³)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)
Olympic Sports Center	0.00	0.44	0.00	0.90	0.00	0.00
Changping Town	0.25	0.70	0.38	0.52	0.46	0.44
Huairou Town	0.50	0.98	0.25	0.00	0.37	0.28
Shunyi New City	0.15	0.79	0.50	0.38	0.71	0.53
Haidian Wanliu	0.30	1.00	1.00	0.24	0.77	0.53
Official Garden	0.25	0.81	0.75	0.52	0.50	0.44
Agricultural Exhibition Hall	0.20	0.86	0.38	0.33	0.52	0.44
Temple of Heaven	0.35	0.91	1.00	0.52	0.62	0.47

Let m_A and m_B be sets of BPAs corresponding to focal elements A_1, A_2, \dots, A_i and B_1, B_2, \dots, B_j , respectively. Let m denote the new evidence after combining m_A and m_B . Then, the Dempster combination rule is expressed as follows:

$$m(\emptyset) = 0 \tag{8}$$

$$m(A) = \frac{1}{1-k} \sum_{(A_i \cap B_j) \neq \emptyset} m_A(A_i) \cdot m_B(B_j) \tag{9}$$

$$k = \sum_{(A_i \cap B_j) \neq \emptyset} m_A(A_i) \cdot m_B(B_j) \tag{10}$$

where K is the conflict coefficient, which reflects the degree of conflict between focal elements. A larger K indicates greater conflict; the combination rule cannot be used when $K = 1$.

2.1.3 Support vector machine (SVM)

We derived a data fusion method to assess air pollution based on comprehensive fuzzy evaluation, the Dempster–Shafer (DS) evidence theory, and the KNN algorithm. To comprehensively evaluate this method, we chose the SVM method as a benchmark for comparison. SVM, a widely used supervised learning method for classification and regression, is based on structural risk minimization. SVM can simultaneously process linear and non-linear data via kernel functions, making it suitable for analyzing our dataset.

2.2 Model overview

2.2.1 Research concept

The atmospheric environment is complex, and air quality is influenced by many factors, including uncertain and fuzzy factors (Seinfeld and Pandis, 2016). To address this environmental complexity, we propose an air pollution concentration evaluation method based on the DS evidence theory corrected by subjective and objective weighting and the extensible KNN. First, the set of air pollution evaluation indicators is determined (Figure 2). Figure 3 presents the air-pollution modeling methodology and workflow. The workflow encompasses each phase from the initial data segmentation to the final evaluation of pollution levels and elucidates the seamless transition between phases. This method thus provides a cohesive approach to quantifying air pollution.

2.2.2 Research procedure

Comprehensive examination and iterative testing have revealed that accurate quantification of air pollution requires consideration of both the pollution intensity for specific contaminants and the spatiotemporal dynamics of the atmosphere. We therefore designed and refined an optimized algorithm combining evidence theory-based data fusion and KNN-based fuzzy evaluation. This algorithm was not chosen arbitrarily; it emerged as the most effective after several experimental iterations and methodological trials. Its efficacy in addressing the complexity of pollution assessment underscores its robustness. This integrated algorithm is deployed by applying evidence theory, the KNN algorithm, and then their synthesis.

2.2.2.1 Evidence theory application

For data segmentation, the air pollution data are initially divided into different intervals, based on national air pollution concentration standards, to ensure consistency and accuracy in evaluation.

Basic Probability Assignment (BPA) values are constructed for each pollutant, based on interval similarity. This step underlies evidence theory fusion, with the key objective being the accurate representation of uncertainty in pollutant concentration.

For each pollutant, evidence theory is applied to sequentially fuse the concentration data from each region, thereby enhancing the model's capacity to handle regional disparities.

2.2.2.2 KNN algorithm application

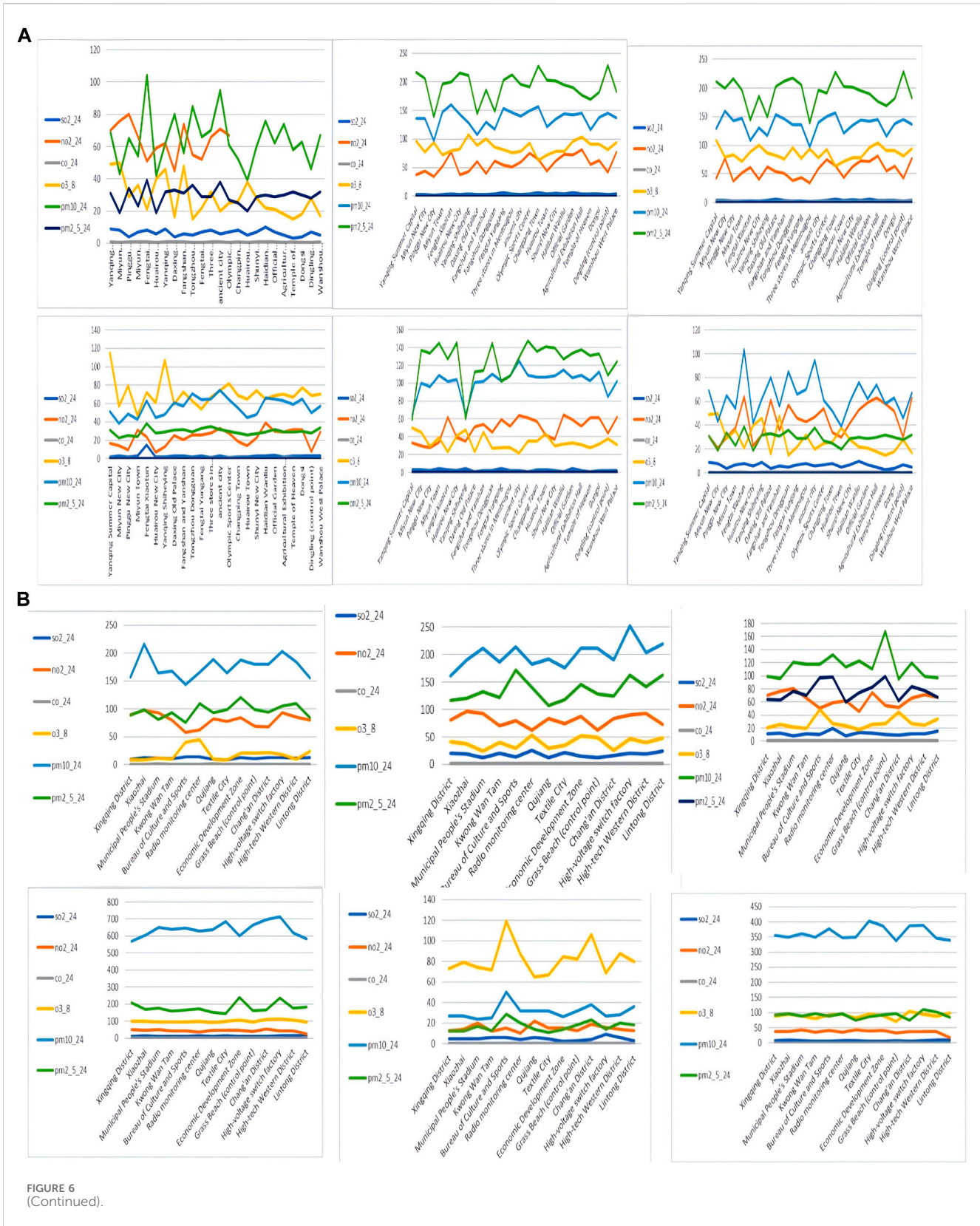
Selecting the value of k : The optimal k value is determined based on cross-validation results to ensure the model's accuracy and generalizability.

Distance Measurement: An improved method of distance measurement is applied to calculate the similarity between data points, aiding in more accurately identifying the nearest neighbors.

2.2.2.3 Synthesis of evidence theory and KNN

The evidence theory fusion results are further corrected using subjective and objective weights to optimize assessment accuracy.

The output of the evidence theory computations is used as input for the KNN algorithm, and the concentrations are estimated by



applying the matter-element extension model. This step accounts for the spatiotemporal dynamics of air pollutants and their environmental and health impacts.

Our model thus enhances the accuracy of air pollution assessment based on multi-source data and effectively fuses the data to improve the comprehensiveness and reliability of the results.

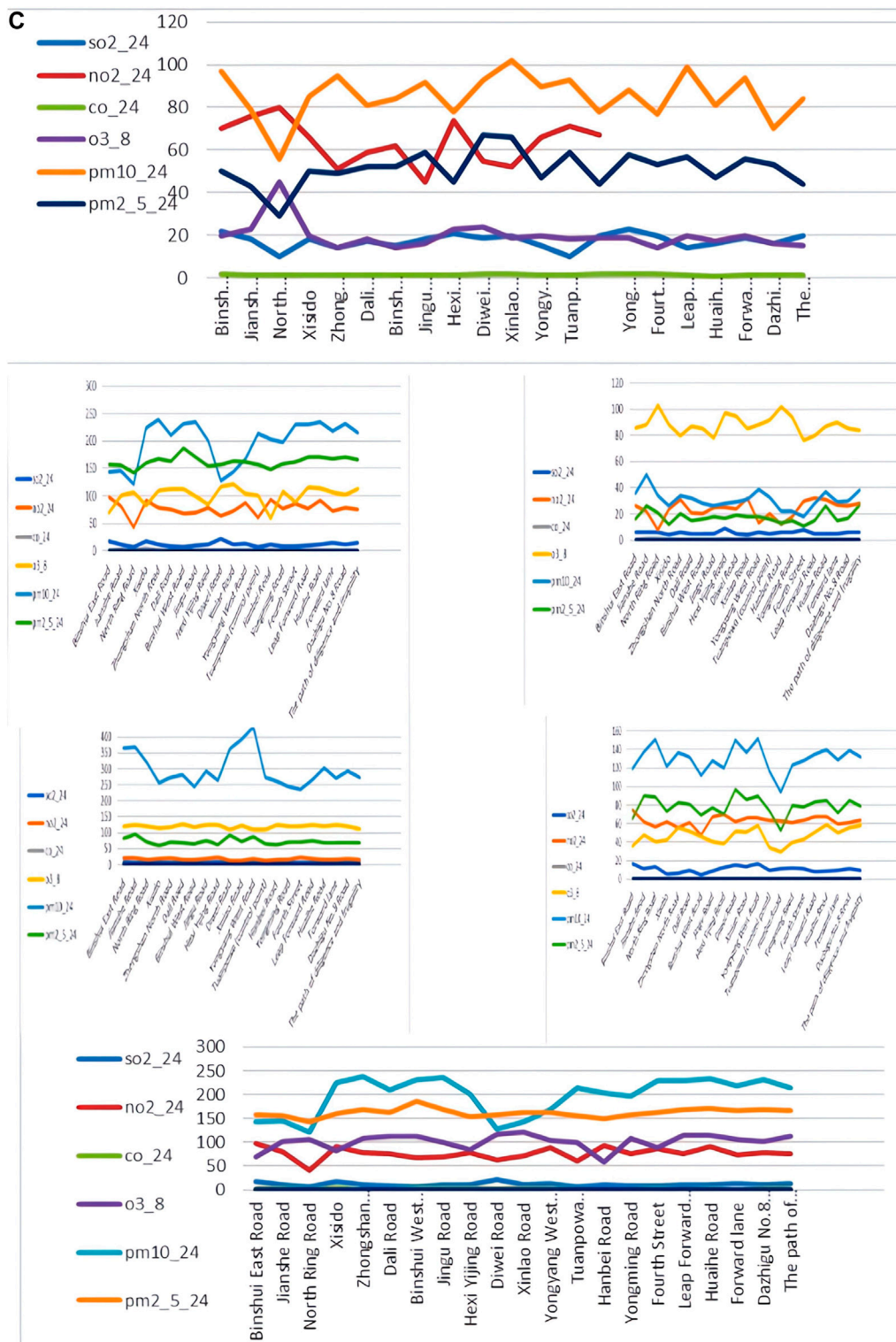


FIGURE 6 (Continued). Comparison of the monitored concentrations of different pollutants in different regions. (A) Xi'an; (B) Beijing; (C) Tianjin.

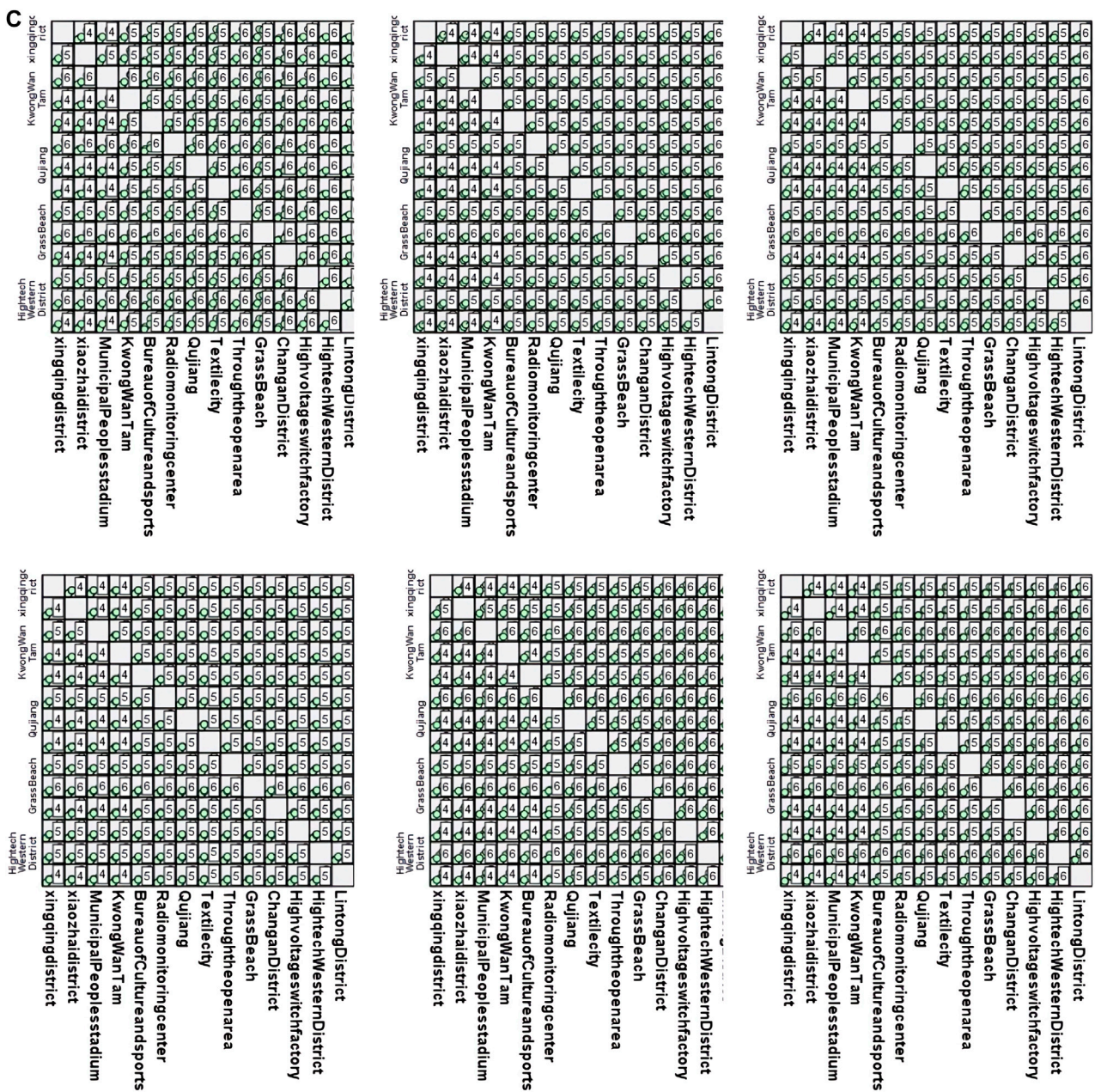


FIGURE 7
(Continued). Correlations among pollutant concentrations in different regions. (A) Xi'an; (B) Beijing; (C) Tianjin.

2.3 Research methods

2.3.1 Fusion of air pollutant concentrations based on interval similarity

The air pollutant concentrations were fused as follows:

- For each urban area within the city, a distinct set (U) of pollutant concentrations was determined:

$$U = \{U_1, U_2 \dots U_N\} \quad (11)$$

Each single factor was then decomposed into subsystem evaluation factors, with the corresponding factor set U_i represented as

$$U_i = \{U_{i1}, U_{i2} \dots U_{iN}\} \quad (12)$$

where U_i is the factor set and U_{iN} is the N th sub-factor indicator in the i th factor set.

- Based on the degree of harm to human health caused by air pollution, the air pollution concentrations were categorized into seven levels, thus determining the evaluation set (Table 2), denoted as

$$V = \{V_1, V_2 \dots V_7\} \quad (13)$$

where V_1 – V_7 correspond to seven levels of air pollution concentration.

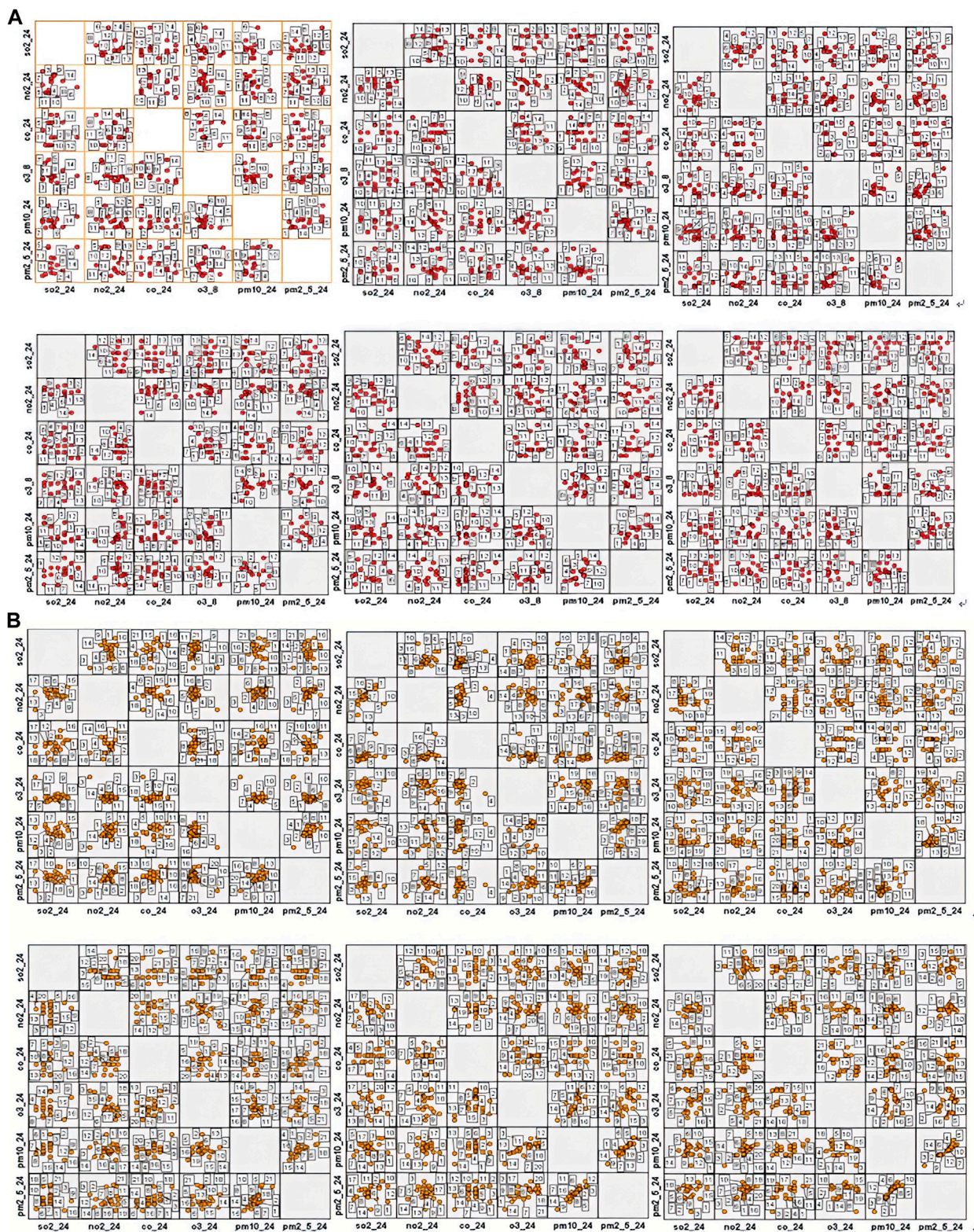


FIGURE 8
(Continued).

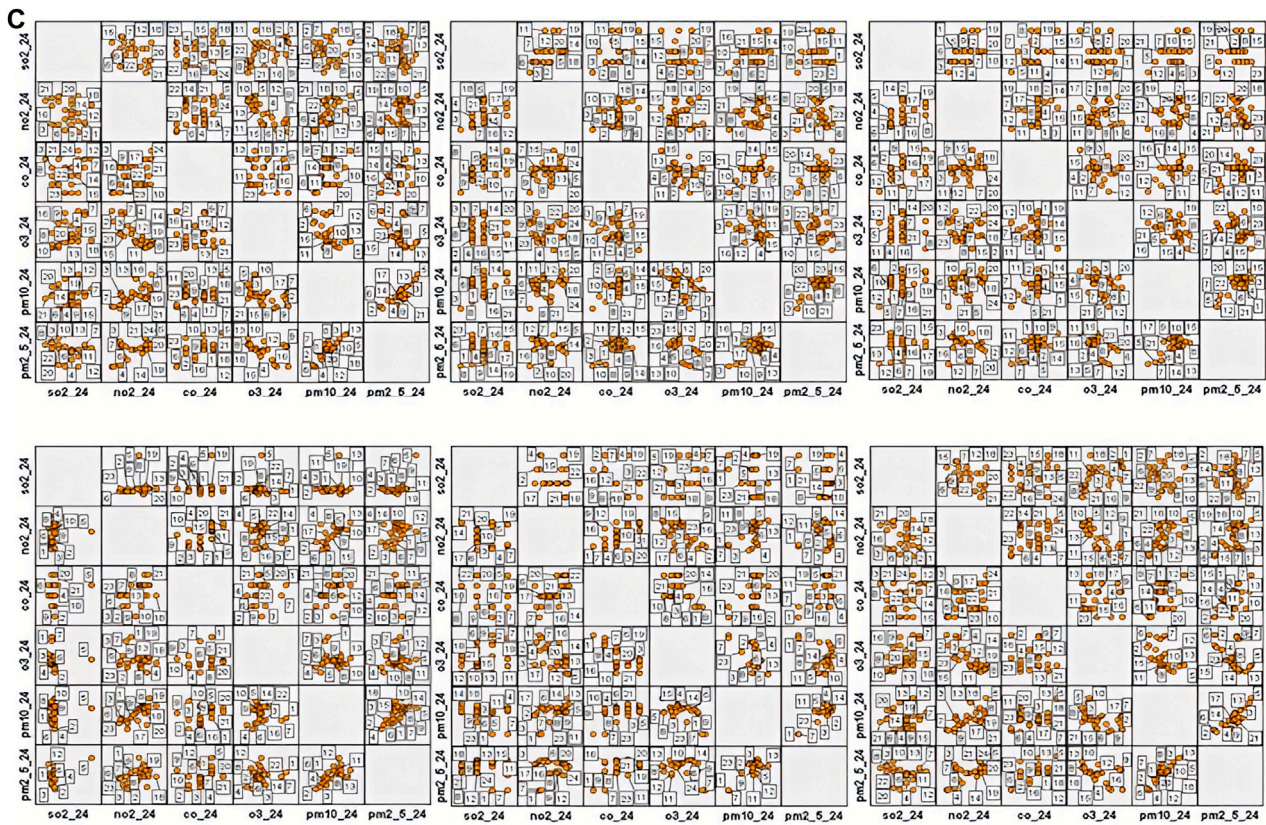


FIGURE 8
Correlations among pollutant concentrations. (A) Xi'an; (B) Beijing; (C) Tianjin.

c) To determine the membership function, based on the air pollution concentration standards of the Chinese Environmental Protection Administration (EPA), the air pollutant concentrations were classified by interval. For each pollutant, the distance to each interval was calculated using Equation (14). A larger distance to the interval corresponds to a lower similarity and a greater impact of pollutants on air pollution concentrations.

$$D(U, V_k) = \frac{\sum_{k=1}^n |U - V_k|}{n} \quad (14)$$

Similarity was then calculated via Equation 15 and was normalized to yield the basic probability distribution function, as follows:

$$S(U, V_k) = \frac{1}{(1 + D(U, V_k))} \quad (15)$$

d) The DS evidence from different regions for each pollutant was then sequentially fused. Following the first law of geography, the law of spatial correlation, which states that “all things are related, but nearby things are more related than distant things” (Tobler, 1970; Li et al., 2016), it is necessary to consider the impact of pollutant concentrations and spatiotemporal influences. During data fusion, data transmission can be delayed. For greater efficiency, we applied sequential data

fusion, fusing the output data items one-by-one, based on their to spatial sequence. Unlike other fusion techniques, sequential data fusion better balances data quality and processing speed, enabling real-time or near-real-time air quality assessments. Figure 4 illustrates BPA-value adjustment using interval similarity via DS evidence theory. This adjustment effectively addresses the inherent uncertainty associated with atmospheric pollution concentrations. Merging the modified BPA values produces the final atmospheric pollution concentrations:

$$U_N = (U_1, U_2, U_3 \dots U_N) \quad (16)$$

2.3.2 Construction of the DS evidence theory model based on comprehensive weighting

To evaluate the air pollution concentrations, concentrations of air pollutants from multiple sources were fused. This analysis requires consideration of individual pollutant concentrations and of the mutual influences among the pollution levels in different regions for each pollutant. Therefore, each indicator has both subjective and objective weights. Subjective (or expert) weighting is derived from expert opinion or experience. Objective weighting is based on the relationships among indicators and relies on specific mathematical methods to calculate the weight of each indicator.

TABLE 10 BPA values across different regions.xi'an.

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Xingqing District	0.1334	0.723,083	0.72041	0.7047	0.68038	0.6795
Xiaozhai	0.1335	0.73219	0.71237	0.7116	0.67454	0.6764
Municipal People's Stadium	0.133	0.738,389	0.71503	0.706	0.72652	0.7176
Kwong Wan Tam	0.1334	0.717,136	0.71771	0.7033	0.719,877	0.6987
Bureau of Culture and Sports	0.1332	0.695,682	0.71771	0.7437	0.719,877	0.7641
Radio monitoring center	0.1345	0.706,962	0.71237	0.713	0.75202	0.7665
Qujiang	0.133	0.711,287	0.71237	0.7088	0.71118	0.6673
Textile City	0.1336	0.687,456	0.72041	0.7006	0.73103	0.7121
Economic Development Zone	0.1335	0.729,129	0.71237	0.7116	0.70269	0.7325
Grass Beach (control point)	0.1332	0.701,277	0.70972	0.713	0.82813	0.7689
Chang'an District	0.1331	0.697,073	0.70972	0.7377	0.67262	0.6733
High-voltage switch factory	0.1334	0.717,136	0.70972	0.713	0.72429	0.7346
High-tech Western District	0.1334	0.724,585	0.71237	0.7102	0.68038	0.7197
Lintong District	0.1338	0.718,614	0.71771	0.7229	0.7265	0.676,476
Xingqing District	0.13472	0.71909	0.71463	0.7074	0.6335	0.671,995
Xiaozhai	0.13508	0.7267	0.72273	0.7102	0.7450	0.681,060
Municipal People's Stadium	0.13484	0.72855	0.72273	0.7189	0.6805	0.68106
Kwong Wan Tam	0.13472	0.72060	0.70,670,944	0.68828	0.654,218	0.69672
Bureau of Culture and Sports	0.13484	0.690,182	0.709,331	0.7524	0.64844	0.68413
Radio monitoring center	0.135,089	0.69437	0.714,633	0.69911	0.8218	0.82859
Qujiang	0.13472	0.71312	0.714,633	0.69365	0.814,504	0.73031
Textile City	0.13447	0.713,121	0.71,197,217	0.6923	0.66209	0.68724
Economic Development Zone	0.13508	0.73041	0.714,633	0.71601	0.7717	0.726,812,618
Grass Beach (control point)	0.134,478	0.708,705	0.71,197,217	0.70187	0.742,507	0.752,074,228
Chang'an District	0.134,844	0.690,182	0.70,670,944	0.73680	0.69557	0.663,168,781
High-voltage switch factory	0.135,212	0.726,707	0.717,313	0.73833	0.75421	0.741,035,957
High-tech Western District	0.135,089	0.722,122	0.72,001,489	0.73987	0.68475	0.723,341,689
Lintong District	0.135,212	0.716,096	0.71,197,217	0.704,659	0.6912	0.732,435,926
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Xingqing District	0.1,347,224	0.719,096	0.714,633	0.707,465	0.633,503	0.67,199,534
Xiaozhai	0.13,508,933	0.726,707	0.722,736	0.71029	0.74502	0.681,060,025
Municipal People's Stadium	0.1,348,445	0.728,558	0.7,227,363	0.718,915	0.680,516	0.681,060,025
Kwong Wan Tam	0.1,347,224	0.720,606	0.70,670,944	0.68828	0.65422	0.696,723,779
Bureau of Culture and Sports	0.1,348,445	0.6,901,826	0.709,331	0.752,404	0.648,435	0.684,136,177
Radio monitoring center	0.1,350,893	0.69437	0.714,633	0.69911	0.821,803	0.828,598,193

(Continued on following page)

TABLE 10 (Continued) BPA values across different regions.xi'an.

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Qujiang	0.1,347,224	0.7,131,217	0.71,463,303	0.69365	0.814,504	0.730,317,018
Textile City	0.1,344,789	0.7,131,217	0.71,197,217	0.6923	0.66209	0.687,240,243
Economic Development Zone	0.1,350,893	0.730,418	0.714,633	0.716,018	0.771,665	0.726,812,618
Grass Beach (control point)	0.1,344,789	0.708,705	0.71,197,217	0.70187	0.742,507	0.752,074,228
Chang'an District	0.1,348,445	0.690,182	0.70,670,944	0.7368	0.695,572	0.663,168,781
High-voltage switch factory	0.135,212	0.726,707	0.7,173,138	0.73833	0.754,212	0.741,035,957
High-tech Western District	0.13,508,933	0.722,122	0.72,001,489	0.73987	0.684,751	0.723,341,689
Lintong District	0.13,521,208	0.7,160,968	0.71,197,217	0.704,659	0.6,912,027	0.732,435,926

2.3.3 Subjective weights

Each air pollutant has a specific toxicity level, with higher toxicity correlating with increased pollution and potential harm to humans. To ensure a well-rounded and credible representation of pollutant toxicity, we derived subjective weights (denoted by W_i) for each pollutant, by consulting a panel of experts and corroborating their input using the literature. This approach both leverages expert experience and ensures a balanced weighting strategy supported by established knowledge.

2.3.4 Objective weights

Objective weighting has the advantage of relying primarily on objective data without requiring human intervention. Therefore, objective judgments are not limited by human subjectivity and are considered true and reliable.

Air pollution levels depend both on differences between pollutants and regional pollutant concentrations. Therefore, to calculate the objective weights, we implemented the CRITIC Importance Through Intercriteria Correlation (CRITIC) method (Wang and Jiang, 2017; Ying and Yunyun, 2017), a sophisticated approach that effectively integrates both contrasting and conflicting indicators. In this context, associations between information are used to quantify the conflict between indicators. More closely associated indicators exhibit less conflict, and are therefore assigned less weight. The contrast between indicators was ascertained using the standard deviation of the information; a larger standard deviation implies a greater contrast. Here, we adopted the "Over-standard Multiple Method" to represent this contrast, using the standard deviation of information. Indicator similarity, conversely, was used to compute the conflict between indicators. Among the various weight-calculation methods, CRITIC was deemed the most suitable for our research objectives; the other methods, in contrast, exhibit greater computational complexity, specific data prerequisites, or inherent limitations. The objective weights were calculated as follows.

a) According to Chinese EPA monitoring data, the regions $P_1, P_2, P_3, \dots, P_j$ are currently monitored for the pollutants SO₂, NO₂, CO, O₃, PM_{2.5}, and PM₁₀. The daily air pollutant concentrations in the different regions are as follows:

$$U = \begin{Bmatrix} U_{11} & U_{12} & U_{13} & \dots & U_{1N} \\ U_{21} & U_{22} & U_{23} & \dots & U_{2N} \\ U_{31} & U_{32} & U_{33} & \dots & U_{3N} \\ U_{41} & U_{42} & U_{43} & \dots & U_{4N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_{i1} & U_{i2} & U_{i3} & \dots & U_{iN} \end{Bmatrix} \quad (17)$$

where N is the number of air pollutants (SO₂, NO₂, CO, O₃, PM_{2.5}, and PM₁₀) and i is the monitoring region.

b) To better demonstrate pollution levels, the pollutant concentration indicators for the different regions and pollutants in the matrix were nondimensionalized. The concentrations of the same pollutant in different regions mutually affect one another. Therefore, range normalization was used for nondimensionalization:

$$U_{iN}' = \frac{U_{iN} - \min}{\max - \min} \quad (18)$$

After nondimensionalization, the relative pollution levels for the same pollutant in different regions are obtained, as follows:

$$U = \begin{Bmatrix} U_{11} & U_{12} & U_{13} & \dots & U_{1N} \\ U_{21} & U_{22} & U_{23} & \dots & U_{2N} \\ U_{31} & U_{32} & U_{33} & \dots & U_{3N} \\ U_{41} & U_{42} & U_{43} & \dots & U_{4N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_{i1} & U_{i2} & U_{i3} & \dots & U_{iN} \end{Bmatrix} \quad (19)$$

where N is the number of air pollutants (SO₂, NO₂, CO, O₃, PM_{2.5}, and PM₁₀) and i is the monitoring region.

c) The concentrations of different pollutants in the same region can also mutually affect one another. When multiple different pollutants are present in greater concentrations, their impact on one another is greater. The contrast among indicators in the CRITIC method can be represented by over-standard multiples, as follows:

TABLE 11 BPA values across different regions.beijing

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Yanqing Summer Capital	0.41,729,175	0.41	0.41737	0.414,353	0.408,899	0.408,415
Miyun New City	0.41,691,635	0.42	0.41893	0.414,353	0.415,314	0.422,296
Pinggu New City	0.41,691,635	0.42	0.415,822	0.418,275	0.413,154	0.413,356
Miyun Town	0.41,654,162	0.41	0.414,285	0.426,345	0.405,764	0.39772
Fengtai Xiaotun	0.41,654,162	0.41	0.415,822	0.415,131	0.406,804	0.415,871
Huairou New City	0.41,654,162	0.42	0.414,285	0.420,663	0.413,154	0.419,702
Yanqing Shiheyang	0.41,691,635	0.42	0.41737	0.414,353	0.417,498	0.42758
Daxing Old Palace	0.41,691,635	0.42	0.41737	0.415,131	0.419,705	0.413,356
Fangshan and Yanshan	0.41,654,162	0.42	0.41893	0.415,913	0.42993	0.415,871
Tongzhou Dongguan	0.41,654,162	0.42	0.415,822	0.427,169	0.412,081	0.400,048
Fengtai Yungang	0.41,691,635	0.41	0.414,285	0.415,913	0.412,081	0.408,415
Three stores in Mentougou	0.41,654,162	0.4	0.414,285	0.423,891	0.407,849	0.405,989
ancient city	0.41,616,756	0.43	0.41737	0.412,805	0.425,325	0.423,604
Olympic Sports Center	0.41,616,756	0.42	0.41737	0.416,697	0.425,325	0.423,604
Changping Town	0.41,654,162	0.41	0.41737	0.413,577	0.427,615	0.420,995
Huairou Town	0.41,691,635	0.42	0.414,285	0.416,697	0.412,081	0.41461
Shunyi New City	0.41,691,635	0.42	0.415,822	0.415,131	0.411,015	0.410,871
Haidian Wanliu	0.41,691,635	0.42	0.41737	0.413,577	0.42993	0.420,995
Official Garden	0.41,654,162	0.43	0.420,501	0.415,131	0.423,059	0.423,604
Agricultural Exhibition Hall	0.41,654,162	0.42	0.414,285	0.415,913	0.423,059	0.428,922
Temple of Heaven	0.41,616,756	0.41	0.41893	0.413,577	0.407,849	0.423,604
Dongsi	0.41,691,635	0.42	0.41893	0.415,913	0.418,598	0.422,296
Dingling (control point)	0.41,616,756	0.42	0.415,822	0.415,913	0.415,314	0.413,356
Wanshou West Palace	0.41,691,635	0.42	0.41737	0.413,577	0.418,598	0.424,921
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
Yanqing Summer Capital	0.41,729,175	0.41	0.41737	0.414,353	0.408,899	0.408,415
Miyun New City	0.41,691,635	0.42	0.41893	0.414,353	0.415,314	0.422,296
Pinggu New City	0.41,691,635	0.42	0.415,822	0.418,275	0.413,154	0.413,356
Miyun Town	0.41,654,162	0.41	0.414,285	0.426,345	0.405,764	0.39772
Fengtai Xiaotun	0.41,654,162	0.41	0.415,822	0.415,131	0.406,804	0.415,871
Huairou New City	0.41,654,162	0.42	0.414,285	0.420,663	0.413,154	0.419,702
Yanqing Shiheyang	0.41,691,635	0.42	0.41737	0.414,353	0.417,498	0.42758
Daxing Old Palace	0.41,691,635	0.42	0.41737	0.415,131	0.419,705	0.413,356
Fangshan and Yanshan	0.41,654,162	0.42	0.41893	0.415,913	0.42993	0.415,871
Tongzhou Dongguan	0.41,654,162	0.42	0.415,822	0.427,169	0.412,081	0.400,048

(Continued on following page)

TABLE 11 (Continued) BPA values across different regions.beijing

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Fengtai Yungang	0.41,691,635	0.41	0.414,285	0.415,913	0.412,081	0.408,415
Three stores in Mentougou	0.41,654,162	0.4	0.414,285	0.423,891	0.407,849	0.405,989
ancient city	0.41,616,756	0.43	0.41737	0.412,805	0.425,325	0.423,604
Olympic Sports Center	0.41,616,756	0.42	0.41737	0.416,697	0.425,325	0.423,604
Changping Town	0.41,654,162	0.41	0.41737	0.413,577	0.427,615	0.420,995
Huairou Town	0.41,691,635	0.42	0.414,285	0.416,697	0.412,081	0.41461
Shunyi New City	0.41,691,635	0.42	0.415,822	0.415,131	0.411,015	0.410,871
Haidian Wanliu	0.41,691,635	0.42	0.41737	0.413,577	0.42993	0.420,995
Official Garden	0.41,654,162	0.43	0.420,501	0.415,131	0.423,059	0.423,604
Agricultural Exhibition Hall	0.41,654,162	0.42	0.414,285	0.415,913	0.423,059	0.428,922
Temple of Heaven	0.41,616,756	0.41	0.41893	0.413,577	0.407,849	0.423,604
Dongsi	0.41,691,635	0.42	0.41893	0.415,913	0.418,598	0.422,296
Dingling (control point)	0.41,616,756	0.42	0.415,822	0.415,913	0.415,314	0.413,356
Wanshou West Palace	0.41,691,635	0.42	0.41737	0.413,577	0.418,598	0.424,921

$$S_j = U_{iN} / \left(\frac{1}{n} \sum_{j=1}^n V_{ij} \right) \quad (20)$$

d) For the same pollutant, the levels are strongly positively correlated among regions. The extent of correlation can be determined from the similarity among indicators. Greater similarity corresponds to less conflict. Here, the indicator similarity matrix was constructed using Pearson correlation coefficients. The level of conflict (q_{ij}) of each indicator in the similarity matrix was calculated as follows:

$$Q_j = 1 - |q_{ij}| \quad (21)$$

e) F_j , the degree of pollution caused by the j th pollutant, was calculated as

$$F_j = Q_j * S_j \quad (22)$$

f) W_j , the objective weight of the j th pollutant, was calculated as

$$W_j = \frac{F_j}{\sum_{j=1}^n F_j} \quad (23)$$

2.3.5 Combined subjective and objective weights

The combined subjective and objective weights of the j th indicator were determined using the normalized combined weights, as follows:

$$W_j^* = w_i w_j / \sum_{j=1}^N w_i w_j \text{ where } j = 1, \dots, N \quad (24)$$

W_i and W_j are, respectively, the subjective and objective weight elements of the n th indicator.

2.3.6 Subjective and objective correction

Subjective and objective correction of the concentrations of different pollutants in the same region was performed as follows:

$$D = W_j^* * U_N \quad (25)$$

2.4 Construction of the corrected KNN model based on the average extensible distance

The KNN algorithm determines the class of a sample by counting the number of nearest neighbors of the same class. However, if the distribution of classified samples is highly scattered and the data distribution is not considered, the results can easily be biased. Yang et al. (2010) proposed using the extensible distance to vote and select the target attribute of the dataset for classification. Tan et al. (2017) developed an air quality evaluation model based on fuzzy matter-element analysis. Xiao and Duan (2013) proposed improving the importance of classes using attribute values. Dai et al. (2018) introduced a method to calculate sample-attribute weights using the analytic hierarchy process and to classify samples based on their weighted distances. Lü et al. (2021) proposed correcting KNN classification using the probability that the test sample belongs to each fault type as the weight.

Pollutant concentration data is typically log-normally distributed, as substantiated by Ott (1990), who provided the physical explanation for this phenomenon. This log-normal distribution is further supported by the method-fusion approach for enhancing long-term air pollution estimates, as discussed by Chastko and Adams (2019). Air pollution concentration data follow a normal distribution (Figure 5). Each

TABLE 12 BPA values across different regions.tianjin

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
Xingqing District	0.385,551	0.374,772	0.386,021	0.399,306	0.392,035	0.390,947
Xiaozhai	0.385,203	0.365,038	0.381,788	0.400,129	0.364,171	0.366,603
Municipal People’s Stadium	0.383,815	0.373,775	0.386,021	0.383,532	0.387,913	0.397,547
Kwong Wan Tam	0.384,855	0.381,901	0.3846	0.388,909	0.377,011	0.374,374
Bureau of Culture and Sports	0.385,203	0.402,857	0.388,896	0.377,566	0.432,226	0.406,366
Radio monitoring center	0.384,508	0.366,944	0.381,788	0.39205	0.362,917	0.366,603
Qujiang	0.385,551	0.375,774	0.387,453	0.396,858	0.384,878	0.393,122
Textile City	0.383,815	0.400,398	0.3846	0.373,931	0.403,837	0.395,322
Economic Development Zone	0.384,508	0.379,837	0.386,021	0.398,487	0.378,947	0.390,947
Grass Beach (control point)	0.384,161	0.397,165	0.383,189	0.373,212	0.40944	0.401,419
Chang’an District	0.384,855	0.388,537	0.383,189	0.378,302	0.388,935	0.386,668
High-voltage switch factory	0.385,203	0.386,249	0.387,453	0.385,818	0.39308	0.386,668
High-tech Western District	0.384,508	0.389,306	0.386,021	0.376,833	0.421,125	0.404,704
Lintong District	0.384,855	0.394,774	0.383,189	0.380,526	0.383,876	0.382,481
Xingqing District	0.385,203	0.377,794	0.3846	0.381,273	0.375,094	0.378,384
Xiaozhai	0.384,161	0.373,775	0.383,189	0.390,473	0.359,207	0.368,516
Municipal People’s Stadium	0.384,855	0.385,493	0.386,021	0.382,776	0.38288	0.386,668
Kwong Wan Tam	0.3859	0.393,984	0.387,453	0.378,302	0.399,464	0.388,796
Bureau of Culture and Sports	0.384,855	0.398,775	0.386,021	0.377,566	0.384,878	0.386,668
Radio monitoring center	0.384,161	0.402,034	0.383,189	0.375,376	0.397,313	0.388,796
Qujiang	0.38347	0.397,968	0.387,453	0.373,212	0.380,904	0.393,122
:	:	:	:	:	:	:
:	:	:	:	:	:	:
:	:	:	:	:	:	:
BinshuiEastRoad	0.477,904	0.47468	0.479,314	0.477,222	0.480,541,073	0.512,594,485
JiansheRoad	0.478,341	0.478,999	0.4825	0.478,138	0.481,408,476	0.507,314,542
NorthRingRoad	0.476,164	0.475,906	0.47204	0.474,497	0.471,201,936	0.443,376,925
Xisido	0.475,299	0.479,622	0.479,314	0.467,378	0.462,218,883	0.469,810,367
ZhongshanNorthRoad	0.474,437	0.476,521	0.483,841	0.470,023	0.463,826,601	0.528,169,295
DaliRoad	0.474,867	0.475,906	0.47204	0.483,705	0.490,257,897	0.470,933,419
BinshuiWestRoad	0.473,149	0.471,642	0.473,837	0.483,705	0.491,160,766	0.47,547,984
JinguRoad	0.478,341	0.479,622	0.479,314	0.479,979	0.490,257,897	0.503,425,418
HexiYijingRoad	0.476,598	0.475,906	0.475,649	0.477,222	0.476,250,528	0.441,390,173
DiweiRoad	0.478,341	0.475,906	0.477,474	0.472,697	0.470,370,892	0.436,500,326
XinlaoRoad	0.474,867	0.465,094	0.475,649	0.470,911	0.464,634,662	0.436,500,326
YongyangWestRoad	0.476,164	0.444,523	0.470,256	0.482,768	0.479,676,791	0.352,427,812
Tuanpowa	0.475,731	0.477,138	0.47204	0.47631	0.475,401,597	0.509,940,846

(Continued on following page)

TABLE 12 (Continued) BPA values across different regions.tianjin

Region	Pollutant concentration					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
HanbeiRoad	0.474,437	0.468,048	0.470,256	0.473,595	0.470,370,892	0.436,500,326
YongmingRoad	0.480,097	0.476,521	0.473,837	0.482,768	0.48,846,208	0.480,114,899
FourthStreet	0.477,468	0.483,397	0.477,474	0.470,023	0.473,712,781	0.486,037,369
LeapForwardRoad	0.476,164	0.485,307	0.479,314	0.477,222	0.475,401,597	0.504,284,506
HuaiheRoad	0.475,299	0.487,232	0.477,474	0.470,911	0.464,634,662	0.507,314,542
Forwardlane	0.475,299	0.47468	0.47204	0.482,768	0.484,909,629	0.480,114,899
DazhiguNo8Road	0.475,731	0.480,247	0.473,837	0.478,138	0.482,279,016	0.504,284,506
Thepathofdiligenceandfrugality	0.475,299	0.4931	0.4825	0.470,023	0.463,021,347	0.513,485,179

TABLE 13 Sequential fusion results.xi'an.

Method	Pollutant					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
D-S Sequential Fusion Results1	9.21366E-13	0.01426	0.01430	0.01428	0.01401	0.01408
		7,936	647	124	5,764	2,376
D-S Sequential Fusion Results2	5.8882	0.01415	0.0141	0.01415	0.01394	0.01416
	6E-13	5,479	85,013	2,786	6,904	668
D-S Sequential Fusion Results2	1.001	0.02054	0.02057	0.0205	0.03142	0.02046
	67E-12	5,705	6,677	60,332	1758	5,258
D-S Sequential Fusion Results2	4.3349	0.01447	0.01450	0.0144	0.0184	0.0144
	5E-13	6,865	1,514	7,182	79,643	17,647
D-S Sequential Fusion Results2	6.015	0.01407	0.01416	0.0137	0.01102	0.01296
	14E-13	9,683	2,273	7,331	1,291	1,657
D-S Sequential Fusion Results2	6.2820	0.01423	0.01427	0.01422	0.01410	0.0141
	2E-13	8,819	6,197	9,081	1,476	6,153

TABLE 14 Sequential fusion results.beijing

Method	Pollutant					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
D-S Sequential Fusion Results1	1.14978E-10	1.35918E-10	1.25262E-10	1.1126E-10	1.44329E-10	1.4529E-10
D-S Sequential Fusion Results2	7.94276E-10	8.48062E-10	8.56334E-10	8.5093E-10	8.35209E-10	8.5155E-10
D-S Sequential Fusion Results2	7.92718E-10	8.45955E-10	8.52807E-10	8.5156E-10	8.36499E-10	8.4969E-10
D-S Sequential Fusion Results2	7.95042E-10	8.52818E-10	8.57729E-10	8.4851E-10	8.50694E-10	8.5444E-10
D-S Sequential Fusion Results2	7.94844E-10	8.5099E-10	8.57506E-10	8.5514E-10	8.43495E-10	8.2224E-10
D-S Sequential Fusion Results2	7.9371	8.5339	8.54802	8.5412	8.5186	8.510
	2E-10	5E-10	E-10	E-10	6E-10	3E-10

TABLE 15 Sequential fusion results.tianjin

Method	Pollutant					
	SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
D-S Sequential Fusion Results1	1.8865E-07	2.07734E-07	2.08179E-07	2.0799E-07	2.06534E-07	2.0539E-07
D-S Sequential Fusion Results2	1.88537E-07	2.06473E-07	2.06968E-07	2.0539E-07	1.90604E-07	2.074E-07
D-S Sequential Fusion Results2	1.89902E-07	2.10487E-07	2.10577E-07	2.1044E-07	2.08663E-07	2.0787E-07
D-S Sequential Fusion Results2	1.89337E-07	2.08476E-07	2.09041E-07	2.0859E-07	2.07962E-07	2.0802E-07
D-S Sequential Fusion Results2	1.89327E-07	2.08873E-07	2.09097E-07	2.0851E-07	2.05306E-07	2.0632E-07
D-S Sequential Fusion Results2	1.89425E-07	2.08922E-07	2.09619E-07	2.0953E-07	1.93548E-07	2.0614E-07

pollutant concentration has a standard interval, according to the Chinese EPA standards. Given these standard intervals, we can determine the air pollution concentrations via DS evidence theory fusion and corrected via subjective and objective weighting. Here, we propose a model algorithm based on KNN corrected using matter-element average extensible distances (Ott, 1990. A Physical Explanation of the Lognormality of Pollutant).

The matter-element extension theory (Xiang, 2008) considers the objective world as comprising matter elements; contradictions in the objective world are treated as contradictions among matter elements. Matter elements have many types of features, each with a value. Here, the matter elements (the evaluation subjects, denoted as U) and their features (the evaluation indicators, denoted as T) are combined with the values Y (which correspond to the features), forming R = (U, T, Y), the three matter-elements. The model calculations are as follows.

- (1) If the described matter U has *n* features T₁, T₂, . . . , T_m with corresponding values Y₁, Y₂, . . . , Y_m, then R is called an *n*-dimensional fuzzy matter-element (Zhang, 1997). Combining *n*-dimensional matter-elements of *m* matters forms *n*-dimensional composite fuzzy matter-elements R_{nm} of *m* matters:

$$R = \begin{bmatrix} U & T_1 & Y_1 \\ & T_2 & Y_2 \\ & \vdots & \vdots \\ & T_n & Y_n \end{bmatrix} \tag{26}$$

$$R = \begin{bmatrix} T_1 & T_2 & \cdots & T_n \\ U_1 & Y_{11} & Y_{12} & \cdots & Y_{1n} \\ U_2 & Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_m & Y_{m1} & Y_{m2} & \cdots & Y_{mn} \end{bmatrix} \tag{27}$$

- (2) The described matter U has a value of Y for each feature. Let us assume that Y is within the range (a, b) belonging to a classified sample; for X (X₁, I.....X_i), the samples to be classified, the aim is to determine the feature interval to which they belong, using the KNN algorithm. Therefore, we set K = 3 and use the quartile method to divide the interval Y of each known classified sample into three adjacent intervals.
- (3) The average distance between the sample to be classified and the three nearest neighbor intervals of different classes is calculated using the interpoint distance:

$$d(x_i, y_n) = \left| x - \frac{(a_m + b_m)}{2} \right| - \frac{(b_m - a_m)}{2} \tag{28}$$

- (4) The average distance between the sample to be classified and the nearest neighbor points of each class is calculated as follow:

$$\bar{d}(x_i, y_n) = \frac{1}{3} \sum \left(\left| x - \frac{(a_m + b_m)}{2} \right| - \frac{(b_m - a_m)}{2} \right) \tag{29}$$

- (5) The average distance between the sample to be classified and each classified sample point is then compared. A smaller distance indicates that the sample to be classified more likely belongs to the corresponding interval.

3 Results

3.1 Evaluation indicator set

In evaluating air pollution, pollutant concentrations were calculated based on regional data for each city. It is assumed that each city has *i* monitoring regions, each with *N* indicators *u*₁, *u*₂ . . . *u*_N, which form the set *u* = {*u*₁, *u*₂ . . . *u*_N}. Each indicator set can be decomposed into subsystem evaluation factors, and the corresponding factor set *U*_{*i*} is represented as *u*_{*i*} = {*u*_{*i1*}, *u*_{*i2*} . . . *u*_{*iN*}}, i.e., the evaluation factor set for each region. Based on the Chinese EPA’s air pollution concentration standards, an evaluation set {I, II, III, IV, V, VI, VII} is defined, where a higher level indicates a higher pollutant concentration and a greater impact on health (Table 2). The data are then nondimensionalized (Table 3).

3.2 Data source

The air quality data used here—the daily air quality data for 2022 for Xi’an, Beijing, and Tianjin—were obtained primarily from the China National Environmental Monitoring Network. Owing to space constraints, only a subset of the data is presented (Tables 4–6).

Additionally, we based our in-depth statistical analyses on the processed and analyzed datasets of the Qingyue Open Environmental Data Center. We standardized the concentrations of the various air pollutants, obtaining

TABLE 16 Subjective and objective weights of the pollutants.xi'an.

Pollutant	Indicator item				
	Indicator variability	Indicator conflict	Objective weight (%)	Subjective weight	Weight
SO _{2_24}	3.082	4.687	5.11	2	0.2511
NO _{2_24}	10.324	6.23	22.75	2	0.6275
CO ₂₄	0.144	5.61	0.29	1	0.2029
O _{3_8}	8.666	5.632	17.26	3	0.7726
PM _{10_24}	19.343	4.972	34.02	4	1.1402
PM _{2.5_24}	13.719	4.239	20.57	5	1.2057
SO _{2_24}	4.312	4.64	7.35	2	0.2735
NO _{2_24}	11.006	5.212	21.09	2	0.6109
CO ₂₄	0.15	3.612	0.20	1	0.202
O _{3_8}	9.725	4.344	15.53	3	0.7553
PM _{10_24}	22.635	3.7	30.78	4	1.1078
PM _{2.5_24}	19.773	3.447	25.05	5	1.2505
SO _{2_24}	1.598	4.488	1.77	2	0.2177
NO _{2_24}	6.796	4.858	8.15	2	0.4815
CO ₂₄	0.107	6.278	0.17	1	0.2017
O _{3_8}	6.188	3.931	6.01	3	0.6601
PM _{10_24}	42.135	4.62	48.06	4	1.2806
PM _{2.5_24}	28.384	5.114	35.84	5	1.3584
SO _{2_24}	1.626	6.091	4.02	2	0.2402
NO _{2_24}	6.627	4.852	13.06	2	0.5306
CO ₂₄	0.141	5.893	0.34	1	0.2034
O _{3_8}	21.658	4.886	42.97	3	1.0297
PM _{10_24}	8.483	5.672	19.54	4	0.9954
PM _{2.5_24}	9.913	4.988	20.08	5	1.2008
SO _{2_24}	1.73	4.985	7.68	2	0.2768
NO _{2_24}	0.109	3.662	0.35	2	0.4035
CO ₂₄	15.079	3.503	47.01	1	0.6701
O _{3_8}	6.826	2.954	17.94	3	0.7794
PM _{10_24}	3.269	4.439	12.91	4	0.9291
PM _{2.5_24}	5.026	3.154	14.11	5	1.1411
SO _{2_24}	1.834	4.694	3.16	2	0.2316
NO _{2_24}	11.926	4.775	20.89	2	0.6089
CO ₂₄	0.156	5.336	0.30	1	0.203
O _{3_8}	11.825	5.931	25.72	3	0.8572
PM _{10_24}	19.685	4.078	29.45	4	1.0945
PM _{2.5_24}	12.065	4.628	20.48	5	1.2048

TABLE 17 Subjective and objective weights of the pollutants.beijing

Pollutant	Indicator item				
	Indicator variability	Indicator conflict	Objective weight (%)	Subjective weight	Weight
SO _{2_24}	1.817	4.749	3.94	2	0.4394
NO _{2_24}	13.721	4.477	28.05	2	0.6805
CO ₂₄	0.149	3.424	0.23	1	0.2023
O _{3_8}	11.461	6.522	34.14	3	0.9414
PM _{10_24}	15.91	3.437	24.98	4	1.0498
PM _{2.5_24}	5.29	3.583	8.66	5	1.0866
SO _{2_24}	1.035	3.476	1.34	2	0.4134
NO _{2_24}	14.387	3.851	20.58	2	0.6058
CO ₂₄	0.146	3.082	0.17	1	0.2017
O _{3_8}	11.098	6.171	25.44	3	0.8544
PM _{10_24}	15.413	3.027	17.33	4	0.9733
PM _{2.5_24}	23.961	3.95	35.15	5	1.3515
SO _{2_24}	0.565	5.204	2.27	2	0.4227
NO _{2_24}	10.23	3.515	27.74	2	0.6774
CO ₂₄	0.098	3.647	0.27	1	0.2027
O _{3_8}	5.291	6.808	27.78	3	0.8778
PM _{10_24}	12.656	3.336	32.56	4	1.1256
PM _{2.5_24}	3.651	3.329	9.37	5	1.0937
SO _{2_24}	2.645	4.194	7.03	2	0.4703
NO _{2_24}	8.704	3.614	19.94	2	0.5994
CO ₂₄	0.096	3.525	0.22	1	0.2022
O _{3_8}	14.689	5.058	47.11	3	1.0711
PM _{10_24}	9.653	3.142	19.23	4	0.9923
PM _{2.5_24}	3.472	2.938	6.47	5	1.0647
SO _{2_24}	0.97	5.749	2.06	2	0.4206
NO _{2_24}	12.105	4.528	20.26	2	0.6026
CO ₂₄	0.114	4.229	0.18	1	0.2018
O _{3_8}	8.016	6.858	20.32	3	0.8032
PM _{10_24}	13.578	4.058	20.36	4	1.0036
PM _{2.5_24}	23.66	4.211	36.83	5	1.3683
SO _{2_24}	0.816	5.535	3.75	2	0.4375
NO _{2_24}	6.283	3.921	20.46	2	0.6046
CO ₂₄	0.118	4.21	0.41	1	0.2041
O _{3_8}	4.934	7.215	29.57	3	0.8957
PM _{10_24}	6.778	3.895	21.93	4	1.0193
PM _{2.5_24}	6.653	4.322	23.88	5	1.2388

TABLE 18 Subjective and objective weights of the pollutants.tianjin

Pollutant	Indicator item				
	Indicator variability	Indicator conflict	Objective weight (%)	Subjective weight	Weight
SO _{2_24}	3.514	3.983	8.81	2	0.4881
NO _{2_24}	8.205	3.637	18.78	2	0.5878
CO ₂₄	0.227	3.529	0.50	1	0.205
O _{3_8}	6.49	7.159	29.24	3	0.8924
PM _{10_24}	10.755	3.635	24.60	4	1.046
PM _{2.5_24}	8.5	3.377	18.07	5	1.1807
SO _{2_24}	3.723	4.747	4.79	2	0.4479
NO _{2_24}	12.772	4.37	15.13	2	0.5513
CO ₂₄	0.523	4.241	0.60	1	0.206
O _{3_8}	16.067	5.435	23.67	3	0.8367
PM _{10_24}	39.322	4.331	46.16	4	1.2616
PM _{2.5_24}	9.25	3.851	9.65	5	1.0965
SO _{2_24}	1.983	4.881	2.56	2	0.4256
NO _{2_24}	2.809	4.754	3.53	2	0.4353
CO ₂₄	0.086	5.276	0.12	1	0.2012
O _{3_8}	5.114	4.755	6.43	3	0.6643
PM _{10_24}	54.359	5.187	74.55	4	1.5455
PM _{2.5_24}	9.805	4.938	12.80	5	1.128
SO _{2_24}	1.155	5.006	4.58	2	0.4458
NO _{2_24}	6.69	5.681	30.12	2	0.7012
CO ₂₄	0.15	3.692	0.44	1	0.2044
O _{3_8}	7.169	5.49	31.19	3	0.9119
PM _{10_24}	6.982	3.816	21.12	4	1.0112
PM _{2.5_24}	4.363	3.629	12.55	5	1.1255
SO _{2_24}	3.232	3.712	8.90	2	0.489
NO _{2_24}	5.579	4.555	18.85	2	0.5885
CO ₂₄	0.126	3.647	0.34	1	0.2034
O _{3_8}	8.549	3.675	23.31	3	0.8331
PM _{10_24}	13.842	2.615	26.84	4	1.0684
PM _{2.5_24}	10.002	2.933	21.76	5	1.2176
SO _{2_24}	3.866	4.411	5.72	2	0.4572
NO _{2_24}	13.185	3.198	14.15	2	0.5415
CO ₂₄	0.235	3.566	0.28	1	0.2028
O _{3_8}	5.698	6.026	11.52	3	0.7152
PM _{10_24}	42.799	2.861	41.07	4	1.2107
PM _{2.5_24}	29.539	2.752	27.27	5	1.2727

TABLE 19 Corrected D-S fusion results.xi'an.

Pollutant					
SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
4.15628E-13	0.00895,313	0.002,902,783	0.011,033,686	0.015,980,774	0.016,979,121
2.78809E-13	0.008,647,582	0.002,865,373	0.010,689,599	0.01,545,038	0.017,715,433
4.18398E-13	0.009,892,757	0.004,150,316	0.013,571,875	0.040,238,703	0.027,800,006
1.90824E-13	0.007,681,425	0.002,949,608	0.014,901,633	0.018,394,637	0.017,312,711
2.87E-13	5.68E-03	9.49E-03	1.07E-02	1.02E-02	1.48E-02
2.71132E-13	0.008,670,017	0.002,898,068	0.012,197,168	0.015,434,065	0.017,061,811

TABLE 20 Corrected D-S fusion results.beijing

Pollutant					
SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
3.28354E-10	5.13756E-10	1.72723E-10	7.27037E-10	8.12909E-10	1.15087E-09
5.05213E-11	9.2492E-11	2.53405E-11	1.04744E-10	1.51516E-10	1.57875E-10
3.47249E-10	5.15963E-10	1.74465E-10	7.65038E-10	8.68307E-10	1.05426E-09
3.35082E-10	5.7305E-10	1.72864E-10	7.47497E-10	9.41563E-10	9.29306E-10
3.73908E-10	5.11179E-10	1.73433E-10	9.08838E-10	8.44144E-10	9.09717E-10
3.34311E-10	5.12807E-10	1.73045E-10	6.86848E-10	8.46532E-10	1.12507E-09

TABLE 21 Corrected D-S fusion results.tianjin

Pollutant					
SO ₂	NO ₂	CO	O ₃	PM ₁₀	PM _{2.5}
9.20799E-08	1.22106E-07	4.26766E-08	1.85607E-07	2.16034E-07	2.425E-07
8.44459E-08	1.13829E-07	4.26355E-08	1.71846E-07	2.40466E-07	2.27418E-07
8.08222E-08	9.16249E-08	4.2368E-08	1.39795E-07	3.22489E-07	2.34472E-07
8.44062E-08	1.46184E-07	4.27279E-08	1.90209E-07	2.10291E-07	2.3413E-07
9.2581E-08	1.22922E-07	4.25303E-08	1.73706E-07	2.19349E-07	2.51221E-07
8.66053E-08	1.13131E-07	4.25107E-08	1.49854E-07	2.34329E-07	2.62356E-07

dimensionless indices, facilitating robust comparisons among different pollutants across various regions (Tables 7–9). The daily pollutant concentrations, shown over the course of the year, varied notably between Xi'an, Beijing, and Tianjin (Figure 6).

3.3 Correlation analysis

Within a particular space, air pollution levels are influenced both by the concentration and toxicity of the pollutants and the interactions between different pollutants and regional

environments. To test this hypothesis, we used SPSS matrix correlation analysis, which is widely recognized and utilized, simple to use, intuitive to interpret, and broadly applicable. Notably, this non-parametric method can robustly handle ordinal data, which is crucial in our research context.

We assessed the correlations using the correlation coefficient (Eq. (30)). This equation ensures reliable and statistically valid results, enabling us to gain deeper insights into the relationships and patterns within the data.

$$\rho_{12} = \frac{\text{cov}(m_1, m_2)}{\sigma_{m_1} \sigma_{m_2}} = \frac{E((m_1 - \mu_{m_1})(m_2 - \mu_{m_2}))}{\sigma_{m_1} \sigma_{m_2}} \quad (30)$$

TABLE 22 Evaluation of the nearest neighbors in KNN.xi'an.

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.056,609,488	0.040,558,004	0.066,487,982	0.054,551,824	1
NO ₂ level 2	0.042,155,944	0.151,585,617	0.196,875,714	0.130,205,758	
NO ₂ level 3	0.190,304,093	0.390,670,328	0.507,058,834	0.362,677,752	
CO level 1	0.118,242,005	0.096,660,921	0.084,676,799	0.099,859,908	2
CO level 2	0.04,377,392	0.101,083,173	0.124,252,279	0.089,703,124	
CO level 3	0.129,984,945	0.481,969,556	0.548,436,671	0.386,797,058	
O ₃ level 1	0.041,178,291	0.24,512,773	0.458,592,325	0.248,299,449	1
O ₃ level 2	0.435,117,685	0.75,777,624	0.886,323,046	0.693,072,324	
O ₃ level 3	0.7,002,692	1.127,437,411	1.245,273,238	1.024,326,617	
PM ₁₀ level 1	0.109,520,385	0.077,700,514	0.100,971,768	0.096,064,222	1
PM ₁₀ level 2	0.09,162,904	0.316,484,973	0.457,855,132	0.288,656,382	
PM ₁₀ level 3	0.599,634,053	0.741,587,328	0.883,630,931	0.741,617,437	
PM _{2.5} level 1	0.113,124,541	0.079,991,266	0.113,333,526	0.102,149,778	1
PM _{2.5} level 2	0.181,487,194	0.265,597,113	0.352,406,719	0.266,497,009	
PM _{2.5} level 3	0.440,322,166	0.528,792,178	0.617,578,473	0.528,897,606	
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.831,836,401	0.799,592,627	0.768,817,483	0.80,008,217	3
NO ₂ level 2	0.733,070,969	0.712,425,697	0.687,232,475	0.710,909,714	
NO ₂ level 3	0.584,922,821	0.57,526,776	0.554,805,356	0.571,665,313	
CO level 1	0.261,714,568	0.238,188,844	0.217,802,665	0.239,235,359	2
CO level 2	0.187,246,482	0.190,368,447	0.185,305,735	0.187,640,222	
CO level 3	0.013,487,617	0.481,969,556	0.461,347,048	0.318,934,741	
O ₃ level 1	0.917,444,775	0.780,772,073	0.685,078,468	0.794,431,772	2
O ₃ level 2	0.523,505,381	0.696,175,417	0.722,294,137	0.647,324,978	
O ₃ level 3	0.258,353,866	0.829,285,459	0.898,629,057	0.662,089,461	
PM ₁₀ level 1	1.494,750,669	1.445,338,396	1.397,798,937	1.445,962,667	3
PM ₁₀ level 2	1.293,601,244	1.219,234,497	1.147,764,378	1.220,200,039	
PM ₁₀ level 3	1.090,254,313	1.04,900,284	1.025,972,973	1.055,076,709	
PM _{2.5} level 1	1.714,876,667	1.659,177,964	1.605,548,347	1.65,986,766	3
PM _{2.5} level 2	1.547,575,525	1.493,219,886	1.442,381,952	1.494,392,454	
PM _{2.5} level 3	1.395,446,239	1.352,818,968	1.314,919,222	1.35,439,481	
SO ₂ level 1	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	1
SO ₂ level 2	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	

(Continued on following page)

TABLE 22 (Continued) Evaluation of the nearest neighbors in KNN.xi'an.

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
NO ₂ level 1	0.956,353,885	0.924,018,745	0.892,941,158	0.92,443,793	3
NO ₂ level 2	0.857,588,453	0.835,114,723	0.808,675,472	0.833,792,883	
NO ₂ level 3	0.709,440,305	0.68,673,658	0.654,206,705	0.683,461,196	
CO level 1	0.39,020,888	0.366,228,386	0.344,162,978	0.366,866,748	2
CO level 2	0.315,740,795	0.307,395,511	0.29,361,696	0.305,584,422	
CO level 3	0.14,198,193	0.481,969,556	0.411,128,385	0.345,026,624	
O ₃ level 1	1.205,672,364	1.064,990,267	0.952,150,227	1.074,270,953	3
O ₃ level 2	0.81,173,297	0.903,940,112	0.895,405,672	0.870,359,585	
O ₃ level 3	0.546,581,455	0.933,638,585	0.971,320,177	0.817,180,072	
PM ₁₀ level 1	3.973,582,944	3.923,617,857	3.874,313,879	3.923,838,226	3
PM ₁₀ level 2	3.772,433,518	3.680,457,518	3.588,590,554	3.680,493,864	
PM ₁₀ level 3	3.500,093,981	3.415,229,814	3.33,427,542	3.416,533,072	
PM _{2.5} level 1	2.72,333,398	2.667,269,329	2.612,460,109	2.667,687,806	3
PM _{2.5} level 2	2.552,335,089	2.494,271,858	2.438,116,946	2.494,907,964	
PM _{2.5} level 3	2.384,005,207	2.33,207,887	2.282,487,093	2.332,857,056	
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.735,220,646	0.703,070,053	0.6,726,076	0.703,632,767	3
NO ₂ level 2	0.636,455,214	0.617,733,032	0.593,880,232	0.616,022,826	
NO ₂ level 3	0.488,307,066	0.493,098,274	0.485,721,858	0.489,042,399	
CO level 1	0.2,701,381	0.246,568,072	0.22,601,261	0.247,572,927	2
CO level 2	0.195,670,015	0.197,604,316	0.191,643,799	0.19,497,271	
CO level 3	0.021,911,149	0.481,969,556	0.457,118,569	0.320,333,092	
O ₃ level 1	1.338,648,154	1.19,676,297	1.079,042,028	1.204,817,717	3
O ₃ level 2	0.94,470,876	1.013,160,666	0.993,458,963	0.98,377,613	
O ₃ level 3	0.679,557,245	1.006,262,957	1.030,623,299	0.905,481,167	
PM ₁₀ level 1	1.789,176,344	1.739,615,972	1.691,594,142	1.740,128,819	3
PM ₁₀ level 2	1.588,026,918	1.508,553,784	1.430,853,337	1.50,914,468	
PM ₁₀ level 3	1.363,601,525	1.308,410,507	1.266,857,561	1.312,956,531	
PM _{2.5} level 1	1.674,604,385	1.618,929,761	1.565,379,105	1.61,963,775	3
PM _{2.5} level 2	1.507,553,451	1.453,455,946	1.402,995,156	1.454,668,184	
PM _{2.5} level 3	1.35,657,697	1.314,629,677	1.277,593,731	1.316,266,793	
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.027,240,659	0.068,581,108	0.11,401,783	0.069,946,532	1
NO ₂ level 2	0.126,006,091	0.206,396,724	0.252,789,344	0.195,064,053	
NO ₂ level 3	0.274,154,239	0.453,178,439	0.568,909,964	0.432,080,881	

(Continued on following page)

TABLE 22 (Continued) Evaluation of the nearest neighbors in KNN.xi'an.

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
CO level 1	0.089,800,641	0.151,648,172	0.186,679,807	0.14,270,954	1
CO level 2	0.263,559,506	0.481,969,556	0.647,814,728	0.46,444,793	
CO level 3	0.015,332,556	0.04,720,816	0.081,772,845	0.04,810,452	
O ₃ level 1	0.140,780,234	0.32,923,158	0.537,396,648	0.335,802,821	1
O ₃ level 2	0.534,719,628	0.83,001,925	0.958,324,115	0.774,354,331	
O ₃ level 3	0.799,871,143	1.198,804,543	1.316,557,244	1.105,077,643	
PM ₁₀ level 1	0.040,047,475	0.103,388,587	0.172,886,878	0.10,544,098	1
PM ₁₀ level 2	0.2,411,969	0.424,297,408	0.565,316,163	0.410,270,157	
PM ₁₀ level 3	0.706,822,133	0.848,571,614	0.99,046,006	0.848,617,936	
PM _{2.5} level 1	0.04,187,612	0.11,367,406	0.192,183,295	0.115,911,158	1
PM _{2.5} level 2	0.276,877,138	0.365,921,565	0.455,171,042	0.365,989,915	
PM _{2.5} level 3	0.544,524,755	0.633,938,628	0.723,390,353	0.633,951,245	
SO ₂ level 1	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	1
SO ₂ level 2	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.834,079,878	0.801,834,207	0.771,052,741	0.802,322,275	2
NO ₂ level 2	0.587,166,298	0.577,228,944	0.556,506,914	0.573,634,052	
NO ₂ level 3	0.735,314,446	0.714,630,635	0.689,410,854	0.713,118,645	
CO level 1	0.190,516,019	0.193,165,568	0.187,746,342	0.190,475,976	1
CO level 2	0.264,984,104	0.241,440,818	0.22,098,739	0.242,470,771	
CO level 3	0.016,757,154	0.481,969,556	0.459,692,083	0.319,472,931	
O ₃ level 1	0.409,110,763	0.873,620,695	0.926,232,573	0.736,321,344	1
O ₃ level 2	0.674,262,278	0.798,637,341	0.804,648,908	0.759,182,842	
O ₃ level 3	1.068,201,672	0.92,912,379	0.822,991,244	0.940,105,569	
PM ₁₀ level 1	1.493,119,192	1.443,707,908	1.396,171,694	1.444,332,931	2
PM ₁₀ level 2	1.08,877,026	1.047,616,982	1.024,716,105	1.053,701,116	
PM ₁₀ level 3	1.291,969,767	1.217,638,136	1.146,211,634	1.218,606,512	
PM _{2.5} level 1	1.649,514,468	1.59,385,546	1.540,356,102	1.594,575,343	2
PM _{2.5} level 2	1.332,403,273	1.290,901,487	1.254,431,847	1.292,578,869	
PM _{2.5} level 3	1.482,626,266	1.428,697,126	1.378,482,774	1.429,935,389	

Figure 7 presents the correlations of the concentrations in each region. Figure 8 presents the correlations among different pollutant concentrations (Figure 8). There were complex interconnections and mutual influences among the diverse pollutants in terms of their concentrations in varying intervals. The dynamics of the air pollutants varied among the regions (Figure 7). This finding offers insight into how regional differences contribute to macro-scale air-pollution trends. Figure 8 illustrates the daily interactions among the pollutants. All of the correlation values exceed 4, indicating that they are statistically significant.

3.4 Case analysis

To verify the rationality and effectiveness of the air pollution data fusion method based on fuzzy comprehensive evaluation and DS evidence theory, this study utilized the original dataset provided by the China National Environmental Monitoring Center (CNEMC) and the Qingyue Open Environment Data Center (Tables 4–6). Systematic analysis and quantification (Tables 7–9) were conducted to consolidate air pollution indices from different regions for the same time period.

TABLE 23 Evaluation of the nearest neighbors in KNN.beijing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,014	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	
NO ₂ level 1	0.131,687,243	0.210,802,444	0.257,127,561	0.199,872,416	2
NO ₂ level 2	0.032,921,811	0.073,615,407	0.118,701,276	0.075,079,498	
NO ₂ level 3	0.279,835,391	0.457,660,536	0.573,302,873	0.436,932,933	
CO level 1	0.024,822,695	0.055,505,233	0.0894,995	0.056,609,143	1
CO level 2	0.09,929,078	0.1,589,428	0.193,871,446	0.150,701,675	
CO level 3	0.273,049,645	0.481,969,556	0.655,337,954	0.470,119,052	
O ₃ level 1	0.151,515,151	0.338,798,178	0.546,295,648	0.345,536,326	1
O ₃ level 2	0.545,454,545	0.838,140,405	0.966,376,892	0.783,323,947	
O ₃ level 3	0.810,606,061	1.206,735,193	1.324,458,238	1.113,933,164	
PM ₁₀ level 1	0.050,287,356	0.112,445,947	0.181,313,642	0.114,682,315	1
PM ₁₀ level 2	0.251,436,781	0.432,588,196	0.573,364,079	0.419,129,685	
PM ₁₀ level 3	0.714,717,621	0.85,636,282	0.998,175,522	0.856,418,654	
PM _{2.5} level 1	0.056,666,667	0.126,710,519	0.204,314,572	0.129,230,586	1
PM _{2.5} level 2	0.288,694,379	0.377,432,849	0.466,488,061	0.37,753,843	
PM _{2.5} level 3	0.555,707,757	0.645,023,686	0.734,400,738	0.64,504,406	
SO ₂ level 1	0.313,386,769	0.298,794,995	0.285,029,225	0.29,907,033	3
SO ₂ level 2	0.253,519,332	0.223,585,614	0.220,175,824	0.232,426,923	
SO ₂ level 3	0.096,367,312	0.000917,272	0.343,602,867	0.146,962,484	
NO ₂ level 1	0.480,833,936	0.449,120,382	0.420,181,322	0.450,045,213	2
NO ₂ level 2	0.1317	0.21,080,244	0.355,661	0.232,716,977	
NO ₂ level 3	0.233	0.31,781,801	0.368,536	0.306,441,699	
CO level 1	0.147,899,935	0.125,555,459	0.110,084,606	0.127,846,667	2
CO level 2	0.07,343,185	0.110,550,934	0.126,372,269	0.103,451,684	
CO level 3	0.100,327,015	0.481,969,556	0.528,420,676	0.370,239,082	
O ₃ level 1	0.575,521,889	0.450,264,983	0.407,527,949	0.477,771,607	2
O ₃ level 2	0.181,582,495	0.552,939,714	0.63,636,368	0.456,961,963	
O ₃ level 3	0.08,356,902	0.827,610,162	0.928,302,504	0.613,160,562	
PM ₁₀ level 1	0.762,621,255	0.714,106,716	0.669,642,365	0.715,456,778	2
PM ₁₀ level 2	0.56,147,183	0.525,020,578	0.503,730,411	0.530,074,273	
PM ₁₀ level 3	0.521,826,732	0.575,606,671	0.656,356,636	0.58,459,668	
PM _{2.5} level 1	1.094,204,967	1.039,084,615	0.98,739,739	1.040,228,991	3
PM _{2.5} level 2	0.933,154,078	0.885,500,181	0.844,682,812	0.887,779,024	
PM _{2.5} level 3	0.811,733,934	0.787,641,638	0.773,234,224	0.790,869,932	
SO ₂ level 1	0.014,966,859	0.033,466,914	0.053,963,778	0.034,132,517	1
SO ₂ level 2	0.074,834,295	0.104,768,013	0.170,648,449	0.116,750,253	
SO ₂ level 3	0.231,986,316	0.3,292,709	0.538,858,893	0.36,670,537	

(Continued on following page)

TABLE 23 (Continued) Evaluation of the nearest neighbors in KNN.beijing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
NO ₂ level 1	0.03,292,181	0.073,615,406	0.118,701,276	0.075,079,497	1
NO ₂ level 2	0.131,687,242	0.210,802,444	0.257,127,561	0.199,872,416	
NO ₂ level 3	0.27,983,539	0.457,660,535	0.573,302,873	0.436,932,933	
CO level 1	0.024,822,695	0.055,505,233	0.0894,995	0.056,609,143	1
CO level 2	0.09,929,078	0.1,589,428	0.193,871,446	0.150,701,675	
CO level 3	0.273,049,645	0.481,969,556	0.655,337,954	0.470,119,052	
O ₃ level 1	0.151,515,151	0.338,798,178	0.546,295,647	0.345,536,325	1
O ₃ level 2	0.545,454,545	0.838,140,405	0.966,376,892	0.783,323,947	
O ₃ level 3	0.81,060,606	1.206,735,192	1.324,458,237	1.113,933,163	
PM ₁₀ level 1	0.050,287,355	0.112,445,946	0.181,313,641	0.114,682,314	1
PM ₁₀ level 2	0.251,436,781	0.432,588,195	0.573,364,078	0.419,129,685	
PM ₁₀ level 3	0.71,471,762	0.856,362,819	0.998,175,521	0.856,418,654	
PM _{2.5} level 1	0.056,666,666	0.126,710,518	0.204,314,572	0.129,230,585	1
PM _{2.5} level 2	0.288,694,378	0.377,432,848	0.46,648,806	0.377,538,429	
PM _{2.5} level 3	0.555,707,756	0.645,023,685	0.734,400,738	0.64,504,406	
SO ₂ level 1	0.022,423,972	0.016,721,705	0.030,861,364	0.02,333,568	1
SO ₂ level 2	0.037,443,464	0.067,377,183	0.142,988,629	0.082,603,092	
SO ₂ level 3	0.194,595,485	0.291,880,069	0.509,772,943	0.332,082,832	
NO ₂ level 1	0.018,196,089	0.036,065,115	0.081,275,272	0.045,178,825	1
NO ₂ level 2	0.080,569,343	0.173,844,103	0.220,301,831	0.158,238,426	
NO ₂ level 3	0.228,717,492	0.418,380,857	0.534,650,438	0.393,916,262	
CO level 1	0.007,479,424	0.040,738,103	0.075,682,951	0.041,300,159	1
CO level 2	0.081,947,509	0.145,803,071	0.180,889,043	0.136,213,207	
CO level 3	0.255,706,374	0.481,969,556	0.641,628,662	0.459,768,197	
O ₃ level 1	0.060,631,355	0.260,697,106	0.473,367,897	0.264,898,786	1
O ₃ level 2	0.454,570,749	0.771,407,102	0.899,971,971	0.708,649,941	
O ₃ level 3	0.719,722,264	1.141,043,592	1.258,893,552	1.039,886,469	
PM ₁₀ level 1	0.034,127,017	0.052,820,212	0.12,054,281	0.069,163,346	1
PM ₁₀ level 2	0.167,022,408	0.367,190,242	0.509,374,706	0.347,862,452	
PM ₁₀ level 3	0.651,580,846	0.793,797,013	0.936,018,637	0.793,798,832	
PM _{2.5} level 1	0.034,305,073	0.060,919,225	0.138,166,242	0.077,796,847	1
PM _{2.5} level 2	0.221,735,932	0.311,054,413	0.400,483,852	0.311,091,399	
PM _{2.5} level 3	0.489,963,498	0.579,470,091	0.668,992,816	0.579,475,468	
SO ₂ level 1	0.319,344,384	0.304,745,278	0.290,954,568	0.305,014,743	3
SO ₂ level 2	0.259,476,947	0.229,543,229	0.225,433,568	0.238,151,248	
SO ₂ level 3	0.102,324,927	0.005,040,343	0.341,947,841	0.149,771,037	
NO ₂ level 1	0.479,884,773	0.448,173,778	0.419,243,912	0.449,100,821	2
NO ₂ level 2	0.1317	0.21,080,244	0.355,661	0.232,716,977	
NO ₂ level 3	0.233	0.31,781,801	0.368,536	0.306,441,699	

(Continued on following page)

TABLE 23 (Continued) Evaluation of the nearest neighbors in KNN.beijing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
CO level 1	0.148,222,082	0.125,871,263	0.110,372,229	0.128,155,191	2
CO level 2	0.073,753,996	0.111,069,296	0.126,433,301	0.103,626,752	
CO level 3	0.100,004,869	0.481,969,556	0.528,208,239	0.370,060,888	
O ₃ level 1	0.535,332,911	0.412,641,387	0.381,827,011	0.443,267,103	1
O ₃ level 2	0.141,393,518	0.547,785,851	0.637,646,321	0.44,227,523	
O ₃ level 3	0.123,757,998	0.836,641,503	0.939,933,043	0.633,444,181	
PM ₁₀ level 1	0.796,244,551	0.747,650,289	0.702,902,409	0.748,932,416	2
PM ₁₀ level 2	0.595,095,126	0.554,771,028	0.52,830,418	0.559,390,112	
PM ₁₀ level 3	0.539,352,214	0.585,796,387	0.660,208,371	0.595,118,991	
PM _{2.5} level 1	1.068,399,516	1.013,318,543	0.96,176,705	1.014,495,037	3
PM _{2.5} level 2	0.907,792,639	0.860,633,606	0.820,571,371	0.862,999,205	
PM _{2.5} level 3	0.788,688,146	0.76,600,591	0.753,356,197	0.769,350,084	
SO ₂ level 1	0.332,282,096	0.317,668,013	0.303,826,544	0.317,925,551	3
SO ₂ level 2	0.27,241,466	0.242,480,941	0.236,965,522	0.250,620,374	
SO ₂ level 3	0.115,262,639	0.017,978,055	0.338,686,952	0.157,309,215	
NO ₂ level 1	0.483,041,049	0.451,321,586	0.42,236,131	0.452,241,315	3
NO ₂ level 2	0.384,275,617	0.375,221,304	0.358,462,099	0.372,653,007	
NO ₂ level 3	0.236,127,469	0.31,932,639	0.369,134,398	0.308,196,086	
CO level 1	0.149,642,357	0.127,263,955	0.111,642,546	0.129,516,286	2
CO level 2	0.075,174,272	0.111,328,075	0.126,711,793	0.104,404,713	
CO level 3	0.098,584,594	0.481,969,556	0.527,272,976	0.369,275,709	
O ₃ level 1	0.613,522,829	0.486,217,991	0.433,858,241	0.511,199,687	2
O ₃ level 2	0.219,583,435	0.560,426,398	0.637,483,248	0.472,497,694	
O ₃ level 3	0.04,556,808	0.820,790,936	0.918,787,912	0.595,048,976	
PM ₁₀ level 1	0.818,019,546	0.769,377,368	0.724,460,049	0.770,618,987	2
PM ₁₀ level 2	0.616,870,121	0.57,426,597	0.544,734,604	0.578,623,565	
PM ₁₀ level 3	0.551,499,107	0.593,319,335	0.663,600,401	0.602,806,281	
PM _{2.5} level 1	0.997,590,959	0.942,629,108	0.891,490,883	0.94,390,365	3
PM _{2.5} level 2	0.838,341,633	0.792,709,598	0.754,986,597	0.795,345,943	
PM _{2.5} level 3	0.726,405,853	0.708,075,284	0.700,799,685	0.711,760,274	

TABLE 24 Evaluation of the nearest neighbors in KNN.tianjing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
SO ₂ level 1	0.014,966,767	0.033,466,832	0.053,963,701	0.034,132,433	1
SO ₂ level 2	0.074,834,203	0.104,767,922	0.170,648,377	0.116,750,167	
SO ₂ level 3	0.231,986,224	0.329,270,808	0.53,885,882	0.366,705,284	
NO ₂ level 1	0.032,921,689	0.073,615,297	0.118,701,175	0.075,079,387	2
NO ₂ level 2	0.131,687,121	0.210,802,349	0.257,127,468	0.199,872,312	

(Continued on following page)

TABLE 24 (Continued) Evaluation of the nearest neighbors in KNN.tianjing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
NO ₂ level 3	0.279,835,269	0.457,660,439	0.573,302,778	0.436,932,829	
CO level 1	0.024,822,652	0.055,505,195	0.089,499,464	0.056,609,104	1
CO level 2	0.099,290,737	0.158,942,767	0.193,871,413	0.150,701,639	
CO level 3	0.273,049,603	0.481,969,556	0.65,533,792	0.470,119,026	
O ₃ level 1	0.54,545,436	0.838,140,264	0.966,376,752	0.783,323,792	2
O ₃ level 2	0.151,514,966	0.338,798,012	0.546,295,493	0.345,536,157	
O ₃ level 3	0.810,605,875	1.206,735,055	1.324,458,101	1.11,393,301	
PM ₁₀ level 1	0.05,028,714	0.112,445,754	0.181,313,462	0.114,682,119	1
PM ₁₀ level 2	0.251,436,566	0.43,258,802	0.573,363,908	0.419,129,498	
PM ₁₀ level 3	0.714,717,454	0.856,362,655	0.998,175,359	0.856,418,489	
PM _{2.5} level 1	0.056,666,424	0.126,710,302	0.204,314,371	0.129,230,365	1
PM _{2.5} level 2	0.288,694,183	0.377,432,658	0.466,487,874	0.377,538,238	
PM _{2.5} level 3	0.555,707,572	0.645,023,503	0.734,400,557	0.645,043,877	
SO ₂ level 1	0.612,472,845	0.515,188,261	0.532,036,473	0.553,232,527	1
SO ₂ level 2	0.769,624,866	0.739,691,148	0.717,448,218	0.742,254,744	
SO ₂ level 3	0.829,492,302	0.814,662,939	0.800,118,714	0.814,757,985	
NO ₂ level 1	1.105,366,352	1.072,949,738	1.041,605,919	1.073,307,336	2
NO ₂ level 2	0.858,452,772	0.825,053,904	0.782,664,951	0.822,057,209	
NO ₂ level 3	1.00660,092	0.982,543,911	0.955,049,965	0.981,398,265	
CO level 1	0.401,532,374	0.377,526,619	0.355,371,797	0.378,143,597	2
CO level 2	0.327,064,289	0.318,133,054	0.303,916,765	0.316,371,369	
CO level 3	0.153,305,423	0.481,969,556	0.408,348,675	0.347,874,552	
O ₃ level 1	1.566,949,046	1.423,520,267	1.299,737,648	1.430,068,987	3
O ₃ level 2	1.173,009,652	1.211,801,257	1.177,887,605	1.187,566,171	
O ₃ level 3	0.907,858,137	1.156,254,011	1.16,124,607	1.075,119,406	
PM ₁₀ level 1	2.354,368,872	2.30,463,022	2.256,037,098	2.305,012,063	3
PM ₁₀ level 2	2.153,219,447	2.067,987,152	1.983,554,754	2.068,253,784	
PM ₁₀ level 3	1.906,008,697	1.836,221,647	1.775,108,958	1.839,113,101	
PM _{2.5} level 1	2.217,512,352	2.161,588,579	2.107,228,935	2.162,109,955	3
PM _{2.5} level 2	2.047,913,844	1.991,230,661	1.93,703,186	1.992,058,788	
PM _{2.5} level 3	1.885,531,689	1.836,957,136	1.791,546,185	1.83,801,167	
SO ₂ level 1	0.57,623,542	0.478,950,836	0.504,071,873	0.519,752,709	1
SO ₂ level 2	0.73,338,744	0.703,453,722	0.681,619,786	0.706,153,649	
SO ₂ level 3	0.793,254,877	0.778,431,914	0.76,390,786	0.77,853,155	
NO ₂ level 1	0.883,326,864	0.851,042,068	0.820,130,621	0.851,499,851	3
NO ₂ level 2	0.784,561,432	0.763,088,232	0.737,327,222	0.761,658,962	
NO ₂ level 3	0.636,413,284	0.620,761,114	0.594,764,283	0.617,312,894	
CO level 1	0.324,389,632	0.315,593,107	0.301,477,285	0.313,820,008	1
CO level 2	0.150,630,767	0.481,969,556	0.408,978,678	0.347,193,001	

(Continued on following page)

TABLE 24 (Continued) Evaluation of the nearest neighbors in KNN.tianjing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
CO level 3	0.398,857,717	0.374,857,792	0.352,723,566	0.375,479,692	
O ₃ level 1	1.246,437,568	1.105,356,024	0.990,880,736	1.114,224,776	3
O ₃ level 2	0.852,498,174	0.936,770,299	0.924,537,901	0.904,602,125	
O ₃ level 3	0.587,346,659	0.95,452,167	0.987,977,376	0.843,281,902	
PM ₁₀ level 1	3.174,600,904	3.124,718,222	3.075,671,033	3.12,499,672	3
PM ₁₀ level 2	2.973,451,479	2.883,858,764	2.794,560,926	2.883,957,056	
PM ₁₀ level 3	2.709,789,892	2.62,998,343	2.555,606,682	2.631,793,335	
PM _{2.5} level 1	2.288,053,493	2.232,106,243	2.177,671,283	2.23,261,034	3
PM _{2.5} level 2	2.118,219,613	2.061,302,393	2.006,769,848	2.062,097,285	
PM _{2.5} level 3	1.954,821,561	1.905,668,884	1.859,533,516	1.906,674,653	
SO ₂ level 1	0.014,966,775	0.033,466,839	0.053,963,708	0.03,413,244	1
SO ₂ level 2	0.074,834,211	0.104,767,929	0.170,648,383	0.116,750,174	
SO ₂ level 3	0.231,986,232	0.329,270,816	0.538,858,826	0.366,705,291	
NO ₂ level 1	0.032,921,665	0.073,615,276	0.118,701,155	0.075,079,365	1
NO ₂ level 2	0.131,687,097	0.21,080,233	0.257,127,449	0.199,872,292	
NO ₂ level 3	0.279,835,245	0.45,766,042	0.57,330,276	0.436,932,808	
CO level 1	0.024,822,652	0.055,505,195	0.089,499,464	0.056,609,104	1
CO level 2	0.099,290,737	0.158,942,767	0.193,871,413	0.150,701,639	
CO level 3	0.273,049,603	0.481,969,556	0.65,533,792	0.470,119,026	
O ₃ level 1	0.151,514,961	0.338,798,008	0.54,629,549	0.345,536,153	1
O ₃ level 2	0.545,454,355	0.838,140,261	0.966,376,749	0.783,323,788	
O ₃ level 3	0.81,060,587	1.206,735,052	1.324,458,097	1.113,933,007	
PM ₁₀ level 1	0.050,287,146	0.112,445,759	0.181,313,467	0.114,682,124	1
PM ₁₀ level 2	0.251,436,571	0.432,588,025	0.573,363,913	0.419,129,503	
PM ₁₀ level 3	0.714,717,458	0.85,636,266	0.998,175,363	0.856,418,494	
PM _{2.5} level 1	0.056,666,433	0.126,710,309	0.204,314,377	0.129,230,373	1
PM _{2.5} level 2	0.28,869,419	0.377,432,665	0.46,648,788	0.377,538,245	
PM _{2.5} level 3	0.555,707,579	0.645,023,509	0.734,400,563	0.645,043,884	
SO ₂ level 1	0.01,774,672	0.012,186,998	0.112,918,074	0.047,617,264	1
SO ₂ level 2	0.077,614,157	0.064,410,331	0.056,297,884	0.066,107,457	
SO ₂ level 3	0.1,394,053	0.236,689,884	0.468,997,091	0.281,697,425	
NO ₂ level 1	0.089,999,811	0.06,589,191	0.070,134,899	0.075,342,207	1
NO ₂ level 2	0.008,765,621	0.138,128,099	0.180,728,192	0.109,207,304	
NO ₂ level 3	0.156,913,769	0.368,148,393	0.484,274,969	0.33,644,571	
CO level 1	0.017,707,586	0.025,822,296	0.059,031,247	0.034,187,043	1
CO level 2	0.056,760,499	0.128,508,664	0.163,482,011	0.116,250,392	
CO level 3	0.230,519,365	0.481,969,556	0.622,042,264	0.444,843,728	
O ₃ level 1	0.022,191,318	0.199,202,146	0.413,156,024	0.211,516,496	1
O ₃ level 2	0.371,748,076	0.715,242,837	0.843,441,924	0.643,477,612	

(Continued on following page)

TABLE 24 (Continued) Evaluation of the nearest neighbors in KNN.tianjing

Pollutant	Neighbor 1	Neighbor 2	Neighbor 3	Average	Level
O ₃ level 3	0.636,899,591	1.084,351,258	1.202,017,625	0.974,422,825	
PM ₁₀ level 1	0.562,374,208	0.703,334,095	0.84,472,128	0.703,476,528	2
PM ₁₀ level 2	0.169,061,528	0.128,981,092	0.121,678,744	0.139,907,121	
PM ₁₀ level 3	0.032,087,897	0.284,288,275	0.422,269,749	0.246,215,307	
PM _{2.5} level 1	0.170,938,446	0.237,717,889	0.315,981,141	0.241,545,825	2
PM _{2.5} level 2	0.194,553,876	0.149,077,139	0.139,431,779	0.161,020,931	
PM _{2.5} level 3	0.399,027,333	0.484,402,713	0.571,063,577	0.484,831,208	
SO ₂ level 1	0.851,085,969	0.836,253,056	0.821,697,661	0.836,345,562	3
SO ₂ level 2	0.791,218,533	0.761,284,814	0.738,817,138	0.763,773,495	
SO ₂ level 3	0.634,066,512	0.536,781,928	0.549,160,487	0.573,336,309	
NO ₂ level 1	1.098,390,041	1.065,976,733	1.034,643,654	1.066,336,809	3
NO ₂ level 2	0.999,624,609	0.97,563,099	0.948,178,819	0.974,478,139	
NO ₂ level 3	0.85,147,646	0.818,494,539	0.77,648,706	0.81,548,602	
CO level 1	0.400,284,714	0.376,281,669	0.354,136,415	0.376,900,933	2
CO level 2	0.325,816,629	0.316,947,957	0.302,778,315	0.315,180,967	
CO level 3	0.152,057,764	0.481,969,556	0.408,640,498	0.347,555,939	
O ₃ level 1	1.347,022,528	1.205,070,424	1.087,081,933	1.213,058,295	2
O ₃ level 2	0.687,931,619	1.011,247,384	1.034,816,307	0.91,133,177	
O ₃ level 3	0.953,083,134	1.020,227,789	0.999,904,147	0.99,107,169	
PM ₁₀ level 1	2.29,300,266	2.243,279,017	2.1,947,336	2.243,671,759	3
PM ₁₀ level 2	2.091,853,234	2.007,090,084	1.923,193,434	2.007,378,918	
PM ₁₀ level 3	1.846,455,106	1.7,778,023	1.718,204,491	1.780,820,632	
PM _{2.5} level 1	2.566,897,862	2.510,870,718	2.456,180,641	2.511,316,407	3
PM _{2.5} level 2	2.396,269,009	2.338,567,532	2.282,922,575	2.339,253,039	
PM _{2.5} level 3	2.229,488,128	2.17,842,686	2.129,909,458	2.179,274,815	

3.4.1 Calculation of BPA using the interval distance similarity

Calculation of interval similarity is critical in assessing the similarity between two intervals of data, particularly in the context of the variability of pollution concentrations. Equations (14) and (15) (Wu and Mendel, 2009), which calculate the similarity between fluctuating concentration values, generate BPA values that encapsulate the relationships among interval data points for atmospheric pollutants (Graziani et al., 2019). This approach ensures the accuracy of these similarity measures and establishes the foundation for data fusion, fostering more robust analysis and a comprehensive understanding of pollution trends.

BPA values corresponding to the atmospheric pollutant concentrations on selected days are illustrated in Tables 10–12. These values, derived from the interval similarity calculations (equations (14) and (15)), facilitate objective assessment of the regional pollution levels for each pollutant, following Qin and Xiao (2019). These BPA values quantify the pollution level and

establish the basis for the advanced stages of data fusion. These empirical insights (Tables 10–12) are paramount, enabling a systematic comparison pollutant levels and thereby highlighting the broader importance and impact of each pollutant.

3.4.2 Sequential fusion via DS evidence theory

Regional pollution levels were estimated via sequential fusion (Figure 4) using DS evidence theory (Tables 13–15). The uncorrected fusion results deviate significantly from the CNEMC AQI values, yielding an error rate of 50%.

3.4.3 Subjective and objective weights

Relying on uncorrected fusion data can lead to misleading conclusions. To improve estimation accuracy, the influence of toxicity and the interactions between pollutants must be considered. We therefore calculated the subjective and objective weights of the pollutants using the CRITIC weighting method (Tables 16–18).

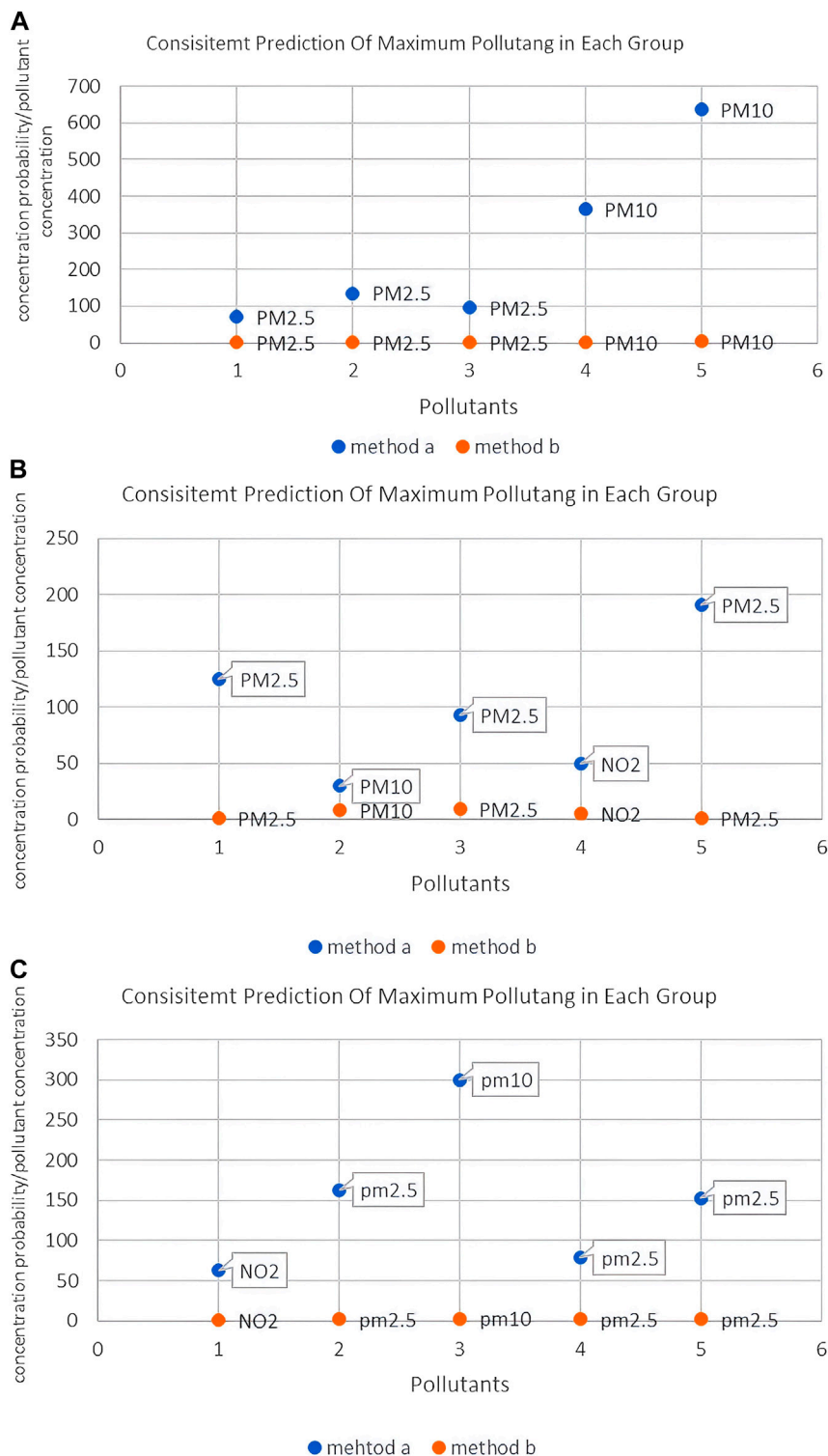


FIGURE 9 Comparison of predicted concentrations of major pollutants. (A) Xi'an; (B) Beijing; (C) Tianjin. (A) Xi'an; (B) Beijing; (C) Tianjin.

3.4.4 DS fusion result correction using objective–subjective weights

The DS fusion results were corrected using objective–subjective weights (Tables 19–21). The high error

rate in the uncorrected data (Table 7) underscores the need for careful interpretation and potential correction. Fusion-based correction achieved a remarkable and significant reduction in error, from the initial 50–11%, suggesting a

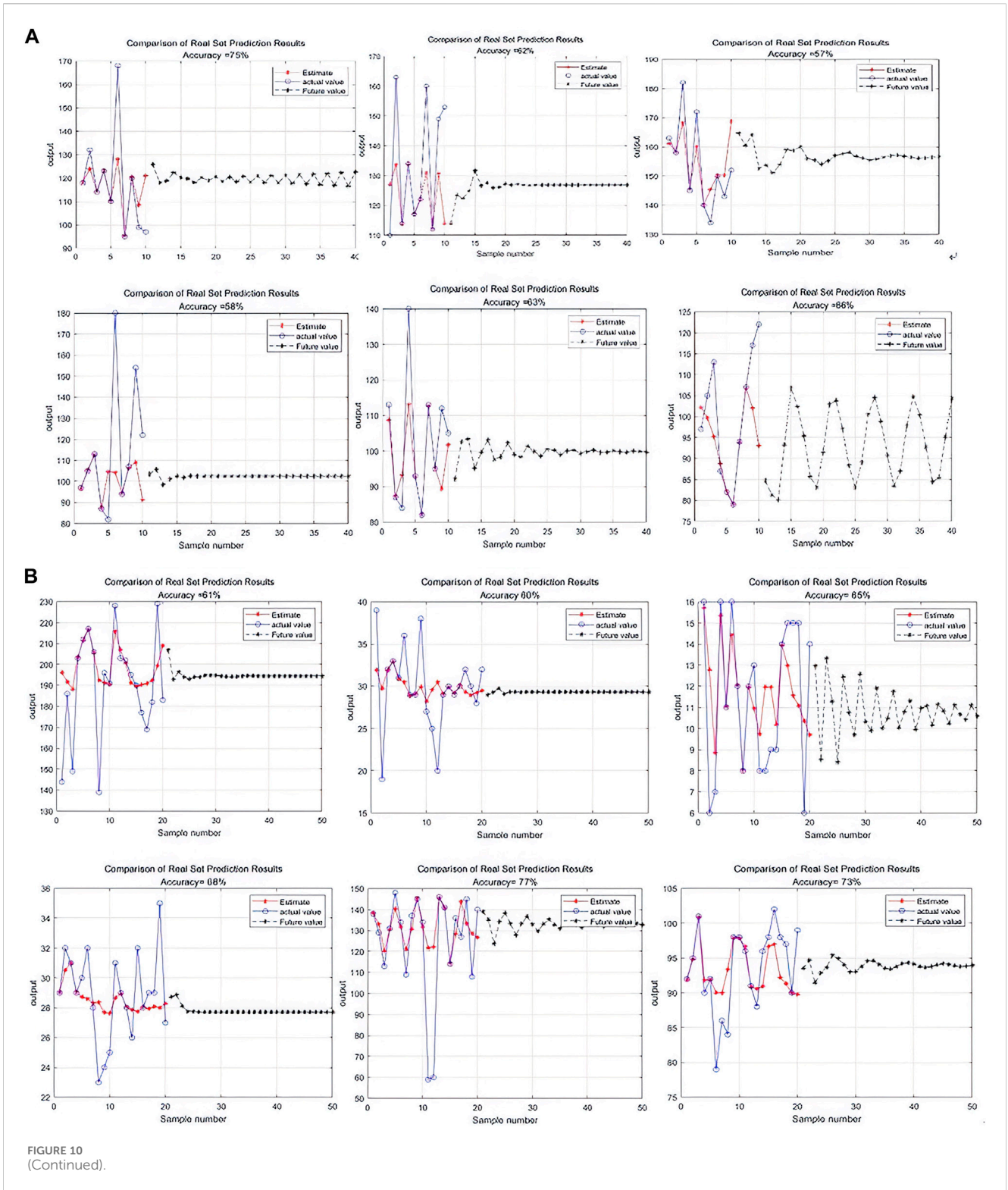


FIGURE 10 (Continued).

close correlation between the corrected results and the AQI values, reflecting the efficacy of the correction approach. This highlights the importance of addressing potential biases and uncertainty in the fusion process, generating more precise and reliable outcomes.

3.4.5 Concentration estimation via on the KNN approach and matter-element extension model

We used an extended KNN algorithm to estimate air pollutant levels, which were then correlated with air-quality levels. Tables 22–24 illustrates the variability in air quality (from good

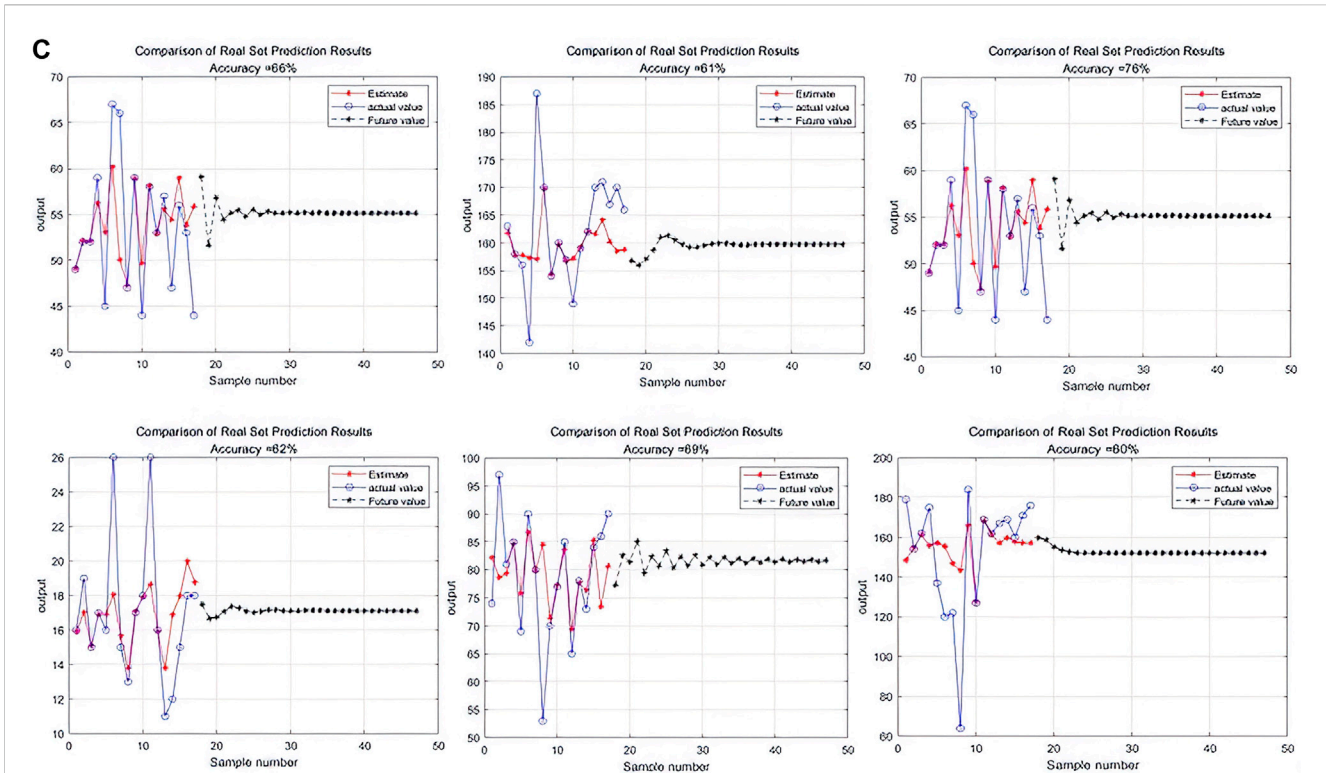


FIGURE 10 Comparison of support vector machine (SVM) training results. (A) Xi'an; (B) Beijing; (C) Tianjin.

to severely polluted) during the study period. These results were consistent with the AQI values, facilitating the identification of major pollutants such as $PM_{2.5}$ and PM_{10} (Figure 9) and validating the accuracy and efficacy of the proposed method.

3.4.6 Further refinement of the DS fusion results

Given that air quality data are influenced by multiple factors, including meteorological conditions, vehicular traffic, and industrial activity, such data are inherently uncertain. Exclusive reliance on the raw DS fusion results might overlook pivotal environmental and socio-economic variables.

Refining DS fusion outcomes extends beyond merely enhancing accuracy, by more comprehensive and accurately portraying air quality and thereby equipping policymakers with evidence-based recommendations. For instance, a raw high pollution reading might trigger unwarranted alarm and panic, while a refined reading could suggest the necessity for additional monitoring sites or improved data fusion methodology. As air quality directly impacts public health, it is imperative to deliver precise data. Using raw values may cause the effects of pollutants to be under- or over-estimated, leading to misguided health advisories and policies. Here, via objective–subjective weighting, were able to integrate environmental and socio-economic aspects into the analysis, achieving more accurate and reliable outcomes. Such refinement is not limited to air quality assessment. As this method can handle uncertainty and complexity, it is applicable in other contexts, such as water and soil quality evaluations, highlighting the versatility and value of this novel approach.

The refinement of DS fusion outcomes were comprehensively validated using expanded datasets from urban environments in Xi'an, Beijing, and Tianjin, cities with distinct environmental and socioeconomic profiles. These findings highlight both the robustness of this refinement process and its applicability in diverse air quality scenarios. By integrating data from these cities, the method can provide substantially more refined and precise air quality assessments. This is critical for informed policy-making and effective public health management.

3.5 Comparison with SVM

To establish the effectiveness of the proposed method, we compared it to SVM. For our dataset, SVM achieved 75% prediction accuracy (Figure 10), whereas our method performed significantly better, achieving 86% accuracy (Figure 11). Further, unlike traditional SVM approaches, our method exhibited remarkable robustness and stability in handling uncertainty and noise. These findings validate the efficacy of our method and highlight its potential for evaluating trends in air pollution.

4 Discussion

In the current context of environmental science, precise estimation is essential for effective policy development. Our novel method aims to characterize air pollution dynamics with greater accuracy by integrating the evidence theory and the KNN algorithm.

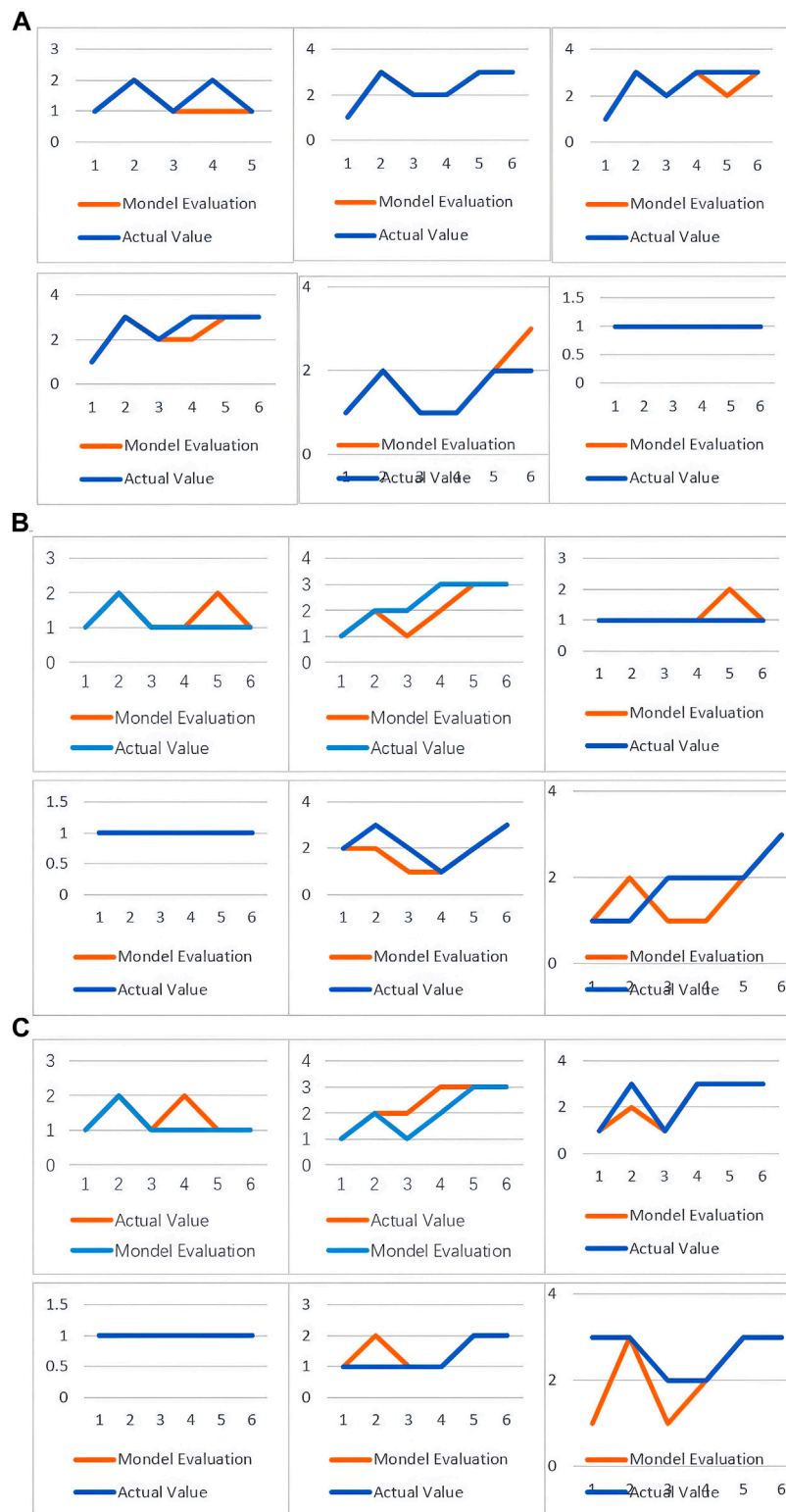


FIGURE 11
 Comparison between values based on fuzzy evaluation of atmospheric pollution concentrations using evidence theory with KNN on the one hand and actual values on the other. (A) Xi'an; (B) Beijing; (C) Tianjin.

BPA, a cornerstone of the DS theory of evidence, effectively addresses the uncertainty inherent in atmospheric pollution data; we therefore used it for decision-making under uncertainty.

Consistent with the findings of Zhong and Liu (2019), integrating BPA into our data fusion strategy reduced the uncertainty, enhancing the precision and reliability of the results.

Given the complexity of atmospheric pollution, a single monitoring point cannot capture the full extent of pollutant diffusion and distribution. Sequential fusion of atmospheric pollution concentrations, however, integrates multi-source monitoring data, estimates pollution levels more accurately and reliably, enhances data accuracy and reliability, and comprehensively describes the spatiotemporal variation in air pollution. By considering spatial correlation, this method more precisely assesses pollutant concentrations. This method achieves several key innovations, in terms of its strategy for handling small datasets, interdisciplinary approach, improved pollution atmospheric monitoring and prediction, and ability to consider multiple environments.

4.1 Innovative strategy for small samples

Limited data availability poses a primary challenge in environmental science, and especially in atmospheric pollution. Therefore, the effective analysis of small samples is important in improving the universality and applicability of models. Here, we aimed to achieve efficient and accurate air quality assessment with limited data availability. Consequently, we did not use the commonly applied neural network approach, primarily because it typically requires large datasets for effective training, and may not achieve optimal performance for small samples. By integrating evidence theory and the KNN approach, our method can handle limited data. Compared to other methods, this approach is more adaptable and precise, extracting meaningful insights from small datasets.

4.2 Interdisciplinary approach

By integrating evidence theory and the KNN algorithm, this model applies an interdisciplinary strategy. It represents a paradigm shift in atmospheric pollution data processing, with applicability in other environmental monitoring domains.

4.3 Enhanced monitoring and prediction

Compared to the traditional AQI approach, our model yields more comprehensive and precise assessments, by integrating multiple data sources and considering the characteristics of multiple pollutants. It thus comprehensively characterizes air pollution-related dynamics, especially for regions with limited data. This approach lays a robust foundation for effective policy development.

4.4 Data quality and methodological accuracy

This hybrid method, integrating interval similarity and subjective-objective weighting, accurately accounts for the inherent variability in pollutant concentrations and toxicological characteristics. By emphasizing data integrity and credibility, this approach will facilitate accurate and reliable assessments.

4.5 Expansion to multiple environments

Expanding our data to include Xi'an, Beijing, and Tianjin, which have varied urban settings, substantially strengthened the robustness and applicability of our findings, demonstrating the model's adaptability and effectiveness across different geographical landscapes. Such broad applicability reveals its potential for accurate and comprehensive air quality assessment for diverse urban contexts. This will enable it to become an essential tool for policymakers and environmental scientists.

4.6 Limitations

Although this method achieved significant results using a small sample data, this dataset was limited in its scope and diversity, focusing on specific urban environments and particular pollutants. This may restrict the model's applicability under a broader range of geographical and environmental conditions. The model does not currently adequately incorporate key external factors such as meteorological conditions, regional industrial activity, and traffic flow, which significantly impact air quality. While the model performs well with historical data, its ability to process real-time data and predict future trends remains to be validated. The current model's complexity may limit its direct application by non-professional users such as policymakers and environmental management agencies.

4.7 Future directions

Data from a wider range of geographical areas and more complex environmental and meteorological factors should be included to enhance the model's generalizability and accuracy. New and optimized algorithms should be used to enhance computational efficiency and accuracy. To validate its cross-field applicability, the method should be applied in environmental monitoring fields such as water quality assessment or forest health monitoring. Long-term research is required to better understand how pollutants change over time and under varying conditions. Finally, research findings should be translated into concrete policy recommendations and practical guidelines to promote environmental protection and sustainable development.

4.8 Conclusion

Our data-fusion evaluation method, which considers spatiotemporal variation, facilitates integration of multi-pollutant measurement data, provides a more granular view of air quality, and improves the evaluation of pollution levels and potential health impacts. It thus provides a tool for air quality evaluation during environmental assessments and in formulating public health policies.

Despite this significant progress, there remains broad scope for future development, particularly in terms of expanding the dataset, diversifying the model's applications, and optimizing the user interface. These improvements are anticipated to enhance the

model's role in air quality assessment and in formulating environmental management policies.

(71874134) and Shaanxi Natural Science Basic Research Plan - Key Project (2019JZ-30).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

All authors contributed to the study conception and design. CB analyzed the data and wrote the Introduction. GH reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China

References

Ai, J., Niu, H., Chen, Z., Li, X., Zeng, Y., Zhuang, X., et al. (2022). D-S multi-information fusion GIL partial discharge joint recognition method based on neural network. *High. Volt. Eng.*, 481–488. doi:10.13336/j.1003-6520.hve.20211249

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46, 175–185. doi:10.1080/00031305.1992.10475879

Carlsaw, D. C., and Rhys-Tyler, G. (2013). New insights from comprehensive on-road measurements of NO_x, NO₂ and NH₃ from vehicle emission remote sensing in London, UK. *Atmos. Environ.* 81, 339–347. doi:10.1016/j.atmosenv.2013.09.026

Chen, G., and Zhu, Y. (2014). An overview on the control of urban air pollution. *Sci. Technol. Rev.* 32, 15–22.

Chen, M., Qin, Q., Liu, F., Wang, Y., Wu, C., Yan, Y., et al. (2022). How long-term air pollution and its metal constituents affect type 2 diabetes mellitus prevalence? Results from Wuhan Chronic Disease Cohort. *Environ. Res.* 212, 113158. doi:10.1016/j.envres.2022.113158

Cover, T., and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory.* 13, 21–27. doi:10.1109/TIT.1967.1053964

Cui, J., Sang, M., Bo, X., Wang, P., Xue, X. D., Guo, J., et al. (2022). Rapid assessment of optimization schemes for atmospheric emission sources based on integer programming algorithm. *Environ. Eng.* 40 (04), 202–208. doi:10.13205/j.hjgc.202204029

Dai, P., Yuming, W., and Feng, Z. (2018). An improved KNN algorithm based on analytic hierarchy process. *J. Liaoning Univ. Petrol. Chem. Technol.* 38, 87.

Dempster, A. P. (2008). "Upper and lower probabilities induced by a multivalued mapping," in *Classic works of the dempster-shafer theory of belief functions*. Editors R. Yager, and L. Liu (Berlin, Heidelberg: Springer), 57–72. doi:10.1007/978-3-540-44792-4_3

Dyer, R. F., and Forman, E. H. (1992). Group decision support with the analytic hierarchy process. *Decis. Support Syst.* 8, 99–124. doi:10.1016/0167-9236(92)90003-8

Eren, B., Aksangür, İ., and Erden, C. (2023). Predicting next hour fine particulate matter (PM_{2.5}) in the Istanbul metropolitan city using deep learning algorithms with time windowing strategy. *Urban Clim.* 48, 101418. doi:10.1016/j.uclim.2023.101418

Feng, X., Wang, B., Gao, G., Gao, S., Xie, C., and Shi, J. W. (2023). Mn₂Co₃O_x bimetallic oxide prepared by ultrasonic technology for significantly improved catalytic performance in the reduction of NO_x with NH₃. *Fuel* 352, 129159. doi:10.1016/j.fuel.2023.129159

Graziani, N. S., Carreras, H., and Wannaz, E. (2019). Atmospheric levels of BPA associated with particulate matter in an urban environment. *Heliyon* 5, e01419. doi:10.1016/j.heliyon.2019.e01419

Acknowledgments

We thank Qingyue Data (data.epmap.org) for providing the environmental data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

He, K., Wang, T., Zhang, F., and Jin, X. (2022). Anomaly detection and early warning via a novel multiblock-based method with applications to thermal power plants. *Measurement* 193, 110979. doi:10.1016/j.measurement.2022.110979

Huihui, X., and Yanming, D. (2013). Improved the KNN algorithm based on related to the distance of attribute value. *Comput. Sci.* 40, 157–159.

Karavas, Z., Karayannis, V., and Moustakas, K. (2021). Comparative study of air quality indices in the European Union towards adopting a common air quality index. *Energy Environ.* 32 (6), 959–980. doi:10.1177/0958305X20921800

Kong, L., Wang, L., Li, F., Li, J., Wang, Y., Cai, Z., et al. (2023). Life cycle-oriented low-carbon product design based on the constraint satisfaction problem. *Energy Convers. Manag.* 286, 117069. doi:10.1016/j.enconman.2023.117069

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, J., et al. (2003). Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmos. Environ.* 37, 4539–4550. doi:10.1016/S1352-2310(03)00583-1

Li, G. D., Yamaguchi, D., and Nagai, M. (2007). A grey-based decision-making approach to the supplier selection problem. *Math. Comput. Modell.* 46, 573–581. doi:10.1016/j.mcm.2006.11.021

Li, J., Zhang, G., Ned, J. P., and Sui, L. (2023). How does digital finance affect green technology innovation in the polluting industry? Based on the serial two-mediator model of financing constraints and research and development (R&D) investments. *Environ. Sci. Pollut. Res. Int.* 30 (29), 74141–74152. doi:10.1007/s11356-023-27593-y

Li, L., Wang, J., Franklin, M., Yin, Q., Wu, J., Camps-Valls, G., et al. (2023). Improving air quality assessment using physics-inspired deep graph learning. *npj Clim. Atmos. Sci.* 6, 152. doi:10.1038/s41612-023-00475-3

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., et al. (2016). Geospatial big data handling theory and methods: a review and research challenges. *ISPRS J. Photogramm.* 115, 119–133. doi:10.1016/j.isprsjprs.2015.10.012

Lü, C., Chen, S., Hua, X., and Xiang, W. (2021). Turbine through-flow fault diagnosis method and application based on the improved KNN algorithm. *Therm. Power Gener.* 50, 7. doi:10.19666/j.rld.202012289

Luo, J., Zhuo, W., Liu, S., and Xu, B. (2024). The optimization of carbon emission prediction in low carbon energy economy under big data. *IEEE Access* 12, 14690–14702. doi:10.1109/ACCESS.2024.3351468

Mo, H., Zhang, L., and Wang, H. (2020). A fuzzy comprehensive evaluation method for air quality assessment. *Environ. Monit. Assess.* 192, 243. doi:10.1007/s10661-020-8211-4

- Murena, F. (2004). Measuring air quality over large urban areas: development and application of an air pollution index at the urban area of Naples. *Atmos. Environ.* 38, 6195–6202. doi:10.1016/j.atmosenv.2004.07.023
- Pope, C. A., III, and Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *J. Air Waste Manag. Assoc.* 56, 709–742. doi:10.1080/10473289.2006.10464485
- Pope, R. J., Coleman, P., and Patterson, K. G. (2022). The limitations of the air quality index: a critical review. *Environ. Sci. Policy* 134, 55–65. doi:10.1007/978-981-10-5792-2_8
- Priti, K., and Kumar, P. (2022). A critical evaluation of air quality index models (1960–2021). *Environ. Monit. Assess.* 194 (4), 324. doi:10.1007/s10661-022-09896-8
- Qin, B., and Xiao, F. (2019). An improved method to determine basic probability assignment with interval number and its application in classification. *Int. J. Distrib. Sens. Netw.* 15, 155014771882052. doi:10.1177/1550147718820524
- Ren, C., Niu, H. X., Chen, G. W., Si, Y. B., and She, Y. M. (2022). Heterogeneous data fusion of high-speed railway signal system based on DS evidence theory. *Transducer Microsyst. Technol.* 41, 54–57. doi:10.13873/J.1000-9787(2022)03-0054-04
- Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* 1, 83–98. doi:10.1504/IJSSCI.2008.017590
- Seinfeld, J. H., and Pandis, S. N. (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. Hoboken, NJ, USA: John Wiley & Sons.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ, USA: Princeton University Press.
- Shang, K., Xu, L., Liu, X., Yin, Z., Liu, Z., Li, X., et al. (2023). Study of urban heat island effect in Hangzhou metropolitan area based on SW-TES algorithm and image dichotomous model. *SAGE Open* 13 (4), 21582440231208851. doi:10.1177/21582440231208851
- Shang, M., and Luo, J. (2021). The Tapio decoupling principle and key strategies for changing factors of Chinese urban carbon footprint based on cloud computing. *Int. J. Environ. Res. Public Health* 18 (4), 2101. doi:10.3390/ijerph18042101
- Shukla, K., Dadheech, N., Kumar, P., and Khare, M. (2021). Regression-based flexible models for photochemical air pollutants in the National Capital Territory of megacity Delhi. *Chemosphere* 272, 129611. doi:10.1016/j.chemosphere.2021.129611
- Sowlat, M. H., Gharibi, H., Yunesian, M., Tayefeh Mahmoudi, M. T., and Lotfi, S. (2011). A novel, fuzzy-based air quality index (FAQI) for air quality assessment. *Atmos. Environ.* 45, 2050–2059. doi:10.1016/j.atmosenv.2011.01.060
- Sun, Q., Zhang, T., Wang, X., and Chen, Z. (2022). Air quality evaluation method based on progressive DS evidence theory. *J. Beijing For. Univ.* 44, 119–128.
- Sun, Q., Zhang, T., Wang, X., Lin, W., Fong, S., Chen, Z., et al. (2022). An ambient air quality evaluation model based on improved evidence theory. *Sci. Rep.* 12, 5753. doi:10.1038/s41598-022-09344-0
- Tan, Q., Zhuang, H., Tie, Z. X., Ding, C. F., and Wu, Q. (2017). An air quality evaluation model for the industrial park based on the fuzzy matter-element analysis. *J. Saf. Environ.* 17, 2019–2023. doi:10.13637/j.issn.1009-6094.2017.05.076
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46, 234–240. doi:10.2307/143141
- Utku, A., Can, Ü., Kamal, M., Das, N., Cifuentes-Faura, J., and Barut, A. (2023). A long short-term memory-based hybrid model optimized using a genetic algorithm for particulate matter 2.5 prediction. *Atmos. Pollut. Res.* 14, 101836. doi:10.1016/j.apr.2023.101836
- Vaidya, O. S., and Kumar, S. (2006). Analytic hierarchy process: an overview of applications. *Eur. J. Oper. Res.* 169, 1–29. doi:10.1016/j.ejor.2004.04.028
- Wang, J., Luo, Q., and Hu, Y. (2021). Locomotive bearing fault diagnosis method based on KNN-EMD algorithm. *Comput. Simul.* 38, 129–132.
- Wang, W., and Klir, G. J. (2009). *Generalized measure theory*. New York, NY, USA: Springer Science and Business Media.
- Wang, X., Wang, T., Xu, J., Shen, Z., Yang, Y., Chen, A., et al. (2022). Enhanced habitat loss of the Himalayan endemic flora driven by warming-forced upslope tree expansion. *Nat. Ecol. Evol.* 6 (7), 890–899. doi:10.1038/s41559-022-01774-3
- Wang, Y., and Jiang, Y. (2017). Air quality evaluation based on improved CRITIC weighting and fuzzy optimization. *Stat. Decis.* 17, 83–87. doi:10.13546/j.cnki.tjyc.2017.17.019
- Wu, B., Gu, Q., Liu, Z., and Liu, J. (2023). Clustered institutional investors, shared ESG preferences and low-carbon innovation in family firm. *Technol. Forecast. Soc. Change* 194, 122676. doi:10.1016/j.techfore.2023.122676
- Wu, D., and Mendel, J. M. (2009). A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. *Inf. Sci.* 179, 1169–1192. doi:10.1016/j.ins.2008.12.010
- Xiao, H., and Duan, Y. (2013). An improved study on the KNN algorithm based on the attribute value-related distance. *Comput. Sci.* 40, 4.
- Xiong, L., Bai, X., Zhao, C., Li, Y., Tan, Q., Luo, G., et al. (2022). High-resolution data sets for global carbonate and silicate rock weathering carbon sinks and their change trends. *Earth's Future* 10 (8), e2022EF002746. doi:10.1029/2022EF002746
- Yang, Y., Xiang, C. C., and Wei, D. J. (2010). Data cluster based on extension K nearest neighbor algorithm. *Comput. Eng. Appl.* 46, 156–159.
- Zadeh, L. A. (1965). Fuzzy sets. *Inf. Control.* 8, 338–353. doi:10.1016/S0019-9958(65)90241-X
- Zhang, B. (1997). *The fuzzy matter-element analysis*. Beijing, China: Petroleum Industry Press.
- Zhang, S., Bai, X., Zhao, C., Tan, Q., Luo, G., Wang, J., et al. (2021). Global CO₂ consumption by silicate rock chemical weathering: its past and future. *Earth's Future* 9 (5), e1938E–e2020E. doi:10.1029/2020EF001938
- Zhang, Y. L., and Cao, F. (2015). Fine particulate matter (PM_{2.5}) in China at a city level. *Sci. Rep.* 5, 14884. doi:10.1038/srep14884
- Zhong, S., and Liu, X. (2019). A new method to determine basic probability assignment based on interval number. *Computing, Communications and IoT Applications*. Shenzhen, China: ComComAp, 316–320. doi:10.1109/ComComAp46287.2019.9018805