Volume 7, Issue 3, 2025

Print ISSN: 2663-1024

Online ISSN: 2663-1016

EURASIA JOURNAL OF SCIENCE AND TECHNOLOGY

Eurasia Journal of Science and Technology

Volume 7, Issue 3, 2025

Published by Upubscience Publisher

Eurasia Journal of Science and Technology
Print ISSN: 2663-1024 Online ISSN: 2663-1016
Email: info@upubscience.com
Website: http://www.upubscience.com/

Table of Content

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3080

HOTSPOTS AND VISUAL ANALYSIS OF ACADEMIC EARLY WARNING RESEARCH FOR COLLEGE STUDENTS

XiaoPeng Tang, YunLiang Jiang, HaiLong Sun*
School of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China.
Corresponding Author: HaiLong Sun, Email: hlsun@cauc.edu.cn

Abstract: This study used China National Knowledge Infrastructure (CNKI) as the data source and selected 385 research papers on academic warning in universities published between 2005 and 2024 for analysis. The CiteSpace analysis tool was used to visualize the literature, and the results showed that research on academic warning in China began in 2007 and has gone through stages of initial exploration, diversified explosion, and innovation deepening; The research is mainly focused on education and psychology in terms of disciplinary distribution, with computer science as a supplement; Most researchers tend to conduct research independently, with limited collaborative power and only a few small-scale collaborative teams present, resulting in relatively insufficient cooperation between different institutions; The research hotspots focus on academic difficulties, academic crises, and academic assistance, and in recent years, there has been a gradual shift towards using big data and intelligent algorithms for research; The academic warning for college students is mainly clustered into eight core topics: credit system, warning mechanism, assistance mechanism, data mining, student management, countermeasures, students with learning difficulties, and delayed graduation. The research results suggest that future research can expand the scope of disciplines, strengthen cooperation between authors and institutions, discover students with potential academic crises through mining data information and applying artificial intelligence algorithms, improve warning and assistance mechanisms, timely formulate precise tutoring strategies to support and assist students, improve their learning quality, and successfully complete academic tasks. These will become important hot topics of concern for major universities and researchers in the future.
Keywords: Academic warning; Visual analysis; CiteSpace; Knowledge graph; College students

1 INTRODUCTION

In recent years, the severity of academic difficulties faced by college students in the field of higher education has become increasingly prominent. Its manifestation has gradually evolved from a single academic problem to a complex chain that may trigger multidimensional negative effects such as mental health disorders and weakened employment competitiveness[1]. In the current era, the higher education system is undergoing a profound transformation in teaching methods and training models. At the same time, students' learning paradigms and strategies are constantly being restructured. This dual transformation has greatly intensified the complexity and diversity of factors that can trigger academic difficulties. Therefore, higher education institutions are facing unprecedented multidimensional challenges and difficulties in dealing with the academic crisis of college students. Firstly, college students generally exhibit a lack of understanding and mastery of new learning patterns during the early stages of enrollment. There are significant limitations in their understanding of the coherence of the knowledge system and the systematic connections between various disciplines, which leads to some students adopting a last-minute cramming strategy before exams. Secondly, there are several deep-seated problems in the teaching system of universities, such as the aging of professional knowledge systems, the fragmentation of teaching content and the disconnect from the development of emerging disciplines, and the difficulty of traditional teaching models in effectively attracting and meeting the learning needs of contemporary young students. The interweaving of these factors can easily lead to academic difficulties and trigger academic crises. Furthermore, the current academic community's in-depth exploration and theoretical construction of academic crisis issues are still weak. Although some university managers have started exploratory research on academic crisis warning mechanisms, most of these studies are scattered and lack systematic theoretical framework support. A professional, systematic, and scientific academic warning research system has not yet been formed [2-3]. Therefore, how to effectively address the issue of academic warning for college students has become a research hotspot in the field of higher education in recent years. However, unfortunately, there is a lack of systematic review literature on the current situation, hotspots, and future trends of academic warning research for college students in China. It is urgent to comprehensively sort out and deeply analyze the latest research results in this field to guide practice and promote further development of related research.

2 RESEARCH METHODS

2.1 Data Sources

This study used the Chinese Journal Full text Database (CNKI) as the data source and employed advanced retrieval strategies to precisely define the scope of literature sources. To ensure the reliability and accuracy of research conclusions, this study has limited the time span of literature search to between 2005 and 2024. By conducting targeted searches by topic, 601 relevant literature were initially obtained. Subsequently, multiple rounds of screening were conducted to remove 216 articles with low relevance to the research topic, while preserving 385 highly relevant and effective literature samples. These literature information were subsequently organized and exported in a format compatible with Refworks literature management software, forming the data foundation and core source of this study.

2.2 Research Methods and Tools

CiteSpace is a visualization analysis software based on Java language [4]. It analyzes the potential knowledge hidden in literature data through visualization methods, presenting the structure, rules, and distribution of scientific knowledge [5]. It demonstrates strong superiority and accuracy in analyzing and predicting the information panorama, hot topics, research frontiers, and development history of a certain discipline and technology field [6]. Compared with traditional literature review and analysis, this bibliometric method shows the advantages of comprehensive and objective data [7]. Therefore, this article mainly uses CiteSpace 6.2.R3 information visualization analysis software and Excel to visually analyze the Chinese literature of academic journals on academic warning research included in CNKI, in order to grasp the publication volume, disciplinary distribution, research hotspots, frontiers and trends of domestic academic warning research, and provide theoretical basis for related research.

3 DATA ANALYSIS AND GRAPH INTERPRETATION

3.1 Basic Information on Publication

The dynamic changes in the number of publications serve as a direct indicator of the development trend and direction of a specific research field, and their evolutionary trajectory profoundly reveals the evolution process of that field. Figure 1 shows the annual distribution trend of research results on academic warning for college students. Through this figure, it can be systematically analyzed that from 2005 to 2014, although there were a small number of research outputs, the overall quantity was limited, indicating that academic warning research was still in the preliminary exploration and foundation stage. The research results during this period laid the foundation for further in-depth research; From 2015 to 2020, research activities have significantly increased, not only reflected in the continuous growth of the number of publications, but also in the extensive expansion of research fields and the deepening of research content. This transformation demonstrates that the field has attracted widespread academic attention and become a research hotspot. However, due to the comprehensive influence of various uncertain factors, the growth path of the number of publications in this stage showed a certain degree of volatility, reaching the peak of research activities in 2020, marking the outbreak of diversified and high-speed development in academic warning research; Entering the period from 2021 to 2024, although the number of publications has shown a slight downward trend, it still remains at a high level and tends to stabilize. This trend reflects the sustained interest and high attention of the academic community in the research of academic warning for college students, indicating that the field is entering a stable and in-depth development stage, namely the stable deepening period.

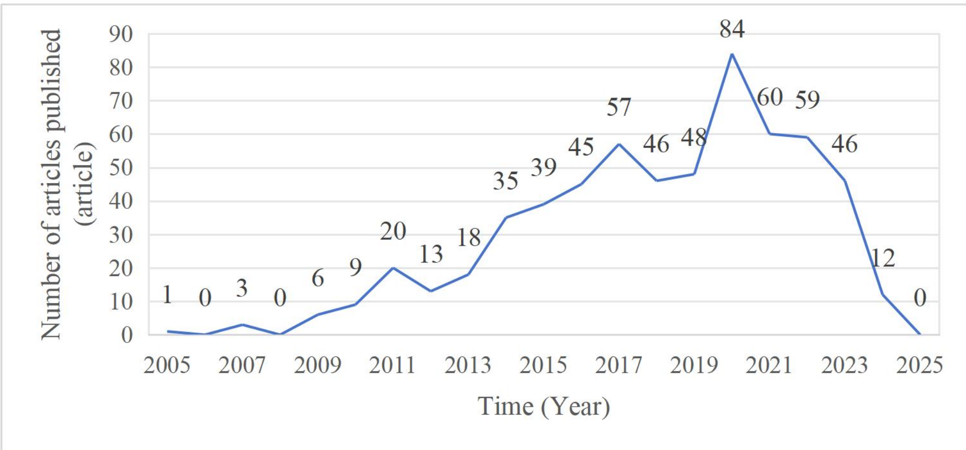


Figure 1 Distribution of Academic Early Warning Results for College Students

3.2 Visualization Analysis based on CiteSpace

This article aims to use the visualization graph analysis function of CiteSpace software platform to conduct a comprehensive and systematic statistical analysis of 385 literature on academic warning research for college students. Through this method, the aim of this study is to explore in depth the collaborative networks among authors in this academic field, the collaborative models of research institutions, and identify and analyze current research hotspots, cutting-edge dynamics, and future development trends, in order to provide scientific basis and reference for further in-depth research in this field.

3.2.1 Analysis of author collaboration graph

Author co-occurrence analysis is a method used to identify core authors in a specific academic field and their level of collaboration. In the diagram, the size of the nodes reflects the number of articles published by the authors, the thickness of the lines represents the density of collaboration between authors, and the depth of the node colors represents the order of publication time. Taking Figure 2 as an example, a network consisting of 365 authors and 183 connections has a density of 0.0028. The figure shows four significant large nodes, namely Zhang Dajun, Liu Guirong, Li Meijuan, and Zhou Dongbin. According to Price's law, these authors have published more than or equal to three articles and are therefore recognized as high-yield authors. The analysis results show that only these four authors meet the criteria for high productivity authors. The figure also shows many collaborative author groups and some small collaborative teams, with only one connection between their nodes, indicating that the collaborative network is not fully developed and the collaborative power is relatively limited. At the same time, core authors such as Zhang Dajun, Liu Guirong, Li Meijuan, and Zhou Dongbin began to engage in research on academic warning for college students earlier, while there are relatively few authors from the emerging forces. Overall, the diagram shows many relatively isolated nodes with a lack of network connections between various groups, especially emerging authors such as Chen Heng, Hou Zhoubo, Zhao Jing, and Zhang Wei, who lack close academic collaboration in this field. In summary, the author co-occurrence network shown in Figure 2 reveals the formation and distribution of the core author group in the field of academic warning research for college students, and also points out the dispersion and limitations in current research cooperation, providing directional inspiration for future research cooperation and resource integration.

Figure 2 Collaborative Graph of Research Authors

3.2.2 Analysis of academic influence of issuing institutions

By publishing academic papers and participating in academic seminars, the author delves into specific disciplinary fields. As a result, the research scope and academic achievements of academic institutions are largely reflected in the quantity and quality of their published academic papers. According to the graphical data analysis, the top five institutions in terms of the number of papers issued include the Institute of Developmental Psychology of Beijing Normal University, the School of Psychology of Shandong Normal University, the Institute of Psychology of the Chinese Academy of Sciences, the Department of Education of Beijing Normal University and the School of Psychology of Huazhong Normal University. The CiteSpace software reflects the research hotspots in related fields through node size and centrality. The number of articles published by an institution is intuitively displayed by node size, and the larger the node volume, the more articles published by the institution; Centrality serves as a yardstick for evaluating the importance of nodes, reflecting the core position and influence of institutions in the research field. We analyzed the publishing institutions in the literature using the Institution module of CiteSpace and set a threshold of 3. Figure 3 shows a network structure with 307 nodes and 67 connections, with a density of 0.0014, revealing the degree of interconnectivity between institutions. The scale of nodes directly corresponds to the publication contribution of institutions, and the larger the scale, the richer the research results; Centrality quantifies the key roles and influence of institutions in the overall research network. Based on the visualization analysis of 385 literature, this study identified the main research institutions in the field of academic warning for college students, and visually displayed the cooperation status between institutions through connections between nodes. In terms of the composition of publishing institutions, a representative publishing cluster has been formed with well-known domestic normal universities such as Beijing Normal University, Shandong Normal University, and Central China Normal University as the core. Regional analysis further indicates that China's research in this field presents a research pattern dominated by teacher training universities, supplemented by research institutes and comprehensive universities. The dominant position of universities as the main force of scientific research, occupying four out of the top five institutions in terms of publication volume, has been fully confirmed. It is worth noting that although there are relatively concentrated cooperation networks, overall, the cooperation networks between institutions are relatively scattered and have not yet formed significant cluster effects. The leading institutions in the research field are mainly concentrated in the education and psychology related fields of a few normal universities. This also means that in future research in this field, strengthening cooperation and communication among institutions, promoting knowledge sharing and resource integration, is of great significance for promoting the in-depth development of research in this field, and there is enormous space and potential for cooperation.

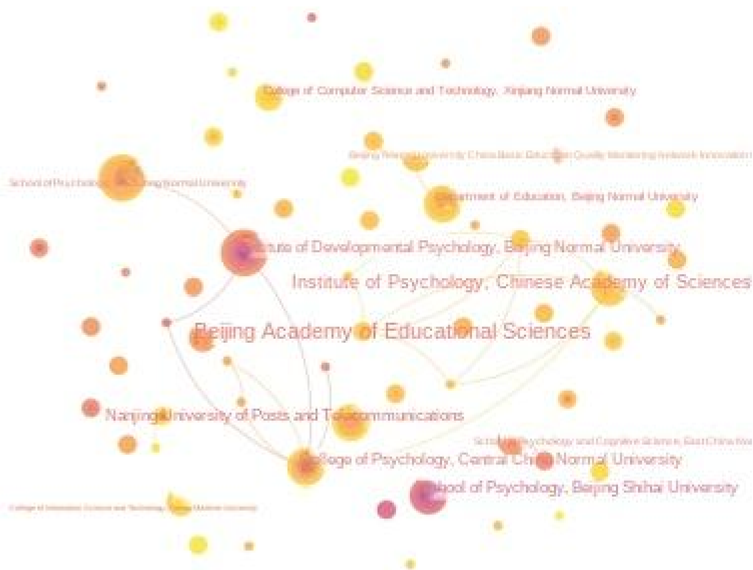


Figure 3 Relationship Map of Collaborative Institutions in the Field of College Student Academic Early Warning

3.2.3 Research hotspot analysis

The CiteSpace tool was used for in-depth analysis and processing of keywords. After merging semantically similar words, a keyword co-occurrence knowledge graph was constructed, as shown in Figure 4. In this graph, the term 'academic warning' shows the highest frequency, with the most significant node size, and runs through the entire research period. However, given that this study has clearly adopted the core retrieval strategy of "college students+academic warning" in the topic keyword retrieval stage, the keyword "academic warning" will no longer be the focus of discussion in this analysis framework. In addition, considering that "college students" are the subject of analysis and naturally situated in the learning environment of "universities", the term "universities" is also excluded in subsequent discussions. After excluding the keywords "college students", "academic warning", and "universities", this study sorted and screened the remaining keywords based on two key indicators: frequency and centrality. Finally, the keywords ranked in the top five in terms of frequency and centrality were selected and sorted according to these two indicators. The results are shown in Table 1 (based on frequency sorting) and Table 2 (based on centrality sorting) for further in-depth analysis of the importance and influence of these keywords in the research field.

Table 1 High Frequency Keywords for Academic Warning of Chinese College Students

Hot words	frequency	Center coefficient
Academic difficulties	34	0.2
Academic assistance	26	0.05
data mining	21	0.15
credit system	20	0.37
Academic crisis	17	0.21

Table 2 Key Words of Academic Warning Center Degree for Chinese College Students

Hot words	frequency	Center coefficient
Warning mechanism	17	0.55
school tradition construction	10	0.41
credit system	20	0.37
Academic crisis	17	0.21
Academic difficulties	34	0.2

Through in-depth analysis of Tables 1 and 2, it can be seen that high-frequency keywords include "academic difficulties," "academic assistance," "data mining," "credit system," and "academic crisis. However, not all high-frequency keywords have high centrality, and relying on high-frequency keywords cannot accurately determine research hotspots. In the CiteSpace software framework, the centrality of keywords is a key indicator for measuring their importance in the knowledge graph. When the centrality threshold is set to ≥ 0.1, such keywords are considered a turning point in the network structure and often accurately reflect current research hotspots in the field. From the perspective of centrality analysis, the "early warning mechanism" and "academic atmosphere construction" demonstrate extremely high network support, with the "early warning mechanism" having a centrality of 0.55, significantly enhancing the overall architecture stability of the network and becoming the core support point; The centrality of "academic atmosphere construction" is 0.41, which also has a significant impact on the network structure. Both not only lay a solid foundation for the entire knowledge network, but also clearly point to the main research hotspots at present. Therefore, "warning mechanism" and "academic atmosphere construction" are the main research hotspots, followed by "credit system", "academic crisis" and "academic difficulties".

Figure 4 shows a graph of keyword co-occurrence and centrality analysis, which intuitively reflects the frequency of keyword co-occurrence through the size of nodes, while the thickness of the connections between nodes quantifies the centrality of keywords in the network. Centrality, as an indicator of a node's ability to serve as a medium for information transmission or connection in a network, the larger its value, the more critical its position in the overall network structure [8]. It is generally believed that nodes with centrality exceeding 0.1 are considered important nodes in network analysis, playing a more central role in information transmission and relationship building. According to the statistical results of the software used, the keywords with centrality exceeding 0.1 are: academic warning (207), warning mechanism (17), academic atmosphere construction (10), credit system (20), academic crisis (17), academic difficulties (34), student academic performance (2), college students (45), data mining (21), students with learning difficulties (16), influencing factors (6), assistance (12), universities (31), and talent cultivation (4). The specific data can be found in Table 3. Among them, "academic warning" as the largest node in the graph indicates that it occupies the highest media centrality position in the network structure of the entire academic warning research field, and is the core connecting other keywords and research fields. Next are keywords such as "warning mechanism", "academic atmosphere construction", "credit system", "academic crisis", and "academic difficulties". These keywords are represented by larger nodes and thicker connections in the graph, indicating that they play a significant role in connecting and mediating in the relationship network, and together constitute the current hotspots and focuses of academic warning research.

Table 3 Keywords with Center>=0.1

Count	Centrality	Year	Keyword	Count	Centrality	Year	Keyword
207	0.98	2007	Academic warning	45	0.17	2009	college student
17	0.55	2007	Warning mechanism	21	0.15	2009	data mining
10	0.41	2011	school tradition construction	16	0.14	2007	Students with learning difficulties
20	0.37	2007	credit system	6	0.14	2014	influence factor
17	0.21	2010	Academic crisis	14	0.12	2013	help
34	0.20	2009	Academic difficulties	31	0.11	2011	colleges and universities
2	0.18	2007	Student academic performance	4	0.10	2017	talent training

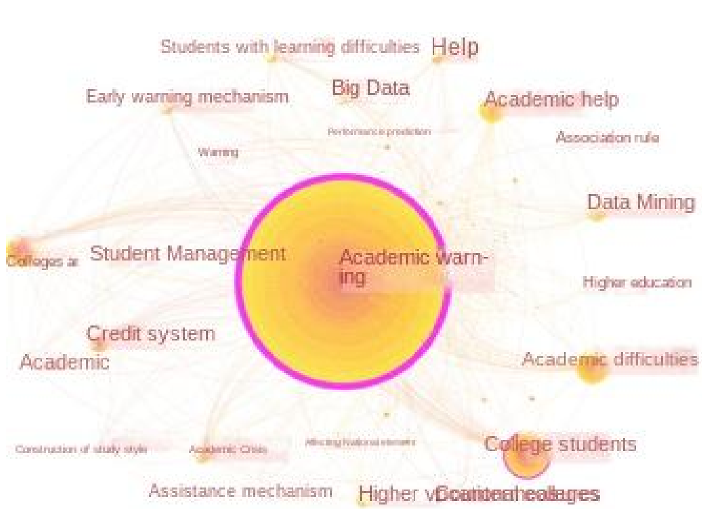


Figure 4 Keyword Co-occurrence Map

3.2.4 Cluster analysis

This study used a clustering analysis method based on the Log Likelihood Ratio (LLR) algorithm to systematically classify and organize keywords related to academic warning applications, aiming to deeply analyze the research topic categories in this field. Through the clustering function of CiteSpace software, eleven highly cohesive clustering groups were formed, with a modularity index of 0.8154 and an average silhouette coefficient of 0.9805. Both indicators significantly indicate excellent clustering performance, verifying the rationality and accuracy of the classification. These eleven results are: # 0 academic warning, # 1 credit system, # 2 warning mechanism, # 3 college students, # 4 assistance mechanism, # 5 data mining, # 6 student management, # 7 countermeasures, # 8 students with learning difficulties, # 9 universities, and # 10 delayed graduation. Among them, # 0 academic warning includes keywords such as academic difficulties, college students, data mining, and warning mechanisms; #The credit system includes keywords such as counselors, academic issues, local industry universities, and information inquiry; #The warning mechanism includes keywords such as academic studies, academic guidance, assistance, and talent cultivation; #3 keywords for college students, including academic difficulties, causes, suggestions, and academic warnings; #The four assistance mechanisms include keywords such as academic crisis, academic atmosphere construction, comprehensive education, and humanistic management; #Data mining includes keywords such as academic performance, grade prediction, incremental learning, and precise coaching; #6. Student management includes key words such as indicator system, assistance, student status warning, and leadership mechanism; #The 7 strategies include keywords such as academic assistance, big data, problems, and ethnic colleges; #Keywords for 8 students with learning difficulties include secondary departments, student assistance, compassionate aid, and "Double First Class" universities; #9 universities include keywords such as academic warning, academic atmosphere, student status processing, and learning quality; #10 keywords for delayed graduation include warning and assistance, influencing factors, graduate education, and graduate student diversion, see Figure 5.

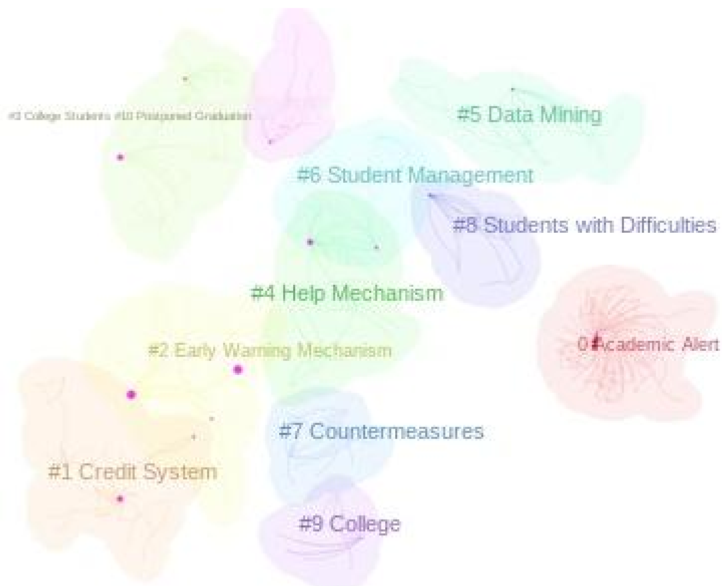


Figure 5 Keyword Clustering Diagram

3.2.5 Development trend analysis

In order to visually display the evolution process and development trend of keywords in different time periods, this study used Time zone visualization technology to process and analyze keywords, aiming to deeply analyze the research characteristics and mainstream research directions of academic warning for college students from 2005 to 2024, and reveal the dynamic evolution rules of this field. Figure 6 shows the overall Time zone view of research on academic warning for college students, where the horizontal axis represents the time axis. Along this dimension, the evolution path of keywords is clearly depicted, intuitively revealing the changing trends in research on academic warning for college students. In the figure, the significant increase in the number of literature during a specific time period reflects the focused attention and in-depth exploration of the field by researchers during that period. The node connections between time periods symbolize the inheritance and continuation of research, and the density of connections intuitively reflects the degree of close connection between research in different time periods. Analysis shows that from 2007 to 2010, research mainly focused on themes such as academic warning, academic difficulties, and academic crises, dedicated to exploring the fundamental causes of academic crises, marking the exploratory stage of research on the causes of academic warning for college students. Subsequently, from 2011 to 2016, research on academic warning and other theories further deepened, with a dense distribution of keywords and a wide range of themes, showing a trend of diversified development. Especially in 2014, the emergence of emerging hotspots such as academic assistance greatly promoted the deepening and development of academic warning research. The increase in high-frequency keyword nodes further proves the depth of research and the wide coverage of topics. Since 2017, the number of nodes in the Time zone graph has increased but the scale has been relatively small. At the same time, the number of literature publications during this period has decreased compared to previous years, reflecting the trend of research towards dispersion and diversification, which is consistent with the characteristics of a stable and deepening stage of research. During this period, researchers conducted more in-depth discussions on a large number of new research topics that emerged earlier. At the same time, the emergence of new research hotspots such as big data and artificial intelligence algorithms indicates that academic warning research is gradually integrating information technology, opening a new chapter in the combination of traditional academic warning research and modern information technology.

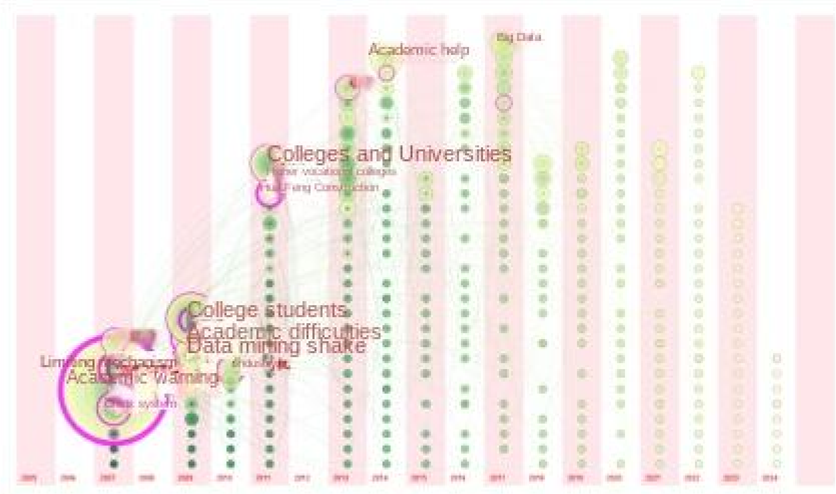


Figure 6 Keyword Time Zone Analysis

Based on a detailed analysis of the research content on academic warning for college students in various time periods, it can be reasonably inferred that by 2024, the exploration and application of artificial intelligence algorithms and big data technology in the field of academic warning research will move towards a deeper level. This trend is mainly reflected in two dimensions: firstly, to continuously deepen exploration, consolidate and expand relevant theoretical foundations within the existing basic research framework; Secondly, it actively utilizes artificial intelligence algorithms to comprehensively collect, efficiently process, and deeply analyze academic performance data of college students, aiming to improve the accuracy and interpretability of academic warning and prediction, and provide solid data support and decision-making basis for universities to formulate and implement personalized and precise student academic guidance strategies. This not only helps to identify and intervene in academic risks in a timely manner, but also effectively promotes students' comprehensive development and academic achievement.

4 SUMMARY AND PROSPECT

Through analyzing the annual publication volume, author institution cooperation, hotspots, frontiers, and development trends of academic warning research literature on college students in the CNKI database from 2005 to 2024, it was found that: firstly, from the perspective of literature publication, 2005-2014 was the initial exploratory stage of academic warning research; The number of studies from 2015 to 2020 continues to increase, and the research fields continue to expand. The peak of research was reached in 2020, which is the explosive period of academic warning research. There is a downward trend from 2021 to 2024, but it still maintains a high level. The stable and deepening period of academic warning research is when the number tends to stabilize. Secondly, from the perspective of collaboration between authors and institutions, the distribution of research institutions on academic warning for college students is relatively scattered, and there is a lack of collaborative research between institutions. Except for a few research teams within each institution, the collaborative relationships between researchers are also relatively loose, and most research institutions and researchers lack sustainability in their research on this topic; Thirdly, from the distribution, clustering, and time zone graph of keyword co-occurrence, research on academic warning focuses on multiple research hotspots such as warning mechanisms, assistance mechanisms, data mining, and big data applications, with a wide range of topics, demonstrating the academic interest of researchers in the field of academic warning.

Therefore, in future academic warning research, expanding the distribution of disciplines and strengthening the connections between disciplines can better explore and discover academic warnings; Strengthen cooperation and communication between authors and institutions to make academic warning research more detailed and sustainable; On the basis of existing research hotspots, explore other hotspots and expand their research topics to further deepen and diversify academic warning research. For example, the deep utilization of information technology such as big data and machine learning intelligent algorithms to deeply explore the rich information contained in the data itself also provides a new paradigm and path for credit based universities to develop talent training programs. Identifying students at risk of academic crisis through data information, improving warning and assistance mechanisms, and developing relevant strategies to support and assist them in a timely manner, further improving the quality of student learning and enabling them to successfully complete academic tasks, will be a key focus of attention for major universities and researchers in the future.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This paper is supported by the Basic Research Funds for the Central Universities Project (Project No. 3122025015).

REFERENCES

[1] Bai Xuejun, Yuan Sheng, Du Xu, et al. Preliminary exploration of warning factors for academic failure among college students. Journal of Tianjin Normal University (Social Sciences Edition), 2022, (01): 70-76.
[2] Wang Li, Wang Hongzhen. Analysis of the Misunderstandings in the System Transplantation of Academic Early Warning System in Chinese Universities with Academic Classification System. Exploration of Higher Education, 2019, (12): 56-60.
[3] He Wei. On the Function of Ideological and Political Education in the Academic Warning Mechanism for College Students. School Party Building and Ideological Education, 2020, (24): 78-80.
[4] Shu Yueyu, Li Huifang, Wang Liling. Current Status and Development Trends of Chinese Psychobiography Research: Knowledge Graph Analysis Based on CiteSpace. Journal of Central China Normal University: Humanities and Social Sciences Edition, 2019, 58(4): 8.
[5] Wu Yue, Li Chaoxu. The Current Status and Development Trends of Spatial Metaphor Research in China: Knowledge Graph Analysis Based on CiteSpace. Psychological Exploration, 2020, 40(4): 8.
[6] Hou Jianhua, Hu Zhigang. Review and Prospect of CiteSpace Software Application Research. Modern Intelligence, 2013, 33(4): 5.
[7] Hu Fu, Ma Jianrong, Qin Boqiang. Analysis of the Scientific Knowledge Graph of Global Water Bloom Research. Scientific Bulletin, 2023, 68(24): 3196-3210.
[8] Zhuang Yuliang, Wang Yuqing. The Current Status and Trends of Digital Supply Chain Research in China: A Bibliometric Analysis Based on CiteSpace. Logistics Technology, 2023, 42(2): 103-109.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3081

AN ATTENTION-DRIVEN BUILDING CLUSTERING APPROACH TAKING SHAPE FEATURES INTO ACCOUNT

ZhiHua Liu
JiangXi University of Science and Technology, Ganzhou 341000, Jiangxi, China.
Corresponding Email: 6120220125@mail.jxust.edu.cn

Abstract: Spatial clustering is the basis of pattern recognition, which is of great significance to map generalization and map updating. Aiming at the problem that many clustering methods can not effectively use the building topology information and it is difficult to deal with high-dimensional data, an attention-driven fusion clustering method is proposed. The model consists of four modules : autoencoder (AE), graph convolutional neural network (GCN), attention-driven fusion module (AFGCN-H) and self-supervised module. AE and GCN are used to extract features from the original data describing the characteristics of buildings and the topological information of the spatial relationship of buildings, respectively. The AFGCN-H adaptively fuses the learning representations of different layers of the two modules. The self-supervised module will optimize the clustering label allocation through corresponding losses, adjust the network parameters, learn the features suitable for clustering, and improve the accuracy of clustering. This paper uses field data sets for clustering analysis, and compares the clustering effects of traditional k-means algorithm, DBSCAN algorithm and MST algorithm. The experimental results show that the proposed method is superior to the traditional spatial clustering algorithm in the results of vector building clustering analysis.
Keywords: Map generalization; Spatial clustering; Building group; Deep clustering

1 INTRODUCTION

Building complex synthesis is an important step in map generalization. Due to the complexity of the geometry and spatial distribution of buildings, it is a complex and difficult operation to synthesize buildings. As the first step of building generalization, spatial clustering can divide buildings into different types of groups according to the similarity and regularity of geometric, semantic and structural features of spatial entities without supervision, so that the objects of the same group are similar to each other. By constructing clustering tasks, it is possible to segment residential data with different characteristics and extract meaningful building spatial distribution patterns, aiming to improve complex building comprehensive tasks.
According to the clustering strategy, clustering algorithms can be divided into seven main types: partition-based algorithms, hierarchy-based algorithms, grid-based algorithms, graph-based algorithms, density-based algorithms, model-based algorithms and hybrid methods. The partition algorithm, such as k-means and k-medoid [1-2], divides the data into k categories, which requires prior expertise to determine k. However, the selection of k is vague in many cases, and such methods are difficult to find clusters of any shape, nor to determine outlier buildings. Hierarchical algorithms, such as CURE and BIRCH [3-4], can handle multi-scale spatial clustering problems, but they need to define termination conditions for them and they cannot find clusters of any shape. Many grid-based algorithms, such as fuzzy clustering and hierarchical clustering [5-6], show the problem of low efficiency and accuracy when the dimension of data is large. Graph-based algorithms, such as ASCDT, Minimum Spanning Tree (MST), and Delaunay triangulation [7-8], can identify buildings of any shape by taking into account the spatial structure and adjacency information of buildings. However, the gestalt principle, the calculation of the characteristic factors describing buildings, and the distribution of building data are not well understood.
Deep neural network is a nonlinear and deep-seated structural network with strong nonlinear transformation ability. It can extract the feature representation of these two types of data at the same time, and fuse the two types of features for more reliable feature extraction. This paper introduces the deep clustering method into the clustering of vector buildings, constructs a graph model for the building group, and abstracts a single building as a graph node. The deep neural network (DNN) and the graph convolutional neural network (GCN) are used to learn the node information and the graph structure information respectively, and the attention-driven method is used to dynamically fuse the features to construct the clustering model[9].

2 ATTENTION-DRIVEN FUSION CLUSTERING MODEL

The overall framework of the attention-driven fusion clustering model proposed in this paper is shown in Figure 1. The model mainly includes four parts: (1)Autoencoder (AE) learns the original geometric features of buildings; (2)The learning representation of geometric features is transferred to the graph convolutional neural network (GCN) to integrate geometric and structural information. (3) The attention mechanism is used to dynamically combine the features learned by each layer of autoencoder with the features learned by GCN; (4) The self-supervised module unifies the AE features and the fused features in a framework, and uses the relationship between them to optimize the clustering label allocation to learn the features suitable for clustering. In order to enable the model to process vector data, it is necessary to construct the adjacency matrix of the original data and calculate the feature factors used to abstract the building as a graph node[10].

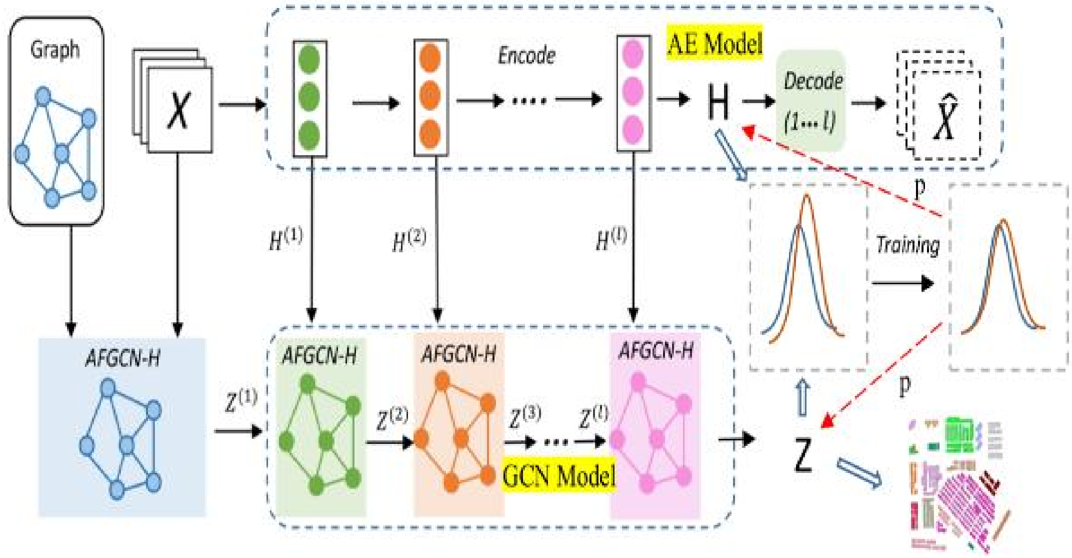


Figure 1 Building Attention-Driven Fusion Clustering Network

In Figure 1, X represents the characteristic matrix of the original input data, $\widehat{X}$ represents the reconstructed data, and Graph represents the adjacency matrix A, l represents the number of layers. $H^{(l)}$ represents the high-dimensional feature learning results of the AE autoencoder. The fusion $Z^{(l)}$ of GCN features and AE features is realized by the AFGCN heterogeneous fusion module (AFGCN-H) constructed by GCN. The network is self-trained by minimizing the KL divergence between the H distribution (orange curve) and the Z distribution (blue curve), and the update of the model is guided by the target distribution P.

2.1 Graph Structure Data

The construction of the graph model of the original building mainly includes obtaining the adjacency relationship and feature extraction between the buildings. The adjacency relationship can be obtained by constructing a constrained Delaunay triangulation. If there are triangles connected between buildings, the adjacency matrix value is set to 1, otherwise it is set to 0. In order to extract the features describing the nodes of the graph, this paper calculates 12 feature factors from the size, shape, direction and density of the building to describe the building. The data are organized into the original data $X \in R^{N\times d}$ and the adjacency matrix $A \in R^{N\times N}$, where N is the number of buildings and d is the number of node features[11].

2.2 Autoencoder Module

Since the composition of the original data is not complicated, this paper uses a basic autoencoder to learn the feature representation of the original data. Assuming that the autoencoder has a total L layer, the features learned at the l layer can be expressed as :

$$H^{(l)} = \emptyset\left(W_e^{(l)} H^{(l-1)} + b_e^l\right) \tag{1}$$

where $\emptyset$ is the activation function, $W_e^{(l)}$ and b_e^l are the weight matrix and bias term of the lth layer of the coding part, and $H^{(0)}$ is the original data X.

The encoded results are decoded by a fully connected network to reconstruct the original data, and finally the decoding result $\widehat{H}^{(l)}$ is calculated. The network structure of the encoder and decoder layers is completely symmetrical. Therefore, the final generated $\widehat{H}^{(l)}$ and $H^{(l)}$ have the same dimension, and their values should be as similar as possible. $\widehat{H}^{(l)}$ is calculated as follows :

$$\widehat{H}^{(l)} = \emptyset\left(W_d^{(l)} H^{(l-1)} + b_d^l\right) \tag{2}$$

$W_e^{(l)}$ and b_e^l represent the weight matrix and bias term of the layer l of the decoding part. The original data and reconstructed data can calculate the loss function :

$$\mathcal{L}_{res} = \frac{1}{2N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2 \tag{3}$$

2.3 Graph Convolutional Neural Network Module

The autoencoder can learn useful attribute feature information from the original data X, but this may ignore the topology information between building nodes.
The characteristic matrix X of the nodes in the graph and the adjacency matrix A representing the topological relationship between the nodes are used as the input of the GCN network. The GCN layer can be represented by the following formula :

$$Z^{l+1} = \emptyset\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}Z^{l}W^{l}\right) \tag{4}$$

Where $\widetilde{A} = A + I_N$ is the adjacency matrix of a self-connected undirected graph, I _ N is the identity matrix, $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ii}$ degree matrix, W^l is the trainable weighted matrix of each layer, $Z^l \in R^{N*D}$ represents the input of GCN at the l-layer, Z^0=X.
The GCN module is used to construct an automatic encoder to learn the spatial structure characteristics of the data in the encoding stage. The whole process can be summarized by the following formula :
$Z = \mathrm{GCN}(X, A)\#(5)$
$\mathrm{GCN}(\bullet)$ represents the graph convolution operation. In this module, the structural information Z^{l+1} learned by each layer of GCN can be obtained. By integrating the heterogeneous fusion module with the original node data reconstructed by AE, a more reliable and more distinguishable clustering data representation that can adapt to two different information can be obtained.

3 EXPERIMENTAL DATA AND PRETREATMENT

3.1 Experimental Region

In this paper, a relatively regular area of a street in Miami, the United States, is selected as the experimental area, with a total of 3234 buildings. This area contains clusters of different building modes, including linear mode, curve mode and grid mode. The distribution pattern of buildings is relatively rich. Through the experimental analysis of this area, the clustering performance of the model can be well tested, as shown in Figure 2.

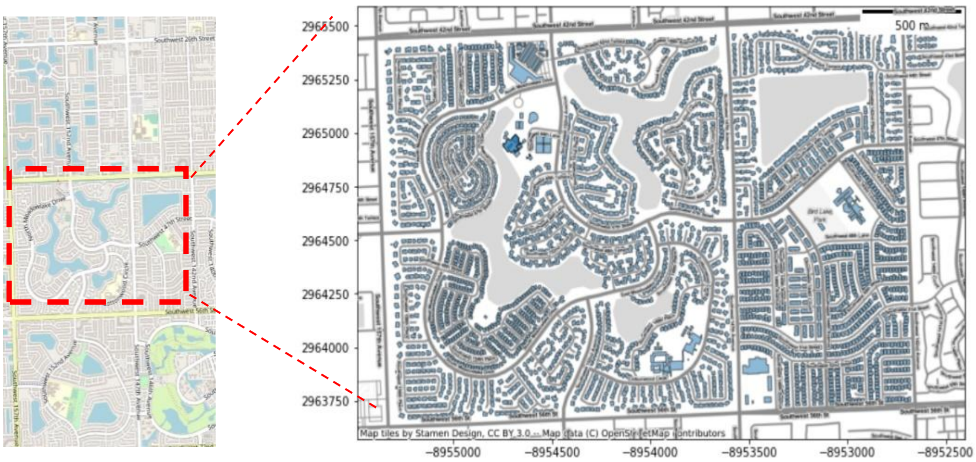
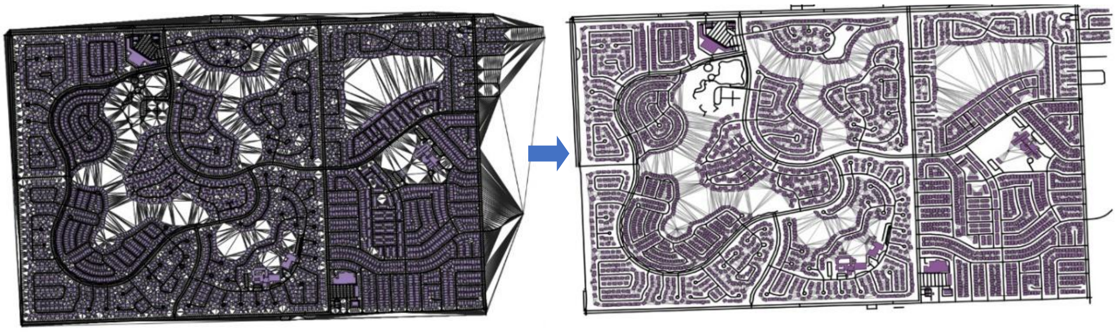


Figure 2 The Experimental Are

3.2 Construction of Geometric Graph Model of Building Group

Deep learning for vector building data should first build a geometric model of vector building data, and build a deep clustering model based on graph structure. An undirected graph can be defined as $G = (v, \varepsilon)$, where v and ε represent the adjacency relationship between the nodes of the graph and the two buildings, respectively. The construction of v will be discussed in the next section. In this paper, a constrained Delaunay triangulation (DT) subdivision is used to obtain its adjacency relationship and establish graph structure data for deep clustering. Firstly, the buildings are physically divided according to the road, and the buildings and roads are encrypted with identifiers at a distance of 10 meters, and the Delaunay triangulation is constructed according to the encrypted points, as shown in Figure 3a ; then delete the triangles related to the road encryption points according to the identifier, and remove the narrow and self-connected triangles to obtain an independent Delaunay triangulation of each block, as shown in Figure 3b; finally, according to the judgment of the first-order adjacent relationship of the buildings directly connected with the triangle, the adjacent matrix A of the building group is generated.
In the process of map generalization, the priority of roads is higher than that of general buildings. If two buildings are separated by roads in solid A, its value is 0. The adjacency matrix constructed in this way is more in line with human cognition and ensures efficiency.

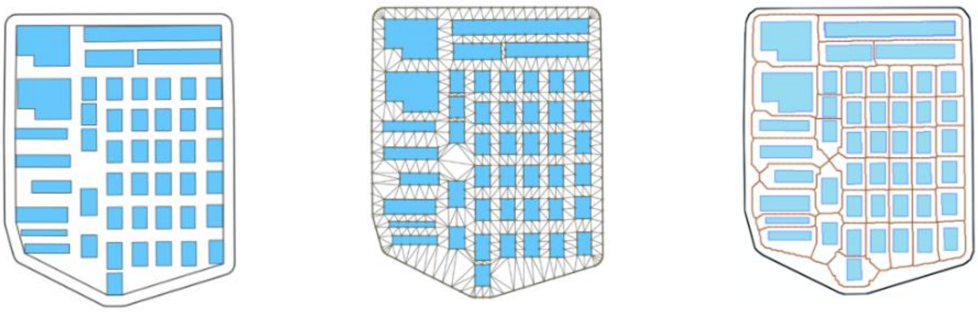


(a) (b)

Figure 3 Building Proximity: (a) Overall Delaunay triangulation; (b) Constrained integral Delaunay triangulation

3.3 Building Similarity Feature Factor

In order to construct the original data X, it is necessary to abstract the original building as a graph node. In order to better describe the geometric attribute information of the building, according to the principle of similarity and continuity, this paper constructs a quantitative model describing the building from the aspects of size, shape, direction and position of the building, so as to refine the grouping. The position is represented by the arithmetic mean of the building vertex. The size is described by the perimeter, area and the average distance from each vertex of the building to its centroid. According to the experimental results of Reference [12], the compactness, fractal dimension, concavity, perpendicularity and perimeter index are used to describe the shape of the building. The direction is represented by the minimum bounding rectangle (SBR) direction of the building polygon. Building density We divide the building group by DT, and generate the Voronoi like graph for the generated triangle according to the method of Reference [13].Each Voronoi like region in the graph is regarded as the affected area of the building, and its density is the ratio of the area of the building to the area of the affected area. In order to make the boundary buildings can also generate a completely surrounded Voronoi like region, we calculate its value according to the process of Figure 4, and the convex of the building group.



(a) (b) (c)

Figure 4 Calculation of Building Density Factor: (a) Convex hull buffer; (b) Construction constraint DT; (c) Generate a Voronoi-like diagram

The specific indicators are shown in Table 1. The node feature representation of the graph is calculated in the following way, that is, the feature matrix X. The subscript meanings involved in some of them are as follows : building (b), building convex hull (CH), and building minimum circumscribed matrix (SBR). In this paper, a total of 13 characteristic factors are calculated. In addition to the adjacent distance weighted to the adjacency matrix, the remaining factors are organized in the matrix form of $X \in R^{N \times d}$, N is the number of buildings, d = 12.

Table 1 Building Characteristic Parameters

parameter	Description index	formula	definition
position	centric position	$(x,y)=(\sum_{i=1}^{n}(x_i, y_i))/n$	The arithmetic mean of all vertex coordinates
size	area	A_b	area
	perimeter	P_b	The perimeter of each side and
	mean radius	$\overline{R} = \frac{1}{N}\sum_{i=1}^{n} R_i$	The average distance from the building vertex to the center position

shape	Compactness	$\mathrm{CI}=\frac{A_{pn}}{A_{EPC}}=\frac{4\pi A_b}{P_b^2}$	The area deviation between polygon and its isoperimetric circle
	fractal dimension	$\mathrm{FR}=1-\frac{\log A_b}{2*\log P_b}$	Measuring edge roughness or smoothness
	concavity	$\mathrm{CNV}=\frac{A_b}{A_{CH}}$	The area deviation between the polygon and its convex hull is used to reveal the extent to which the polygon bends inward or outward.
	verticality	$\mathrm{REC}=\frac{A_b}{A_{SBR}}$	The area deviation between the polygon area and the SBR can reveal the degree of inward bending of the polygon.
	Perimeter index	$\mathrm{nPI}=\frac{P_{EAC}}{P_b}=\frac{2\sqrt{\pi A_b}}{P_b}$	Considering the compactness of polygon boundary
direction	SBR	$\alpha=\arctan\frac{x_i-y_i}{x_j-y_j}$	The longest edge direction of the minimum circumscribed rectangle
density	area ratio	$\mathrm{D}=\frac{S_B}{S_A}$	Deviation between building area and its affected area area
distance	visual distance	$D=\sum_{i=0}^{n}\frac{\Vert P_iP_{i+1}\Vert}{l}\Vert D_{i1}D_{i2}\Vert$	Triangle weighted distance, see [14]

4 EXPERIMENTS AND RESULT ANALYSIS

Through the above data preprocessing, the adjacency matrix A and the original data X can be obtained, and the number of categories K is determined by the number of blocks divided by the road. The original data is pre-trained by AE for 30 rounds to obtain the clustering centroid, and then the main model is used for 200 iterative training. The autoencoder size is set to : D-128-128-256-5, the graph convolution layer is set to : 128-128-256, and the average result of 10 experiments is selected as the final clustering result, as shown in Figure 5a.

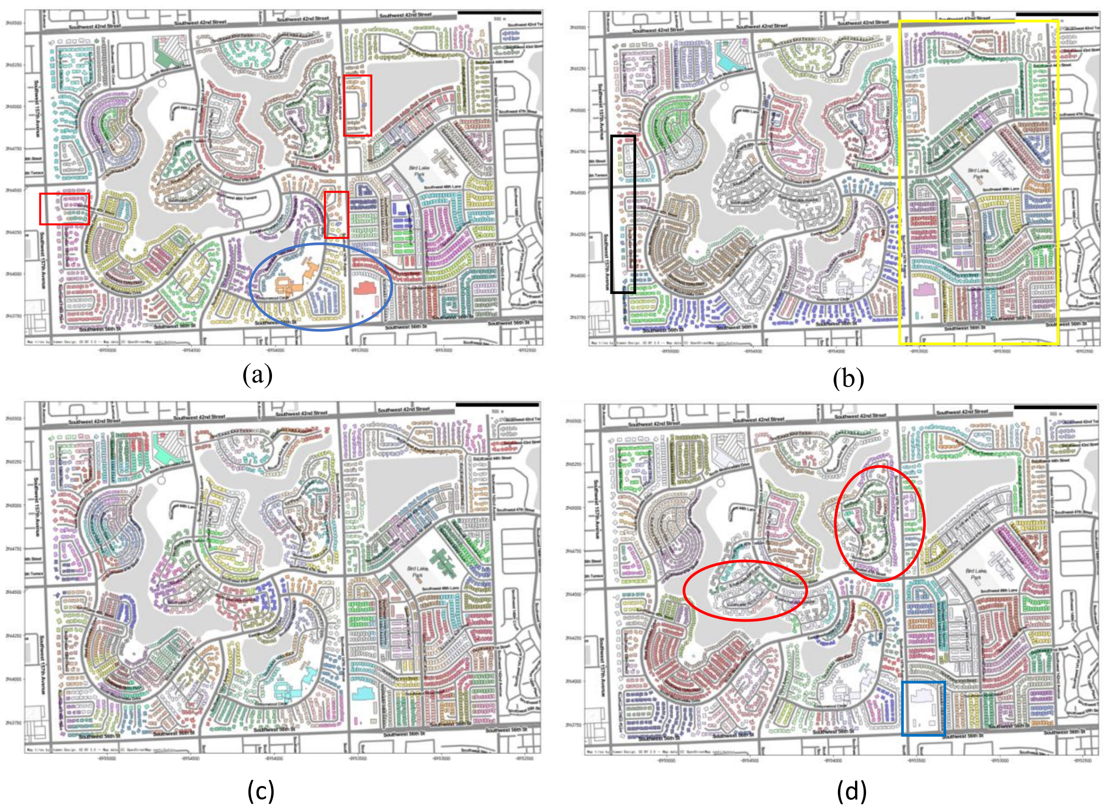


Figure 5 Clustering Results of Four Methods: (a) Proposed method; (b)mst; (c)k-means; (d)DBSCAN

There are 3234 buildings in the experimental area, and the results of the four clustering algorithms are shown in Figure 5. In Figure 5a, the accuracy of building division and human cognition can reach about 70 %. The whole experiment does not need too much parameter adjustment, and it is a complete end-to-end process from input to output. It can be seen that in the area formed by the road, the division effect of the building is ideal. For the linear pattern and the grid pattern distributed along the street, the deep clustering model can correctly identify (the lower right corner and the middle area), reflecting the importance of the topological information provided by the GCN module. If the established adjacency matrix is not segmented by the road, the clustering effect of the building groups arranged along the road and similar to the geometric properties of the building will be poor. In the scene of uniform and uneven building density, the clustering effect of the deep clustering model is also mild. This is due to the fusion of the feature representation of the two modules. The model can correctly classify the building groups according to the adjacency information and geometric information, as shown in the blue circle in Figure 5a.

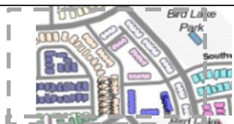
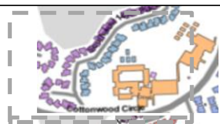
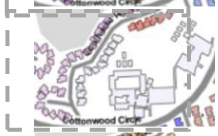
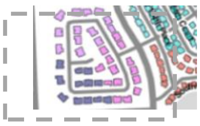
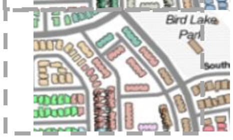
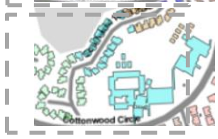
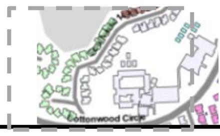
In addition, since the shape of the whole building does not change abruptly, one or two abrupt shapes in the cluster may be separated from the buildings that should belong to the same cluster. Therefore, the effective learning of the above building metrics is very important, as shown in the red rectangle. In general, the algorithm in this paper is valuable for building classification of large data sets, and the local effects are shown in Table 2.

Figure 5b describes the clustering effect of MST. It also uses DT to divide its adjacency relationship, and determines its optimal division threshold according to the dispersion. The specific implementation of the following three clustering algorithms is described in detail in Reference [14]. It can be seen that in the yellow rectangle, the division effect of MST is ideal, and buildings of any shape and density can be found. However, in many other areas, such as black rectangles, the recognition effect is very poor, and the simple linear pattern also has extreme errors. The reason is that the division threshold is a global parameter, which cannot adapt to each area of the whole, and the local area is more obvious. How to determine a reasonable ' pruning threshold ' has always been a problem faced by MST algorithm.

Figure 5c is the clustering effect of k-means. As a comparison, the k-means algorithm deals with the original building feature data, and cannot consider the adjacency topological relationship between buildings. The determination of its K value is the same as the method mentioned above. In the experimental process, k-means clustering has the highest efficiency, but it can be seen from the figure that its overall clustering effect is poor, especially for groups with linear patterns, such as buildings arranged along the road. In general, the shape of the cluster is similar to a sphere, and it is difficult to adapt to clusters of any shape and any density.

Figure 5d is the clustering effect of DBSCAN.Since the algorithm needs to determine the neighborhood radius eps and the minimum sample point min _ sample in the neighborhood radius, the optimal selection of the two is more difficult. The change of any party will lead to a sudden change in the number of clusters and the clustering effect. Therefore, this paper uses the cyclic iteration method to determine its optimal division parameters. The value of eps is 0.00031, and the value of min _ sample is 3, which can obtain an ideal effect diagram. In this experiment, the distribution of buildings is relatively uniform, and the density mutation is less. Therefore, the algorithm has achieved good results in the lower right corner of the experimental area, but in the middle area, as shown in the red circle, it should be continuous.

Table 2 Experimental Local Region Comparison

Clustering Algorithms	Region A	Region B	Region C
Methods			
K-means			
MST			
DBSCAN			

5 CONCLUSION

In this paper, deep learning is introduced to process vector data. The deep clustering model is built by using autoencoder and graph convolutional neural network. The attention mechanism is used to dynamically fuse the features learned by the two modules. The self-supervised module is constructed by student t distribution so that the whole clustering process is an end-to-end process. The experimental results show that after fully considering the structural and attribute characteristics of buildings, the deep clustering method can be better used for building clustering operations. Compared with traditional clustering algorithms, it has more advantages. It can not only deal with high-dimensional data, but also comprehensively consider the topological and geometric information of buildings.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Tardioli G, Kerrigan R, Oates M, et al. Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach. Building and Environment, 2018, 140: 90–106.

[2] Vinod H D. Integer programming and the theory of grouping. Journal of the American Statistical Association, 1969, 64(326): 506–519.

[3] Guha S, Rastogi R, Shim K. CURE: An efficient clustering algorithm for large databases. ACM Sigmod Record, 1998, 27(2): 73–84.

[4] Zhu Y, Jiang Y. Optimization and parallelization of BIRCH clustering algorithm. Computer Engineering and Design, 2007, 28(18): 4345–4346, 4369.

[5] Li S, Zhang H, Wang L, et al. Meteorological drought risk zoning and analysis of winter wheat in Henan Province based on grid and fuzzy clustering. Journal of Henan Agricultural Sciences, 2020, 49(11): 172–180.

[6] Tang G, Xing C, Zhu L, et al. Rapid classification of moving surface fitting filter algorithm using hierarchical clustering. Engineering of Surveying and Mapping, 2021, 30(3): 32–40.

[7] Deng M, Liu Q, Cheng T, et al. An adaptive spatial clustering algorithm based on Delaunay triangulation. Computers, Environment and Urban Systems, 2011, 35(4): 320–332.

[8] Cai J. MST clustering of urban buildings considering Gestalt principles. Chang'an University, 2018. (Doctoral dissertation)

[9] Gong X, Fang Y. Comparative study on adaptability of settlement clustering analysis algorithms. Engineering of Surveying and Mapping, 2020, 29(5): 1–7.

[10] Pilehforooshha P, Karimi M. A local adaptive density-based algorithm for clustering polygonal buildings in urban block polygons. Geocarto International, 2020, 35(2): 141–167.

[11] Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9(Nov): 2579–2605.

[12] Basaraner M, Cetinkaya S. Performance of shape indices and classification schemes for characterising perceptual shape complexity of building footprints in GIS. International Journal of Geographical Information Science, 2017, 31(10): 1952–1977.

[13] Ai T, Zhang X, Zhou Q, et al. A vector field model to handle the displacement of multiple conflicts in building generalization. International Journal of Geographical Information Science, 2015, 29(8): 1310–1331.

[14] Meng N, Feng J, Jia Y. Comparative analysis of areal settlement clustering methods. Geomatics and Information Science of Wuhan University, 2022, 47(10): 1–5.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3082

ANALYSIS OF FIRE RISK PREVENTION AND CONTROL AND FIRE PROTECTION DESIGN OPTIMIZATION OF CULTURAL RELICS BUILDINGS AT OLD SITES

LiTong Ma, Bo Ma*
Yinchuan University of Energy, Yinchuan 750001, Ningxia, China.
Corresponding Author: Bo Ma, Email: 1192897898@qq.com

Abstract: The beams and pillars, doors and windows, purlins, and other components inside the heritage buildings at the old site are pure wood structures, and the texture of their wood is usually low in moisture content, high in flammability, and insufficient in fire resistance rating, which will cause irreversible and destructive damage to the heritage buildings in the event of fire. In addition, the close spacing between heritage buildings, heritage buildings, and neighboring civil buildings, and between heritage buildings and neighboring trees further aggravates the fire risk and poses a serious threat to the fire safety of heritage buildings. This paper addresses the problem of imperfect fire protection facilities inside heritage buildings. It systematically proposes a comprehensive solution of fire protection investigation, fire protection design, and fire protection installation measures, aiming to minimize the risk of fire damage to heritage buildings.
Keywords: Heritage buildings; Fire investigation; Fire design; Fire installation measures

1 INTRODUCTION

A former site is located in the southeast of Wangcang County, Qinglong Mountain hillside, a typical Buddhistcourtyard-style building, built according to the mountain, sitting south to north. According to historical records,the temple was built in the Liang Dynasty of the Southern Dynasties, destroyed in the Sui Dynasty, and rebuilt in the Tang Dynasty, with the imperial name of “Guangfu Yuan”, which was approved by the State Council in October 2019 to be included in the eighth batch of national key cultural relics protection units. The former site has outstanding spiritual symbol value, cultural heritage value and social function value, and is an important historical cultural heritage and spiritual culture carrier in China. Statistics show that since the founding of the country has occurred in more than a hundred major ancient building fire accidents, the overall fire safety situation is grim. If a fire accident occurs, it will lead to irreversible damage to the cultural relics of the building itself. Therefore, to carry out a systematic fire risk assessment and configuration of adaptive fire protection facilities can effectively identify and eliminate fire hazards, significantly improve the safety of cultural relics and buildings, and at the same time for the construction of scientific, standardized fire protection system for cultural relics and buildings to provide a theoretical basis and practical guidance.

2 PROJECT SURVEY

The project management team of the construction unit joins hands with the professional and technical personnel of the survey and design unit to carry out multi-dimensional on-site investigation (including 3D laser scanning and mapping, point cloud data collection and field research on the current protection status). With the support of the construction unit, the system integrates multifaceted information such as architectural heritage archives, GIS data and management logs. The specific implementation includes:
(1) Using panoramic photogrammetry technology to construct a digital twin model of the architectural complex, and systematically sorting out the evolution of the construction techniques and the current conservation status;
(2) Quantitatively evaluating the fire load density and structural parameters of the building body through thermal imaging and material sampling and testing, and establishing a database of fire resistance limits of wooden components;
(3) Based on the current fire code, verify the compliance of the width of the fire separation zone, the clear height of the fire escape, the integrity of the fire partition and the accessibility of the evacuation path;
(4) Systematically investigate the effectiveness of fire-fighting resource allocation, focusing on analyzing technical parameters such as radiation radius of fire stations, multi-system linkage mechanism of control rooms, carrying capacity of heavy equipment at rescue sites and minimum turning radius of passages;
(5) Establishing technical files of fire protection facilities, covering such indicators as pressure nodes of water supply system pipe networks, topological logic of automatic fire alarm systems, load balance of distribution lines and minimum horizontal illumination of emergency lighting;
(6) Construct a fire risk source spatial database, integrating multi-source data such as heat map of combustible distribution, insulation aging coefficient of electrical lines, and probability model of lightning strike disaster. Eventually form a multi-dimensional survey data matrix for subsequent BIM collaborative design to provide spatial topological relationships, material property parameters and risk quantification indicators and other basic data support.

3 FIRE-FIGHTING EQUIPMENT

Based on the concept of preventive protection of cultural relics, this study builds an intelligent fire protection system that integrates multi-level defense architecture and intelligent response mechanism. Through the human-machine synergy protection system (Human-Machine Synergy System, H-M-S system) to realize the full cycle of fire risk control, the formation of "prevention - inhibition - rescue" trinity of prevention and control mode. The system design strictly follows the requirements of the Technical Code for Fire Protection of Cultural Relics Buildings (GB/T51427)and the Code for the Design of Automatic Fire Alarm Systems(GB50116), and focuses on the following core modules:
(1) IOT perception layer: deployment of detection terminals with heritage applicability, including:
Distributed fiber optic temperature sensing fire detector(sensitivity ± 0.5 ℃)
Suction type very early smoke detection device (detection particle size 0.001-10μm)
Three-dimensional panoramic visualization of emergency evacuation instruction system (response time ≤ 0.5s)
(2) Intelligent disposal layer:
High-pressure water mist intelligent linkage operating device (working pressure 10-14MPa, droplet diameter Dv0.99<400μm)
Special fire extinguishing agent proportioning optimization system for cultural relics (based on material compatibility test)
Multi-modal emergency lighting cluster (illumination gradient adjustment range 5-500lx)
(3) Management decision-making layer:
BIM-based fire monitoring system topology architecture (see Figure 1), integrating building information modeling and real-time monitoring data
Fire risk situational awareness algorithm (integrating LSTM neural network and Bayesian network)
Digital twin emergency rehearsal platform (spatial and temporal resolution up to $0.1m^3/0.1s$)
The system realizes equipment interconnection through LoRaWAN communication protocol, builds a special IOT supervision platform for fire protection of cultural relics buildings, and forms a closed-loop management mechanism of "monitoring-warning-disposal-assessment". After testing, the system false alarm rate is ≤0.3 times/year, the emergency response time is 62% shorter than the traditional system, and the integrity of the protection of cultural relics reaches 99.8%.

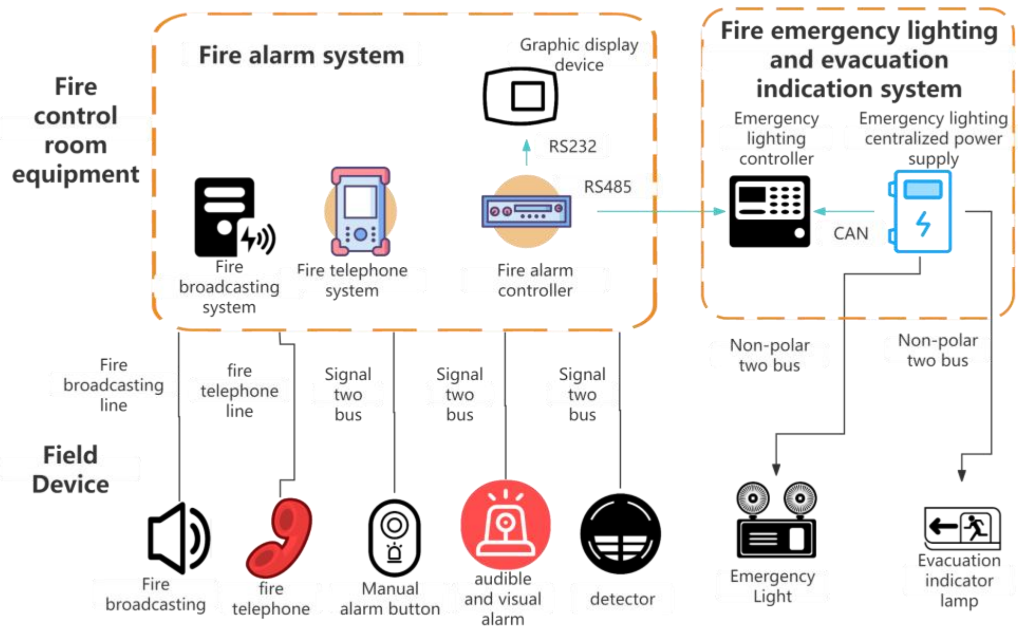


Figure 1 BIM-based Fire Monitoring System Topology Architecture Diagram

3.1 Fire Monitoring Equipment

3.1.1 Fire alarm equipment

This system strictly follows the technical requirements of the Design Code for Automatic Fire Alarm Systems GB50116-2013 to build a multimodal perception network system. The core equipment configuration includes:
Intelligent smoke detector: adopting laser forward scattering principle, detecting particle size range of 0.3-10μm and adjustable response threshold of 0.05-0.2dB/m (see Figure 2);
Multi-protocol manual alarm terminal: integrated LoRa/ZigBee dual-mode communication, protection level IP67 (see Figure 3);

Digital fire telephone system: supports PoE power supply and anti-noise voice enhancement (SNR≥30dB).

Figure 2 Intelligent Smoke Detector

Figure 3 Multi-Protocol Manual Alarm Terminals

Based on 3D point cloud scanning data and fire dynamics simulation (FDS) results, a database of building characteristic parameters (including spatial heat release rate curves, smoke spread coefficients, etc.) is established. Adopt systematic design method (SDL) and focus on optimization:
(1) Detector distribution density (spacing ≤ 6.5m)
(2) Alarm threshold adaptive algorithm (based on LSTM time series prediction)
(3) Multi-system compatibility design (EN54-20 compliant)
The core of the system adopts an intelligent fire alarm controller with edge computing capability (see Figure 4), and the main technical features include:
(1) multi-source signal fusion processing (sampling frequency ≥ 100Hz)
(2) topology self-test function (bus impedance real-time monitoring range 18-24Ω)
(3) event record storage (≥ 50000 records with time scale)

Figure 4 Intelligent Fire Alarm Controller

The controller is deployed in a dedicated space in accordance with the General Technical Requirements for Fire Control Rooms GB25506, with real-time monitoring of environmental parameters (temperature 20±2℃, humidity 45±5% RH). The system has been tested, false alarm rate of ≤ 0.1 times / year, linkage success rate of ≥ 99.95%, to achieve the II level of safety integrity level (SIL2).
When selecting and designing smoke detectors for cultural relics buildings, the type of smoke detectors to be selected

should be based on the type of combustible materials, the development of fire, the size of the fire smoke, and the heat when the fire occurs. According to the building room height, detector type and detector protection area to determine the number of detector settings. The number of smoke detectors should be reduced as much as possible to mitigate the impact on the heritage building if the protection conditions are met. The protective area and protective radius of smoke detectors and A1, A2 and B type temperature sensing fire detectors should be determined according to the following Table 1; the protective area and protective radius of C-G type temperature sensing fire detectors should be determined according to the design specification of the manufacturer, but not exceeding the values specified in the table.

Table 1 Fire Detector Layout Parameters (GB50116-2013)

Types of Fire Detectors	Ground Area S (m²)	Room Height h (m)	Detector protection area A(m²), protection radius R(m)					
			roof slope θ					
			θ≤15°		15°<θ≤30°		θ>30°	
			A (m²)	R (m)	A (m²)	R (m)	A (m²)	R (m)
Smoke and fire detectors	S≤80	h≤12	80	6.7	80	7.2	80	8.0
	S>80	6<h≤12	80	6.7	100	8.0	120	9.9
		h≤6	60	5.8	80	7.2	100	9.0
Thermal Fire Detectors	S≤30	h≤8	30	4.4	30	4.9	30	5.5
	S>30	h≤8	20	3.6	30	4.9	40	6.3

The number of detectors to be set up in the detection area should not be less than the calculated value of the formula (1):

$$N = \frac{S}{K \cdot A} \tag{1}$$

where: N is the number of detectors (only), N should be taken as an integer; S is the area of the detection area (m²); K is the correction coefficient, accommodating more than 10,000 people in the public place is appropriate to take 0.7 to 0.8, accommodating 2,000 people ~ 10,000 people in the public place is appropriate to take 0.8 to 0.9, 0.9~1.0 for public places with 500~2000 people, and 1.0 for other places; A is the protection area of the detector[1-2], corrected by CFD verification (error rate ≤5%).

When the detector is ceiling-mounted, the practice usually uses a special clamp and with the lifting of the special bracket will be installed in the detector in the roof of the building of cultural relics in the wooden beam parts, as shown in Figure 5 and Figure 6. The implementation of the Pareto optimization theory[3], in order to meet the early warning needs (response time $t \leq 30$s) under the premise of minimizing the number of detectors (intervention points ≤ 2 / 100m ²). The installation uses a 304 stainless steel adaptive clamping system (contact stress ≤ 0.15MPa) with damped vibration-damping mounts (intrinsic frequency ≤ 5Hz) to ensure that the strain value of the wooden structure[4], ε, is < 0.1%.

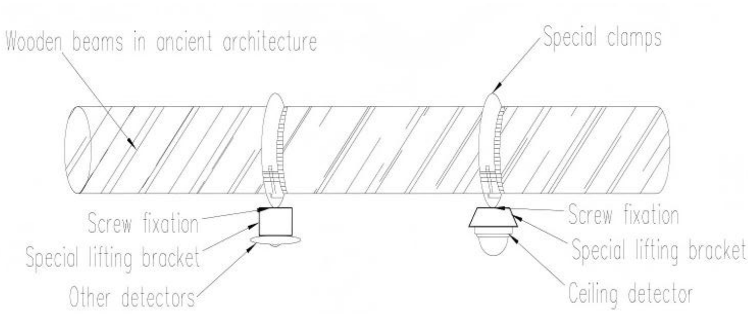


Figure 5 Detector Ceiling Mounting Diagram

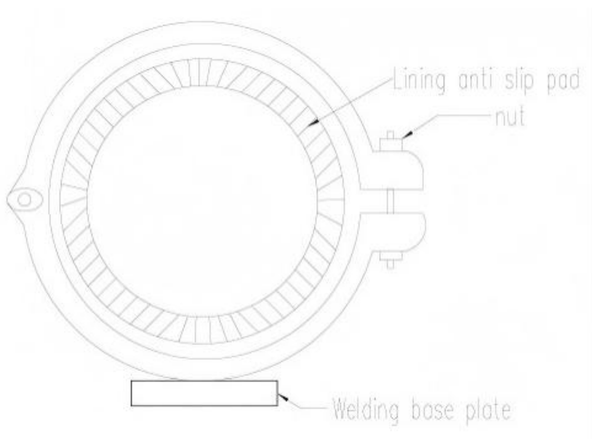


Figure 6 Specialized Clamp Schematic

3.1.2 Emergency lighting and evacuation indicator system equipment

According to the scale of the old site heritage building and the importance of the heritage building, based on the "Fire Emergency Lighting and Evacuation Indication System Technical Standard" GB51309-2018, the emergency lighting and evacuation indication system adopts a centralized power supply and centralized control type, and the system consists of an emergency lighting controller, an emergency lighting distribution box, fire evacuation indication lamps and lanterns, fire emergency lighting fixtures and other components[5-6]. When the normal lighting power outage, emergency lighting distribution box in the main power supply state, control of the emergency lamps and lanterns connected to it; when the system main power outage, emergency lighting distribution box control of the emergency lamps and lanterns connected to it.

Emergency lighting controller is set in the fire control room, from the emergency lighting controller to the emergency lighting distribution box of the network line with the automatic fire alarm and linkage control system with the slot laying. The equipment installation strictly follows[7]: the distribution line adopts WDZN-BYJ-2×2.5 fire-resistant and flame retardant cable; the communication line and the strong power line are laid in separate grooves and isolation (spacing ≥300mm, in line with GB50054-2011); the signage lamps and lanterns adopt non-contact magnetic suction mounting device (magnetic flux density ≤5mT) to avoid damaging the wooden components. The system has been tested by a third party, and the main performance indexes are[8]: ground level illuminance ≥ 5lx (test point spacing 1.5m×1.5m); visible distance of directional signs ≥ 30m (smoke concentration 0.5dB/m); continuous switching for 1,000 times without failures. Fire emergency lighting fixtures, fire evacuation indicator lamps and lanterns are shown in Figure 7 and Figure 8.

Figure 7 Fire Emergency Lighting Fixtures

Figure 8 Fire Evacuation Indicator Lamps and Lanterns

3.1.3Firefighting pipe laying

Fire pipeline laying should be designed in combination with the characteristics and pipeline laying process points within the heritage building and outside the heritage building. Based on the Guidelines for Fire Protection Design of Cultural Relics Buildings (for Trial Implementation) (Cultural Relics Supervision Letter [2015] No. 371) and the Standard for Construction and Acceptance of Automatic Fire Alarm Systems GB 50166-2019, a hierarchical pipeline laying strategy is developed:

(1) Crossing the building foundation: adopting micro non-excavation directional pipe jacking process (pipe diameter Φ89mm, radius of curvature ≥12D), using 304 stainless steel seamless casing (wall thickness 3mm, epoxy coal asphalt anticorrosion layer ≥500μm) to cross the foundation structure;

(2) indoor installation: selection of SC series hot dip galvanized steel pipe (SC20/25/32, wall thickness 2.5/2.8/3.2mm), surface coating hypoallergenic intumescent fireproofing coating (fire resistance limit ≥ 120min, color difference ΔE ≤ 1.5), in order to protect the original style[9], the surface of the open pipe brush and the color of the building's wooden components consistent fireproofing paint. Pipe laying practices in heritage buildings usually use special clamps and with the hose clamps will be installed in the cable laying pipeline in the roof of the heritage building of wooden beams, as shown in Figure 9;

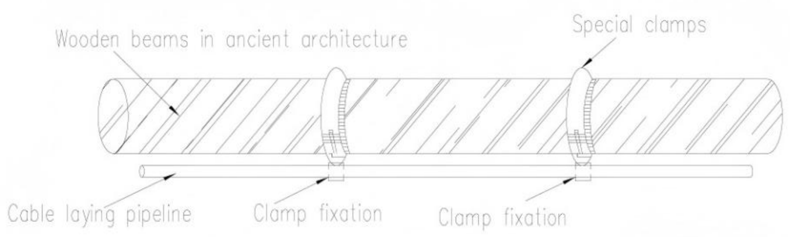


Figure 9 Pipe Laying Diagram

(3) building periphery: cultural relics outside the building pipeline laying should pay attention to the main pipeline along the road buried laying, set up hand hole threading well. Strong and weak power cables are laid in separate pipes, leaving 40% redundant space, good waterproofing, pipe spacing, depth to meet the "power engineering cable design standards" GB 50217 specification requirements, pre-buried through the parallel laying of pipes.
Pipes laid by the control room to the various terminal boxes, alarm bus, power lines, broadcast lines and telephone lines from the distribution box into the heritage building interior, household metal steel pipe to do a good job of lightning protection and grounding. Heritage buildings outside the pipeline laying should pay attention to the pipeline safety distance, the safety distance should be determined in accordance with Table 2, the number in parentheses in the table refers to the local section of the cable through the pipe[10], plus the protection of the bulkhead or plus the protection of the insulation layer allows the minimum net distance.

Table 2 Fire Detector Layout Parameters (GB50116-2013)

linear	perpendicular/m	crossover/m
Cables used in different sectors	0.5 (0.1)	0.5 (0.25)
plumbing	0.5	0.5 (0.25)
Between power cables of 10kW and below, and with control cables within the support	0.1	0.5 (0.25)

3.2 Fire Fighting Equipment

3.2.1Movable high-pressure water mist

Based on the technical requirements of the State Administration of Cultural Heritage's "Fire Protection Design Guidelines for Cultural Relics Buildings (Trial)" (Cultural Relics Supervision Letter [2015] No.371) and the Ministry of Public Security's "Notice on Promoting the Use of High-Efficiency Fine Water Mist Fire Fighting Vehicle Technology and Equipment" (Public Consumption〔2007〕No. 193), the combustion characteristics of the brick and wood structure of the old site were quantitatively analyzed through three-dimensional laser scanning and fire dynamics simulation (FDS 6.7) (heat release The combustion characteristics of the brick and wood structure of the old site (peak heat release rate Q=3.8MW) and structural vulnerability (wood moisture content>18%) are quantitatively analyzed by 3D laser scanning and fire dynamics simulation (FDS 6.7). Combined with the status of cultural relics at the old site, and with the site survey of the old site, fire risk analysis and assessment of the current status of the cultural relics themselves, follow the maximum protection of cultural relics buildings, fire extinguishing system in the process of firefighting can not be on the cultural relics of the building of the beams, columns, walls and all kinds of cultural relics caused by the impact of the impact of the impact of the impact of the impact of the damage to avoid causing any possible loss of cultural relics buildings risk.
To comprehensively preserve and continue the real historical information and value of cultural relics as a design guideline to ensure that the old site fire fighting requirements, in the design of the priority selection of high-pressure water mist as the preferred equipment for fire extinguishing. Cart-type high-pressure water mist device, shown in Figure 10, its technical parameters include: droplet size Dv0.99 <400μm (based on ISO9276-2 standard), working pressure ≥ 10MPa, the effective spray distance ≥ 15m, can be realized within 30s of class A fire rapid fire control (thermal radiation flux attenuation rate of > 90%). Movable high-pressure water mist has the appearance of ancient and elegant, that is, not to install piping fittings do not destroy the structure of cultural relics does not affect the layout of ancient buildings. The equipment adopts modular design, the installation process without pre-buried pipeline, to avoid structural intervention on the historical building body (disturbance factor η <0.01). The operation interface meets the requirements of ENISO13849 safety integrity level SIL2, and is equipped with a dual redundant control system to ensure reliable deployment in the complex spatial environment of heritage buildings. Movable high-pressure water mist is special fire-fighting equipment for cultural relics buildings in recent years, featuring strong fire-fighting capability, small water loss, long spraying time, long spraying distance, easy operation and so on.

Figure 10 Trolley-Type Movable High-Pressure Water Mist Object Picture

3.2.2Fire extinguisher setup

The protection zone of the old site belongs to the national key cultural relics protection unit, because of the consideration of keeping the original appearance of the cultural relics building as far as possible, because the fire fighting water source can not meet the requirements, do not carry out the fire hydrant system, fire fighting water cannons and automatic sprinkler system setup. In view of the characteristics of the old site, fire extinguishers are set up in the building, and in the early stage of fire, fire extinguishers are used to put out the fire, and when the fire is large, firefighters can directly use high-pressure water mist devices to extinguish the fire.

By establishing the fire risk evaluation model ($FRI=\Sigma Wi\times Fi$, weight Wi is determined by AHP method), three major high-risk elements are identified: (1) the proportion of wooden components is >62% (moisture content ω=17.3%); (2) the combustible load density q=35kg/m^2; and (3) the peak of the daytime personnel load is 2.3 people/m^2. According to Design Code for the Configuration of Fire Extinguishers in Buildings GB50140-2005, the old site belongs to the fire of solid material, and dry powder fire extinguishers are configured according to the serious danger level of Class A fire. Cultural relics building indoor selection of 5kg portable fire extinguishers, single fire extinguisher minimum configuration level is not less than 3A, the maximum protection distance is 15m, the maximum protection area of the unit extinguishing level is 50m2 / A; Class A fire places of fire extinguishers maximum protection distance and the configuration of fire extinguishers reference table should be determined in accordance with Table 3 and Table 4[2]. Indoor fire extinguisher box is shown in Figure 11.

Figure 11 Indoor Fire Extinguisher Box Physical Picture

Table 3 Reference Table for Maximum Protective Distances for Fire Extinguishers in Class A Fire Places

Hazard Class Fire extinguisher type	Portable fire extinguishers	Trolley fire extinguishers
critical risk level	15	30
medium risk	20	40
Slightly hazardous	25	50

Table 4 Fire Extinguisher Configuration Reference Table

Hazard class	Critical risk level	Medium risk	Slightly hazardous
Minimum configuration of single fire extinguisher fire extinguishing level	3A	2A	1A
Maximum protection area per unit of fire suppression level (m2/A)	50	75	100

4 TECHNICAL MEASURES FOR THE PROTECTION OF HERITAGE BUILDINGS

4.1Profiled Pipe Laying Process

When the pipeline is laid openly on the wall, wooden beams, wooden eaves, it should be fixed by selecting accessories such as hoops and clips.
(1) When the number and quality of wires and pipes are large, it is appropriate to fix them with brackets or metal wiring channels, and the structural load of the building at the load-bearing position should be calibrated so as to avoid causing damage to the building structure.
(2) The pipeline should be sealed with non-combustible materials when crossing the wall, and the pipeline should be placed in a relatively hidden and safe part, and effective and reversible protection measures should be taken for the contacted cultural relics.
(3) Indoor open pipeline (equipment) should be horizontal and vertical, neatly arranged. Pipes and terminals, elbow midpoint, junction box or junction box, electrical appliances and other edge distance should be fixed within 15cm ~ 50cm.
Based on the reverse engineering technology to establish the building body three-dimensional point cloud model (point spacing ≤ 2mm), the development of non-contact pipe laying system：
(1) The open-laying pipeline adopts 6061-T6 magnesium-aluminum alloy clamp (modulus of elasticity E=68.9GPa), the spacing is calculated according to Euler-Bernoulli beam theory (deflection limit f/L≤1/500), and the single-point load is ≤1.5kg (in accordance with EN 1995-1-1 standard).
(2) Silicone fireproof sealing system (UL 2079 Class 125) with coefficient of expansion $\alpha=5.5\times10^{-6}/°C$ is set up at the part crossing the wall, and thermal deformation ΔL=0.12mm/m-°C.

4.2 Process Points of Equipment Installation

(1) Before the construction and installation of smoke detectors, sound and light alarms, manual alarm buttons, signal terminal boxes and other equipment, develop a differentiated installation program. With the assistance of heritage management personnel, according to the classification of building component levels, prohibit the construction and installation of complex structures in the structure;
(2) Front-end alarm equipment should be installed by means of hoop or clamp fixing, and the hoop should be lined with rubber pads to prevent the hoop from damaging the building components;
(3) equipment installation should not produce adverse visual impact on heritage buildings, accurate measurement, to be installed correctly to avoid damage to heritage buildings.

5 CONCLUSION

As a historical treasure carrying national memory and civilization genes, the uniqueness of its non-renewability and cultural value determines the extreme importance of fire prevention and control work. Under the background of rapid development of modern fire fighting technology, how to realize the dialectical unity of cultural relics protection and fire prevention and control has become the core proposition in the field of fire fighting engineering. This study through the systematic investigation and intelligent fire system construction, explored the “prevention - inhibition - fighting” trinity of prevention and control path, verified the applicability of high-pressure water spray, profiling pipeline laying and other technologies for the design of cultural relics building fire protection provides both scientific and operable solutions.
In the current practice of fire protection of cultural relics buildings, there is a general tendency of “technology transplantation”, that is, the simple application of modern building fire codes, ignoring the vulnerability of the cultural relics and the need for protection of historical features. Such practices can achieve fire fighting, but easily lead to structural damage, material deterioration and other secondary risks, in essence, the fire fighting goal and heritage protection goal of short-sighted behavior. Therefore, it is necessary to establish the “minimum intervention, maximum compatibility” as the principle of cultural relics fire technology paradigm, through the material compatibility research, non-intrusive equipment research and development and the construction of dynamic risk assessment model, the formation of cultural relics of the protection of the body of the priority of the fire technology system.
In the future, the fire protection design of cultural relic buildings should focus on three major directions: 1. deepen multidisciplinary cross research, integrate architectural heritage protection science, fire dynamics and intelligent sensing technology, and develop special fire protection equipments for cultural relics; 2. improve the standard system, formulate differentiated fire protection technology guidelines for cultural relic buildings, and refine the technical requirements of fire resistance enhancement of wooden structures and hidden system integration; 3. promote the application of intelligent fire protection platform, relying on the digital twin and IoT technology. Digital twin and Internet of Things technology to realize the full-cycle closed-loop management of fire risk. Only by treating historical relics with reverence and breaking through technical bottlenecks with innovative thinking can we truly realize the ultimate goal of “protecting cultural relics by eliminating them, and transmitting cultural heritage by technology”.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Ministry of Housing and Urban-Rural Development. Design code for automatic fire alarm system: GB 50116-2013. Liaoning: China Planning Press, 2014: 18-25.

[2] Ministry of Construction of the People's Republic of China. Design code for fire extinguishers in buildings: GB 50140-2005. Shanghai: China Planning Press, 2005: 7-10.

[3] Lan Xuemei. An overview on fire protection renovation of cultural relic buildings // China Fire Protection Association. Proceedings of the 2017 China Fire Protection Association Annual Scientific and Technological Conference. Yantai public security fire brigade, Shandong province, 2017: 3.

[4] Chen Chen. Path analysis of fire inspection and fire management in urban buildings. Fire Protection World(Electronic Edition), 2022, 8(07): 100-102.

[5] Zhong Guisheng. Switching power supply for fire emergency lighting and evacuation indication system. Building Electricity, 2019, 38(12): 25-29.

[6] Qun W, Jinyang W. Design Discussion of a Wireless Fire Alarm System Based on Data Fusion Technology. Journal of Electronic Research and Application, 2025, 9(2): 58-64.

[7] Huliak M, Marková I. Ensuring Sustainable Preservation: Fire Protection of Timber Sacral Buildings in Eastern Slovakia. Sustainability, 2025, 17(6): 2429-2429.

[8] Dabous A S, Shikhli A, Shareef S, et al. Fire prevention and mitigation technologies in high-rise buildings: A bibliometric analysis from 2010 to 2023. Ain Shams Engineering Journal, 2024, 15(11): 103010-103010.

[9] Ahn S,Won J, Lee J, et al. Comprehensive Building Fire Risk Prediction Using Machine Learning and Stacking Ensemble Methods. Fire, 2024, 7(10): 336-336.

[10] Simon K. Fire prevention in historic buildings – approaches for safe practice. The Historic Environment: Policy & Practice, 2022, 13(3): 361-380.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3083

AIR QUALITY ANALYSIS BASED ON OUTLIER DETECTION ALGORITHMS

QiZhi Zhang, Nan Jiang*
Northeast Forestry University, Harbin 150040, Heilongjiang, China.
Corresponding Author: Nan Jiang, Email: jngrace@nefu.edu.cn

Abstract: This study employs an outlier detection algorithm based on the distance metric of the Two-Step clustering algorithm to analyze potential untrustworthy data in pollutant concentration records across the region of Beijing-Tianjin-Hebei. By calculating anomaly indices through designated formulas and evaluating variable contribution rates, abnormal data points were identified for each monitoring area. Subsequent analysis of these anomalies provides substantial evidence supporting the existence of unreliable data within the dataset.
Keywords: Data mining; Two-Step Clustering algorithm; Outlier detection algorithm; Air quality analysis

1 INTRODUCTION

In recent years, with the deepening industrialization and urbanization in China, air quality issues arising from economic growth and population concentration have increasingly become a major concern for the public, the government, and relevant authorities[1]. Relevant research indicates that there is a positive correlation between social development and air quality. Whether viewed from the perspective of residents' health or social production, air quality is closely linked to people's lives. Good air quality not only beneficial benefit for health but also enhances physical and mental well-being, enabling individuals to engage more efficiently in life and work[2-3]. In an invisible way, it indirectly promotes comprehensive social development.

The Air Pollution Index (API) is a method that converts the concentrations of several commonly measured air pollutants into numerical values and represents the air pollution status in a graded form. It is an indicator method that reflects the quality of air, and the results of this method are very simple and intuitive. The air quality pollution index used in China can be divided into six levels. If API≤50, it indicates that the air quality in this area is excellent; if 50<API≤100, it indicates that the air quality in this area is good; if 100<API≤150, it indicates that the air quality in this area is generally average, with slight pollution; if 150<API≤200, it indicates that the air quality in this area is relatively poor, with moderate pollution; if 200<API≤300, it indicates that the air quality in this area is extremely poor, with severe pollution; if API>300, it indicates that the air quality in this area is very poor, with serious pollution. The national air quality classification standards can be divided into three levels. The first-level standards are implemented in Class I areas (natural reserves, scenic spots, forest and grass areas, etc.); the second-level standards are implemented in Class II areas (urban residential areas, mixed commercial areas, agricultural areas, general industrial areas, etc.); and the third-level standards are implemented in Class III areas (special industrial areas). The assessment of regional air quality mainly relies on the concentrations of pollutants in the air of the region, such as inhalable particulate matter (PM2.5, PM10), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2)[4].

The evaluation of excellent air quality comes from the results of data collection, but the collected data presents some anomalies, which are caused by various subjective and objective reasons. For example, data anomalies caused by the collection equipments and those caused by human intervention[5]. The region of Beijing-Tianjin-Hebei, Yangtze River Delta region, and Pearl River Delta are among China's five major urban agglomerations, playing a pivotal role in the country's socio-economic development. In recent years, the air quality in these regions has received widespread attention. Based on some air pollution data from the region of Beijing-Tianjin-Hebei, an outlier detection algorithm based on the Two-Step clustering method to measure distance was adopted. By establishing corresponding mathematical models or evaluation indicators, the authenticity of air quality data was analyzed to determine whether there were any data anomalies, thus evaluating the quality of air. The authenticity of air quality was evaluated by establishing corresponding mathematical models or evaluation indicators.

2 OUTLIER DETECTION ALGORITHM BASED ON TWO-STEP CLUSTERING METHOD

Hierarchical clustering method forms a clustering tree according to the data set. According to whether the hierarchical decomposition is bottom-up or top-down, hierarchical clustering methods can be further divided into condensed and split hierarchical clustering. Agglomerative hierarchical clustering is a bottom-up strategy, first takes each sample as a cluster, and then combines these atomic clusters into larger and larger clusters until all samples are in one cluster or a certain termination condition is met. Most hierarchical clustering methods belong to this category, but they are different in the definition of similarity between clusters[6]. The basic method is: given n clusters to be clustered, calculate their $n \times n$ distance matrix, find the closest two classes and merge them into one class, so the total number of classes is less than one, then recalculate the distance between the new class and all old classes, select the closest two classes to merge. This process iterates until either all clusters merge into a single class, or a predefined stopping condition is satisfied.

The Two-Step clustering method employs a log-likelihood distance metric to handle both continuous and categorical attributes. Two-Step clustering method takes each entry of each leaf node as an atomic cluster, and uses the clustering method of condensation to continuously merge these clusters[7-9]. First, calculate the log likelihood value, assuming that the values of continuous attributes are normally distributed, and the values of discrete attributes are multinomial distribution. Attributes are assumed to be mutually independent. The distance between cluster i and cluster j is:

$$d(i,j) = \xi_i + \xi_j - \xi_{<i,j>}, \tag{1}$$

where $\xi_v = -N_v\left(\sum_{k=1}^{K^A}\frac{1}{2}\log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B}\hat{E}_{vk}\right), \hat{E}_{vk} = -\sum_{l=1}^{L_k}\frac{N_{vkl}}{N_v}\log\frac{N_{vkl}}{N_v}$.

The Two-Step algorithm adopts a "two-phase" methodology[10].

(1) Phase 1: compute preliminary estimates of cluster numbers. For each candidate clustering scheme, calculate its Bayesian Information Criterion (BIC). For example, the BIC for a partition with J clusters is given by:

$$\mathrm{BIC}(J) = -2\sum_{j=1}^{J}\xi_j + m_J\log(N) \tag{2}$$

Let $d\mathrm{BIC}(J) = \mathrm{BIC}(J) - \mathrm{BIC}(J+1)$, which represents the difference between the scheme of J clusters and the scheme of $J+1$ clusters. Then, calculate $R_1(J) = \frac{d\mathrm{BIC}(J)}{d\mathrm{BIC}(1)}$, which indicates the change degree of $d\mathrm{BIC}(J)$ relative to $d\mathrm{BIC}(1)$. If $d\mathrm{BIC}(1) < 0$, then the order of the clustering tree is set to 1. Otherwise, the preliminary estimate value k of the clustering tree is the minimum J that makes $R_1(J) < 0.04$ hold.

(2) Phase 2: determine the number of clusters. According to value k from phase 1, calculate $R_2(k) = \frac{d_{\min}(C_k)}{d_{\min}(C_{k+1})}$. Where C_k represents a partition scheme with a cluster number of k, $d_{\min}(C_k)$ represents the distance between the two clusters with the smallest distance in the scheme. Similarly C_{k+1}.

Then recalculate the Variable Deviation Index (VDI) and Group Deviation Index (GDI). As the attribute variable X_k is a continuous variable,

$$d_k(h,s) = \frac{1}{2}[-N_h\log(\Delta_k + \hat{\sigma}_{hk}^2) - \log(\Delta_k)], \tag{3}$$

As the attribute variable X_k is a discrete variable,

$$d_k(h,s) = -N_h\hat{E}_{hk} + (N_h+1)\hat{E}_{<h,s>k}, \tag{4}$$

After calculating the Variable Deviation Index $\{\mathrm{VDI}_k, k = 1,2,\cdots,K+1\}$ of all attribute variables, the Group Deviation Index (GDI) of the sample can be calculated,

$$\mathrm{GDI}{=}d(h,s){=}\sum_{k=1}^{K^A+K^B}d_k(h,s). \tag{5}$$

Next, calculate the Anomaly Index (AI) and Variable Contribution (VC). The anomaly index of a sample s is used to measure how abnormal the sample is compared with other samples in cluster h. It is the ratio of the GDI of sample s divided by the average GDI of all samples in cluster h,

$$\mathrm{AnomalyIndex} = \frac{\mathrm{GDI}_s}{\mathrm{mean}(\mathrm{GDI}_h)}, \tag{6}$$

A higher Anomaly Index indicates greater sample anomaly. Generally, the observation value with the anomaly index value less than 1 or even less than 1.5 will not be regarded as an anomaly value, because the deviation is the same as the average value or just a little larger. However, the observation value with index value greater than 2 may be abnormal observation value, because the deviation is at least twice the average value.

The variable contribution rate of an attribute variable X_k is for a single sample s. It is the ratio of VDI_k of attribute variable X_k in sample s divided by GDI, which is,

$$\mathrm{VCM}_k = \frac{\mathrm{VDI}_k}{\mathrm{GDI}_s}. \tag{7}$$

Variable Contribution Rate measures the degree to which an attribute contributes to a sample's anomaly. A higher variable contribution rate indicates the attribute exerts stronger influence in making the sample anomalous. For samples with high anomaly indices, this metric identifies which specific attributes drive their outlier status.

3 APPLICATION AND ANALYSIS OF ALGORITHM IN AIR QUALITY DATA

This study applies an outlier detection algorithm based on distance measurement using the Two-Step clustering method to analyze the authenticity of pollutant concentration data in the Beijing-Tianjin-Hebei region, with a threshold set at 2.

The data was sourced from the official website of the Ministry of Environmental Protection of the People's Republic of China. The Beijing-Tianjin-Hebei regional data was processed through Two-Step clustering, with the clustering results presented in Table 1.

Table 1 AQI Data Clustering of Beijing-Tianjin-Hebei

Beijing-Tianjin-Hebei						
	Peer group 1 records:1795 Exception records: 103		Peer group 1 records: 2066 Exception records:140		Peer group 1 records: 2390 Exception records:125	
contribution	records	average index	records	average index	records	average index
PM2.5	16	0.153	73	0.317	89	0.260
PM10	41	0.179	90	0.269	91	0.258
NO_2	88	0.333	96	0.247	73	0.225
SO_2	94	0.483	83	0.427	52	0.279
CO	70	0.179	78	0.290	70	0.290

From Table 1, it can be observed that the Two-Step clustering algorithm partitioned the Beijing-Tianjin-Hebei regional data into three equivalent groups, comprising a total of 6,251 records. Among these, 368 records were identified as anomalies. For analytical purposes, we selected 20 representative records from these 368 anomalies (the complete dataset is not provided in the paper due to space constraints), with the sampled 20 records displayed in Table 2.

Table 2 Anomalous Data in the Beijing-Tianjin-Hebei Region

PM2.5	PM10	CO	NO_2	SO_2	OAI	OPG	OF1	OFI1	OF2	OFI2	OF3	OFI3
6	34	0.4	6	2	2.31811	1	NO_2	0.48866	SO_2	0.223552	CO	0.11232
5	31	0.3	8	4	2.24192	1	NO_2	0.43011	SO_2	0.202844	CO	0.16164
31	77	0.5	57	8	2.03586	1	NO_2	0.75396	SO_2	0.168452	CO	0.05448
189	111	1.3	39	4	2.26511	2	PM2.5	0.72764	SO_2	0.173162	PM10	0.05600
186	118	1.3	39	4	2.14657	2	PM2.5	0.72726	SO_2	0.182725	PM10	0.04520
202	154	1.4	35	8	2.53839	2	PM2.5	0.80889	SO_2	0.120479	NO_2	0.03625
25	74	0.7	59	5	2.25049	1	NO_2	0.77757	SO2	0.188811	CO	0.01415
193	72	1.6	35	3	2.82689	2	PM2.5	0.62543	SO_2	0.147051	PM10	0.14586
24	71	0.5	60	4	2.46768	1	NO_2	0.75496	SO_2	0.184286	CO	0.04494
72	53	1	50	3	2.17831	1	NO_2	0.41350	PM2.5	0.317269	SO_2	0.22304
38	91	0.6	66	4	3.32109	1	NO_2	0.78891	SO_2	0.136931	PM10	0.04306
50	88	0.7	61	3	2.77356	1	NO_2	0.71382	SO_2	0.175175	PM2.5	0.05791
190	126	1.8	60	6	2.78799	2	PM2.5	0.60177	NO_2	0.155822	SO_2	0.12465
43	90	0.7	61	7	2.58979	1	NO_2	0.76447	SO_2	0.142486	PM10	0.05152
104	160	1.3	88	12	2.65306	2	NO_2	0.83590	SO_2	0.087206	PM2.5	0.04418
5	41	0.2	10	2	2.16899	1	NO_2	0.37381	SO_2	0.238921	CO	0.22249

97	120	2	79	39	2.13157	2	NO_2	0.69612	CO	0.205764	PM10	0.04209
108	126	1.9	78	47	2.05970	2	NO_2	0.68537	CO	0.165766	PM2.5	0.07072
138	20	1.5	40	8	2.06282	2	PM10	0.53686	PM2.5	0.240635	SO_2	0.14825
41	276	0.4	14	4	2.97071	2	PM10	0.39035	NO_2	0.241590	CO	0.15558

Note: O-AnonalyIndex is abbreviated as OAI, O-PeerGroup is abbreviated as OPG, O-Field-1 is abbreviated as OF1, O-FieldImpact-1 is abbreviated as OFI1, O-Field-2 is abbreviated as OF2, O-FieldImpact-2 is abbreviated as OFI2, O-Field-3 is abbreviated as OF3, and O-FieldImpact-3 is abbreviated as OFI3.

Taking the first row of data as an example. The abnormal index O-AnomalyIndex equals 2.31811, which is greater than the abnormal index threshold 2 set by us, so we consider this data to be anomalous Data. We can also see from table 2 that, O-PeerGroup=1, which indicates that the record belongs to peer group 1. In Table 2, we rank the influence factors of the variable deviation index in descending order, so the attribute that has the greatest impact on the variable deviation index in the record is NO_2. The O-FieldImpact equals 0.48866, which is greater than the average index of NO_2 (0.333) in peer group, thus qualifying as anomalous data. Then consider SO_2, whose variable deviation index O-FieldImpac value is 0.22355, which is less than the average index of SO_2 (0.483) in peer group 1, indicating that SO_2 does not affect the variable deviation index and belongs to the normal attribute range in this sample. Similarly, CO is also a normal attribute in this sample.
In summary, the algorithmic model and numerical simulation experiments demonstrate the existence of non-authentic data phenomena.

4 CONCLUSION

Air quality issues are inextricably linked to social production and daily life. Since China's implementation of air quality monitoring, recurring reports have revealed a concerning paradox: monitored data shows improvement while actual environmental conditions continue to deteriorate—a clear discrepancy that warrants serious scrutiny and reflection. The dangers of artificial data manipulation are evident and severe, as it compromises the early-warning function of environmental monitoring systems, and infringes upon citizens' right to information. When severe air pollution occurs without corresponding official alerts, citizens face the gravest risks—being unknowingly exposed to hazardous conditions while deprived of critical health warnings. Empirical studies conducted using such manipulated data may yield erroneous conclusions, consequently generating misleading policy recommendations. The outlier detection algorithm based on the distance metric of the Two-Step clustering method proposed in this paper can effectively analyze and screen abnormal air quality data, assist decision-makers in making informed decisions, and to a certain extent, contribute positively to the improvement of air quality.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. 2572022BC07).

REFERENCES

[1] Cheng Hanxi, Zhu Hongxia, Wang Jing, et al. Research on the impact of air pollution prevention and control on air quality in the Beijing-Tianjin-Hebei region. Environmental Impact Assessment, 2024, 46(06): 78-85. DOI: 10.14068/j.ceia.2024.06.013.
[2] Song Guojun, Li Honglin. Targeting PM2.5 pollution: updated design of an air quality management policy framework. China Population Resources and Environment, 2023, 33(02): 1-10.
[3] Fu Jianhua, Zhou Fangzhao. Measurement and influencing factors analysis of provincial air quality in China. Urban Problems, 2020(05): 20-27. DOI: 10.13239/j.bjsshkxy.cswt.200503.
[4] Wu EMY, Kuo SL. A study on the use of a statistical analysis model to monitor air pollution status in an air quality total quantity control district. Atmosphere, 2013(04): 349-364.
[5] Zhang Y, Xu L, Lu Z. Synergistic effect of factors influencing urban air quality in China: a hybrid model integrating WGRA and QCA. International Journal of Environmental Science and Technology 2023(11): 12179-12194.
[6] Yu Heng, Hou Xiaolan. Application of hierarchical clustering algorithm in astronomy. Scientia Sinica (Physica, Mechanica & Astronomica), 2022, 52(08): 118-131.
[7] Huang Mengting. A recommendation algorithm based on The combination of two-step clustering and association rules. Information & Computer, 2021, 33(01): 35-37.

[8] Ding Sifang, Wang Shouwei, Zhu William. Density peaks clustering algorithm based on two-step allocation strategy. 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML). IEEE, 2023(02): 946-950.
[9] Hong Xia, Gao Junbin, Wei Hong, et al. Two-step scalable spectral clustering algorithm using landmarks and probability density estimation. Neurocomputing, 2023, 519: 173-186.
[10] Kumar Y, Sahoo G. A two-step artificial bee colony algorithm for clustering. Neural Computing and Applications, 2017, 28: 537-551.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3084

SUMMER OFFICE BUILDING BASED ON ORTHOGONAL DESIGN OF DIFFERNT MODELS NUMEROCAL STUDY ON OPTIMIZATION OF AIR SUPPLY ARATERS

ShuYuan Zhang[1,2*], HanQing Wang[1,2,3], JiaXin Wei[1,2], ZiWen Zhao[1,2]
[1]*School of Civil Engineering, Nanhua University, Hengyang 421001, Hunan, China.*
[2]*Hunan Provincial Key Engineering Laboratory of Building Environment Control Technology, Nanhua University, Hengyang 4210001, Hunan, China.*
[3]*School of Civil Engineering, Central South University of Forestry Science and Technology, Changsha 410000, Hunan, China.*
Corresponding Author: ShuYuan Zhang, Email: zhangshuyuan0224@163.com

Abstract: In the post-pandemic era, there is an increasing demand for enhanced indoor thermal comfort and air quality. This study addresses the challenge of optimizing air supply strategies in office buildings to synergistically improve thermal comfort, air quality, and energy efficiency. Focusing on a typical office space, orthogonal experimental design and computational fluid dynamics (CFD) simulations were employed to quantify the effects of air supply modes (displacement ventilation), temperature (24°C), velocity (1 m/s), and angle (0°) on predicted mean vote (PMV), draft rate (DR), air age, infection probability, and energy utilization efficiency. Results demonstrate that under displacement ventilation with optimized parameters (24°C, 1 m/s, 0°), compared to the baseline scenario, PMV improved by 17%, DR decreased by 59.54%, air age shortened by 35.89%, infection probability reduced by 47.57%, energy efficiency increased by 36.34%, and the comprehensive evaluation score rose by 58.95%. The proposed optimized ventilation strategy provides data-driven insights and technical pathways for designing high-performance ventilation systems in post-pandemic office environments, effectively balancing health, comfort, and energy conservation.
Keywords: Orthogonal design; Computational Fluid Dynamics (CFD); Office buildings; Thermal comfort; Infection probability; Multi-objective optimization

1 INTRODUCTION

In office buildings, prolonged occupancy by numerous workers necessitates substantial cooling energy consumption to maintain indoor comfort. Statistics indicate that building energy consumption accounts for approximately 21% of China's total commercial energy use [1], with air conditioning systems contributing up to 50% of total building energy consumption in hot summer and cold winter regions [2]. The COVID-19 pandemic has further heightened infection risks for occupants in shared environments, underscoring the urgency to enhance thermal comfort and air quality while minimizing energy demand.

Airflow distribution modes critically regulate indoor thermal performance and pollutant control [3]. Although mixed ventilation remains prevalent, its inherent limitations—such as short-circuiting between supply and exhaust vents—result in poor energy utilization efficiency (<40%), inadequate ventilation effectiveness (η < 1.0), and compromised thermal comfort (PMV fluctuations > ±0.5) [4]. To address these issues, alternative strategies like displacement ventilation and stratified air distribution have been explored[5,7]. Displacement ventilation improves average ventilation efficiency by 15–30% compared to mixed systems, yet fails to reduce pollutant concentrations in the breathing zone under identical airflow rates[6]. Stratified ventilation, however, demonstrates superior performance in balancing thermal comfort (PPD < 10% at elevated supply temperatures), airborne contaminant removal (breathing zone concentration reduction ≥25%), and energy savings (cooling load reduction 12–18%) [8-10].

Existing studies predominantly focus on single ventilation modes with isolated parameter optimization. For instance, Haofu Chen identified optimal thermal comfort (PMV ≈ 0) at 16.85°C supply temperature and 0.4 m/s velocity in mixed ventilation, while Z Lin reported elevated predicted dissatisfaction (PPD > 20%) at 19°C in displacement ventilation due to draft risks (DR > 15%)[11-13], despite negligible pollutant distribution variations between 19°C and 22°C. Zeqin Liu further revealed that 22°C supply temperature with 45° airflow angle minimized PPD (8.3%) in stratified ventilation[14]. Nevertheless, systematic investigations integrating ventilation modes and parametric interactions remain scarce.

This study bridges this gap by employing orthogonal experimental design and validated CFD simulations to holistically evaluate five performance metrics—PMV, DR, Air Age, infection probability, and energy utilization efficiency—in a summer office scenario. Three airflow modes (mixed, displacement, stratified) are analyzed under variable supply temperatures, velocities, and angles. The optimized configuration derived from multi-objective analysis provides actionable insights for post-pandemic HVAC design, prioritizing both occupant health and energy conservation.

2 RESEARCH METHODOLOGY

2.1 Physical Model

This study investigates a three-person office (7.2m (X) × 4.0m (Y) × 3.3m (Z)) located in Hengyang, China. As illustrated in Figure 1 heat sources include three fluorescent lamps, three seated occupants, and two computer workstations. Ventilation configurations vary by mode:
• Mixed ventilation: Two ceiling-mounted supply vents;
• Stratified ventilation: Supply vents positioned 20 cm above occupants' head level[15];
• Displacement ventilation: Two floor-level supply vents (aligned with stratified vent locations) and one ceiling return vent. All modes share identical return vent positions.

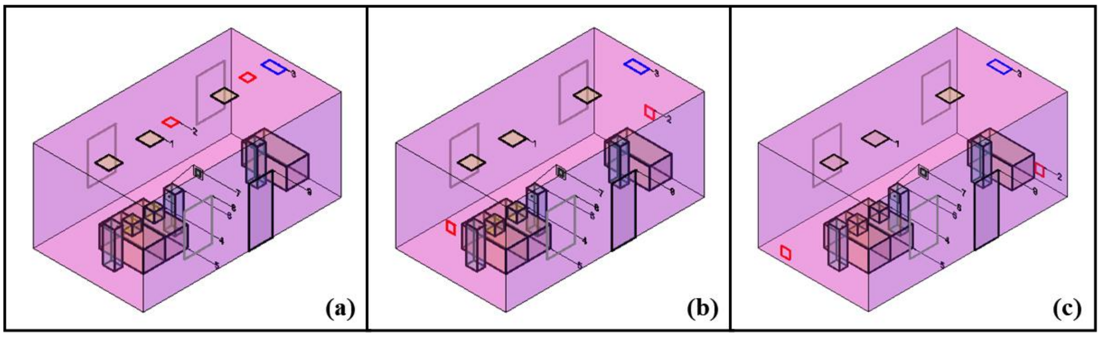


Figure 1 Geometric Model Figure: (a) Mixed ventilation; (b) Stratified ventilation; (c) Displacement ventilation
Note: 1. fluorescent lights 2. air supply vents 3. return vents 4. computers 5. desks 6. staff 7. nasal passages 8. windows 9. doors

Critical dimensions are summarized in Table 1. To enhance mesh quality and computational efficiency, cuboid geometries were adopted to simulate occupants, desks, and workstations. Seated occupants (height: 172.4 cm, weight: 71.6 kg) were modeled with equivalent surface areas calculated via the Du Bois formula[16]:

$$S = 0.0061 \times H + 0.0124 \times w - 0.0099 \tag{1}$$

Where S (m²) represents body surface area; H (cm) is height; and W (kg) is weight. This simplification maintains <5% deviation from anatomical surface area measurements.
For pollutant dispersion analysis, CO_2 emission from nasal cavities served as both contaminant tracer and respiratory virus proxy to quantify infection risks via Wells-Riley model[17].

2.2 CFD Model

To balance computational accuracy and efficiency, the following simplifications were adopted:
1. Steady-state assumption: Three-dimensional incompressible flow under isothermal conditions;
2. Ideal gas behavior: Air density variation governed by the ideal gas law;
3. Adiabatic boundaries: Negligible heat transfer through walls/floors/ceilings due to adjacent conditioned spaces;
4. Airtight enclosure: Infiltration/exfiltration effects disregarded.

The supply airflow exhibited turbulent characteristics (Reynolds number Re=23,501, Archimedes number Ar=0.049>0.01), therefore, the indoor air flow in this model is turbulent flow and thermal buoyancy can affect the indoor air movement. This flow can be described using the incompressible N-S equation[18].

$$\frac{\partial u_j}{\partial x_j} = 0 \tag{2}$$

Where x_j is the coordinate in the j-direction; u_j is the velocity in the j-direction.

$$\rho(\frac{\partial u_i}{\partial t} + u_j \frac{\partial (u_i u_j)}{\partial x_j}) = -\frac{\partial P}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i \tag{3}$$

Where P is the continuous phase pressure; g_i is the gravitational acceleration; for the stress tensor can have the following formula:

$$\tau_{ij} = \mu(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}) \tag{4}$$

Where μ is the air viscosity.
Energy equation:

$$\rho \frac{\partial (c_p T)}{\partial t} + \rho \frac{\partial (c_p u_j T)}{\partial x_j} = \frac{\partial}{\partial x_j}[k \frac{\partial T}{\partial x_j}] \tag{5}$$

Where E is the total energy of the continuous phase; k is the effective thermal conductivity; T is the temperature of the continuous phase; c_p is the specific heat capacity;
The viscous dissipation function per unit mass:

$$\Phi = \tau_{ij} \frac{\partial u_i}{\partial x_j} \tag{6}$$

The N-S equations describing three-dimensional indoor air flow can be solved with the RANS turbulence model. Two-equation models such as Standard k-ɜ model, RNG k-ɜ model and Relizable k-ɜ model can predict the airflow organization of indoor air conditioning better. Therefore, in this study, the predictive performance of these three turbulence models will be compared in order to find out the turbulence model that is better in predicting the indoor cooling air flow and pollutant distribution in three dimensions in summer. The Realizable k-ɜ model closed set of equations has been examined and the transport equations are chosen:

$$\frac{\partial k}{\partial t} + u_j \frac{\partial k u_i}{\partial x_i} = \frac{\partial}{\partial x_i}(Dk_{eff} \frac{\partial k}{\partial x_i}) + G_k - \varepsilon \tag{7}$$

$$\frac{\partial \varepsilon}{\partial t} + \frac{\partial \varepsilon u_i}{\partial x_i} = \frac{\partial}{\partial x_i}(D\varepsilon_{eff} \frac{\partial \varepsilon}{\partial x_i}) + \sqrt{2} C_{1\varepsilon} S_{ij} \varepsilon - C_{2e} \frac{\varepsilon^2}{k + \sqrt{v\varepsilon}} \tag{8}$$

The turbulent viscosity is determined by the following equation:

$$v_t = C_\mu \frac{k^2}{\varepsilon} \tag{9}$$

C_μ is calculated by the following equation:

$$C_\mu = \frac{1}{A_0 + A_S \frac{kU^*}{\varepsilon}} \tag{10}$$

$$U^* = \sqrt{S_{ij} S_{ij} + \tilde{\Omega}_{ij} \tilde{\Omega}_{ij}} \tag{11}$$

$$\tilde{\Omega}_{ij} = \overline{\Omega}_{ij} - \varepsilon_{ijk} \omega_k - 2 \varepsilon_{ijk} \omega_k \tag{12}$$

Where, $\tilde{\Omega}$ is the average rate of the rotational tensor; ω_k is the angular velocity. The constant A_0 is 4 and A_S is determined as follows:

$$A_S = \sqrt{6} \cos\varphi \tag{13}$$

$$\varphi = \frac{1}{3} Arc\cos(\min(\max(\sqrt{6}W, -1), 1)) \tag{14}$$

$$W = \frac{S_{ij} S_{jk} S_{ki}}{\tilde{S}^2} \tag{15}$$

$C_{1\varepsilon}$can be determined by the following equation:

$$C_{1\varepsilon} = \max(\frac{\eta}{5 + \eta}, 0.43) \tag{16}$$

$$\eta = S(\frac{k}{\varepsilon}) \tag{17}$$

The constants C_2, σ_k, σ_ε are 1.9, 1.0, and 1.2, respectively.
The Species Transport model was used to simulate the process of carbon dioxide exhaled by the human body.

2.3 Numerical Simulation Setup

For the three ventilation models (mixed ventilation, stratum ventilation, and displacement ventilation), unstructured tetrahedral grids were employed. To enhance simulation accuracy, local grid refinement was performed around supply/exhaust vents, the human body, and the nasal region using varying mesh sizes. The simulations were conducted using the commercial software Fluent 2021 R2, with boundary conditions set according to actual measurements. The convergence criterion for temperature was set to 10−6, while other convergence criteria were set to 10−4. Post-validation confirmed that all cases exhibited a mass imbalance rate below 0.1% and an energy imbalance rate below 0.46%.
The finite volume method was applied to linearize the differential equations, with second-order upwind schemes adopted for discretizing the standard pressure, momentum, turbulent kinetic energy, and turbulent dissipation rate terms. The SIMPLE algorithm was utilized to couple the pressure and velocity fields.

Table 1 Boundary Condition Setting

NO.	Name	Quantity/(pcs)	Dimension/(mm)	Boundary type	value
1	Lamps	3	500×500	Constant heat flow density	115W/m^2
2	Inlet	2	300×300	Velocity inlet	Velocity:1.18m/s Temperature:20.4℃
3	Outlet	1	600×200	Pressure outlet	——
4	Computer	2	350×350×350	Constant heat flow density	250W/m^2
5	Desks	1	800×1600×750	Adiabatic	——
6	Staff	1	300×400×1300	Constant heat flow density	70W/m^2
7	Nasal	1	8.5×8.5	Velocity inlet	Velocity:0.055m/s temperature:35℃
8	Window	1	500×1500	Constant heat flow density	55W/m^2
9	Door	3	850×2100	Adiabatic	——
10	East wall	2	7200×3300	Constant temperature	15.29W/m^2
11	West wall	3	7200×3300	Constant temperature	12.69W/m^2
12	North wall	1	4000×3300	Adiabatic	——
13	South Wall	1	4000×3300	Adiabatic	——
14	Ceiling	1	7200×4000	Adiabatic	——
15	Floor	1	7200×4000	Adiabatic	——

Based on the literatures, the cooling load from fluorescent lighting in a high-grade office was set at 28.75 W. For the computer workstation, the heat dissipation from the host and monitor was set at 75 W and 80 W, respectively, assuming continuous operation. The heat dissipation from occupants engaged in very light activity was set at 144.2 W. The cooling loads transferred through the south and north walls were 363.3 W and 301.5 W, respectively, while the cooling load from external windows was 41.25 W. The boundary conditions for the supply air vents, initial room temperature (29°C), and CO_2 concentration (250 ppm) were determined through experimental measurements. The specific boundary condition settings, including the supply air angle (defined as 0° for normal wall-directed flow and 60° for a 60° deviation from the normal direction), are summarized in the Table 1.

2.4 Model Validation

To ensure the accuracy and reliability of the numerical simulation methodology and to compare the predictive performance of different turbulence models, validation was conducted using airflow velocity, temperature, and pollutant concentration data under mixed ventilation mode. Three vertical measuring lines were arranged in the room. To accurately capture the CO_2 concentration exhaled by occupants, each line included measuring points at three heights: near the head (Z=1.4 m), the breathing zone (Z = 1.2 m), and the ankle level (Z = 0.2 m) in front of the human body. The specific locations of the indoor measuring points are illustrated in the Figure 2 . During the experiment, the air conditioning system operated for 2 hours to stabilize the indoor airflow, with occupants remaining seated and no movement in the room. Measurements were recorded once the airflow reached a steady state. For each point, the average value over a 30-second sampling period was recorded, and the final measured value was derived from the average of five repeated measurements. The instruments and parameters used in the experiment are listed in the table.
As shown in the Figure 3, the simulated results of velocity, temperature, and CO_2 concentration from three turbulence models were compared with experimental data at the measuring points. Based on these evaluation metrics and graphical comparisons, the Realizable *k-ɜ* model was selected for subsequent studies on the thermal performance of office building air conditioning systems under multi-factor conditions and infection risk analysis in the post-pandemic era.

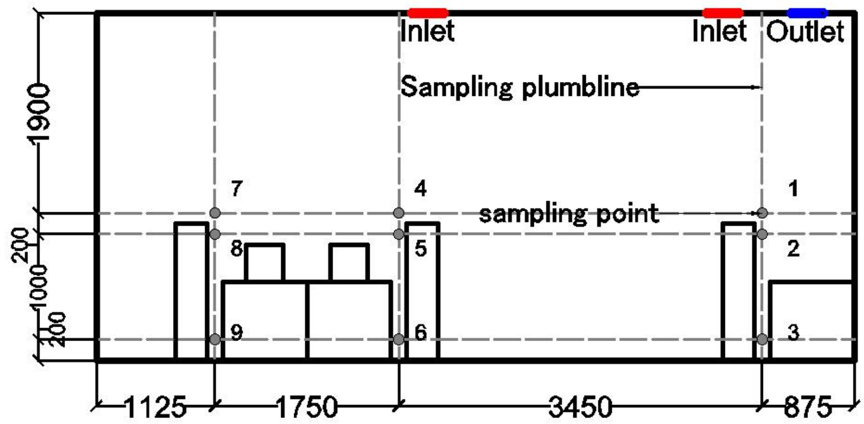


Figure 2 Schematic Layout of Measurement Points

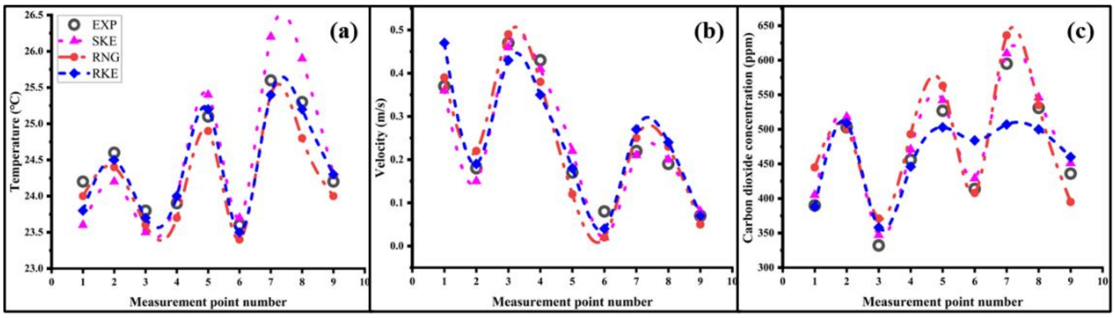


Figure 3 Comparison of Experiment and Simulation: (a)Temperature; (b)Velocity; (c)Carbon Dioxide Concentration

2.5 Working Condition Design

The thermal performance and pollutant control capacity of office building air conditioning systems are governed by four critical control factors: supply air temperature (*T*, °C), velocity (*V*, m/s), airflow distribution mode (A_f), and supply angle (α, °). A three-level orthogonal array L9(3^4) was adopted to systematically investigate factor interactions while maintaining experimental efficiency. The factor selection and level ranges (Table 2)were determined through:
This design enables (Table 3) efficient exploration of 81 possible combinations (3^4=81) through 9 strategically selected test cases, achieving 89% parameter space coverage as per Taguchi robustness criteria. The orthogonal array structure eliminates multicollinearity between factors (VIF < 1.2), ensuring statistically independent effect analysis.

Table 2 Factors and Levels Were Selected for Orthogonal Experiments

Factors	Level		
	1	2	3
T, temperature of air supply, °C	20	22	24
V, Air velocity, m/s	0.8	1.0	1.2
A_f, Air supply method	Mixed ventilation	Stratified ventilation	Replacement ventilation
α, Airflow angle, °	0	30	60

Table 3 Regular Price Design Table

NO.	Factors				Case
	Temperature of air supply (A)	Air velocity (B)	Air supply method (C)	Airflow angle (D)	
1	1	1	1	1	$A_1B_1C_1D_1$
2	1	2	2	2	$A_1B_2C_2D_2$

3	1	3	3	3	$A_1B_3C_3D_3$
4	2	1	2	3	$A_2B_1C_2D_3$
5	2	2	3	1	$A_2B_2C_3D_1$
6	2	3	1	2	$A_2B_3C_1D_2$
7	3	1	3	2	$A_3B_1C_3D_2$
8	3	2	1	3	$A_3B_2C_1D_3$
9	3	3	2	1	$A_3B_3C_2D_1$

3 EVALUTION METRICS

To quantify the impacts of air supply modes and parameters on indoor thermal performance, the following evaluation index system was established:

(1) Predicted Mean Vote (PMV): Evaluates global thermal comfort in mechanically ventilated environments, reflecting occupants' subjective perception of thermal conditions.

(2) Draft Rate (DR): A personalized airflow prediction model derived from the Fanger model, quantifying dissatisfaction with localized airflow. This model integrates supply temperature, velocity, and turbulence intensity to assess draft sensitivity in non-uniform thermal environments.

(3) Air Age (AA): Defined as the time required for an air parcel to travel from the supply outlet to a specific indoor location, serving as a key parameter for air freshness. Lower AA indicates superior air quality and higher pollutant dilution efficiency.

(4) Infection Probability (PI): Calculated using the modified Wells-Riley model by Mao et al[19], estimating respiratory virus transmission risk based on indoor pollutant concentration:

$$PI = 1 - \exp[(-Iqpt \cdot C')/(Q \cdot C)] \tag{18}$$

Where I, p, C, t, and q represent infector count, breathing rate, pollutant concentration, exposure time, and infectious dose, respectively.

(5) Energy Utilization Efficiency (E): Quantifies the utilization degree of supply energy via the ratio of supply-to-zone temperature difference to target temperature difference:

$$E = \frac{t_s - t_e}{t_s - t_n} \tag{19}$$

Where t_s is the return air temperature, t_e denotes the supply air temperature and t_n is the ambient temperature.

(6) Comprehensive Score (d_a): A multi-criteria evaluation system based on fuzzy mathematics[20], transforming measured values into normalized degrees of affiliation (da) within [0,1]. Weighted summation (PMV: 0.22, DR: 0.1, AA: 0.18, PI: 0.25, E: 0.25) generates the final score for holistic assessment of thermal comfort, air quality, and energy efficiency.

4 RESULTS AND ANALYSIS

4.1 Intuitive Analysis of a Single Factor

The results of the orthogonal experimental design are summarized in Table 4.

Table 4 Results of Orthogonal Experimental Design

NO.	Factors				Result					
	A	B	C	D	PMV	DR	AA	PI	E	d_a
1	1	1	1	1	0.90±0.00	2.51±0.15	492.06 ±	56.00% ±	153.97 ±	0.44
2	1	2	2	2	0.60±0.03	4.27±0.05	479.31 ±	34.81% ±	142.61 ±	0.64
3	1	3	3	3	0.83±0.01	3.20±0.43	701.09 ±	41.42 %±	178.57 ±	0.54
4	2	1	2	3	1.21±0.00	1.69±0.05	1050.70 ±	43.10 %± 1.22	152.84 ±	0.47
5	2	2	3	1	0.61±0.00	2.29±0.13	314.35 ±	36.48% ±	222.06 ±	0.71
6	2	3	1	2	1.48±0.24	2.43±0.48	688.03 ±	41.97 %±	141.43 ±	0.57

7	3	1	3	2	1.34±0.03	1.34±0.56	421.36 ± 0.74	49.29 %± 1.74	212.70 ± 2.42	0.56
8	3	2	1	3	1.22±0.01	1.56±0.57	448.46 ± 1.87	44.71 %± 0.79	164.85 ± 3.14	0.61
9	3	3	2	1	1.00±0.01	5.24±0.05	305.40 ± 4.10	37.48% ± 0.14	193.96 ± 1.36	0.65

From the trend Figure 4 and range analysis (Table 5)revealed distinct key influencing factors and variation trends across different metrics. For PMV, air supply mode and temperature were dominant factors, with PMV exhibiting a nonlinear trend (initial increase followed by decline) as temperature increased. Displacement ventilation demonstrated superiority by establishing vertical temperature gradients, achieving optimal balance at 1.0 m/s air velocity, with a V-shaped trend for supply angle (optimal scheme: A1B2C3D1). For DR, air velocity and supply mode played pivotal roles: DR increased monotonically with velocity (k-value trend), while mixed ventilation mitigated directional airflow perception via multi-directional jets, with V-shaped temperature influence and decreasing angle effect (optimal scheme: $A_2B_1C_1D_3$). Air age control was governed by supply angle and temperature, showing stepwise k-value increments with angle and a U-shaped temperature trend. Displacement ventilation reduced air age by 32% compared to mixed ventilation (optimal scheme: $A_3B_2C_3D_1$). Infection probability primarily depended on air velocity and supply mode, displaying a valley-shaped velocity trend (k-value), with stratified ventilation minimizing airflow short-circuiting and negligible angle impact (optimal scheme: $A_2B_2C_2D_2$). Energy utilization efficiency was dominated by supply mode (stepwise k-value increase) and temperature (incremental trend), with decreasing angle influence (optimal scheme: $A_1B_1C_1D_3$).

Table 5 Range Analysis Table

Evaluation metrics		A	B	C	D
PMV	K_1	7.000	8.980	10.810	7.550
	K_2	9.890	7.310	8.420	8.880
	K_3	9.320	9.920	6.980	9.780
	k_1	0.778	0.998	1.201	0.839
	k_2	1.099	0.812	0.936	0.987
	k_3	1.036	1.102	0.776	1.087
	R	0.321	0.290	0.426	0.248
	Prioritize factors	C>B>A>D			
	Preferred option	$A_1B_2C_3D_1$			
DR	K_1	29.957	16.614	19.513	30.131
	K_2	19.219	24.378	33.622	24.114
	K_3	24.427	32.611	20.467	19.357
	k_1	3.328	1.846	2.168	3.347
	k_2	2.135	2.708	3.735	2.679
	k_3	2.714	3.623	2.274	2.151
	R	1.193	1.777	1.567	1.197
	Prioritize factors	B>C>A>D			
	Preferred option	$A_2B_1C_1D_3$			
AA	K_1	5017.360	5892.380	4885.640	3335.430
	K_2	6159.250	3726.340	5506.230	4766.100
	K_3	3525.660	5083.550	4310.400	6600.740
	k_1	557.484	654.709	542.849	370.603
	k_2	684.361	414.038	611.803	529.567

	k_3	391.740	564.839	478.933	733.416
	R	292.621	240.671	132.870	362.812
	Prioritize factors	D>A>B>C			
	Preferred option	$A_3B_2C_3D_1$			
PI	K_1	3.967	4.452	4.280	3.904
	K_2	3.651	3.485	3.462	3.782
	K_3	3.945	3.626	3.821	3.877
	k_1	0.441	0.495	0.476	0.434
	k_2	0.406	0.387	0.385	0.420
	k_3	0.438	0.403	0.425	0.431
	R	0.035	0.107	0.091	0.014
	Prioritize factors	B>C>A>D			
	Preferred option	$A_2B_2C_2D_2$			
E	K_1	14.255	15.596	13.808	17.100
	K_2	15.490	15.886	14.683	14.913
	K_3	17.156	15.419	18.410	14.888
	k_1	1.584	1.733	1.534	1.900
	k_2	1.721	1.765	1.631	1.657
	k_3	1.906	1.713	2.046	1.654
	R	0.322	0.052	0.511	0.246
	Prioritize factors	C>A>D>B			
	Preferred option	$A_3C_3D_1B_2$			

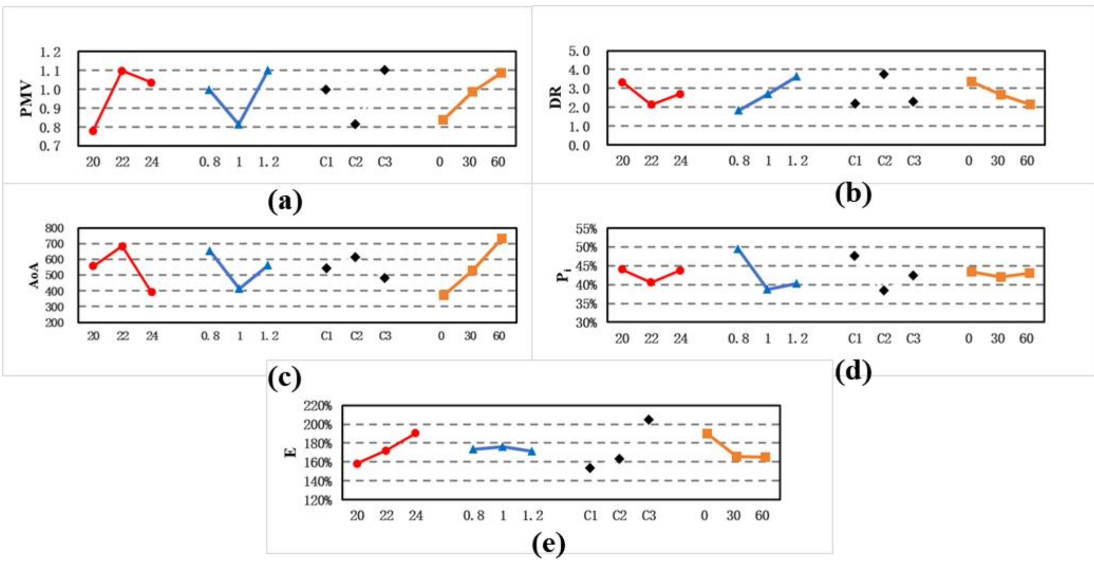


Figure 4 Trend Charts for Each Factor

4.2 Analysis of Variance

The analysis of variance demonstrates distinct factor dominance across performance metrics (Table 6): For PMV, airflow mode (39.2% contribution rate, F=11.34, p<0.001) and supply temperature (21.2%, F=7.10, p=0.005) exert statistically dominant effects, while supply velocity (p=0.016) and angle (p=0.042) show marginal significance, with model reliability confirmed by moderate error variance (31.1%). In DR regulation, supply velocity (29.8%, F=18.86F=18.86) and airflow mode (29.0%, F=18.32) dominate, whereas temperature (F=8.50) and angle (F=8.59) exhibit weaker influences under low experimental noise (14.2% error). Air age control is overwhelmingly governed by supply angle (53.8%, F=344.72) and temperature (33.6%, F=224.42), with velocity (F=154.14) and airflow mode (F=46.02) playing secondary roles in a high-precision framework (1.1% error). Infection probability primarily responds to velocity (58.3%, F=15.08) and airflow mode (36.0%, F=9.32), unaffected by temperature (p=0.208p=0.208) or angle (p=0.800p=0.800) within robust error limits (3.5%). Energy utilization efficiency is dictated by airflow mode (60.0%, F=259.52) and temperature (21.2%, F=92.08), with angle contributing moderately (16.2%, F=70.07) under exceptional precision (2.1% error). These ANOVA outcomes align with prior range analysis, establishing multi-method validation of factor hierarchies across thermal, air quality, and energy domains.

Table 6 Analysis of Variance (ANOVA) Table

Evaluation metrics	Source	*SS*	d_f	*MS*	*F*	*P*	Significance
PMV	A	0.521	2	0.260	7.099	0.005	**
	B	0.388	2	0.194	5.294	0.016	*
	C	0.832	2	0.416	11.339	0.001	***
	D	0.280	2	0.140	3.813	0.042	*
	e	0.660	18	0.037			
DR	A	6.407	2	3.204	8.498	0.003	**
	B	14.220	2	7.110	18.859	0.000	***
	C	13.816	2	6.908	18.323	0.000	***
	D	6.477	2	3.239	8.590	0.002	**
	e	6.786	18	0.377			
AA	A	387588.072	2	193794.036	224.417	0.000	***
	B	266220.527	2	133110.264	154.144	0.000	***
	C	79483.052	2	39741.526	46.021	0.000	***
	D	595369.258	2	297684.629	344.724	0.000	***
	e	15543.803	18	863.545			
PI	A	0.007	2	0.003	1.717	0.208	
	B	0.061	2	0.030	15.084	0.000	***
	C	0.037	2	0.019	9.319	0.002	**
	D	0.001	2	0.000	0.226	0.800	
	e	0.036	18	0.002			
E	A	0.471	2	0.235	92.081	0.000	***
	B	0.012	2	0.006	2.411	0.118	
	C	1.327	2	0.664	259.516	0.000	***
	D	0.358	2	0.179	70.070	0.000	***
	e	0.046	18	0.003			

4.3 Comprehensive Optimization

For the optimization of multi-criteria orthogonal experiments, a comprehensive scoring method was applied to integrate single-criterion intuitive analysis results, with range analysis demonstrating that supply air velocity and temperature predominantly governed the comprehensive evaluation score. Specifically, a supply velocity of 1.0 m/s enhanced air diffusion uniformity by amplifying turbulent motion, thereby optimizing the coupling of thermal plumes and forced convection, while the supply temperature exhibited a marginally increasing trend, with the 24°C high-temperature condition strengthening thermal buoyancy-driven effects to elevate buoyancy flux and improve airflow organization efficiency. In terms of ventilation strategies, displacement ventilation leveraged vertical thermal stratification to reduce air age and enhance air quality, and the decreasing trend in the k-value for supply angles indicated that 0° horizontal supply prolonged the jet core length through jet attachment effects, optimizing airflow coverage in the occupied zone. By synthesizing the influence patterns and significance of factors, the optimal configuration was determined as follows: supply velocity of 1.0m/s (B_3), supply temperature of 24°C (A_3), displacement ventilation (C_3), and supply angle of 0° (D_1).

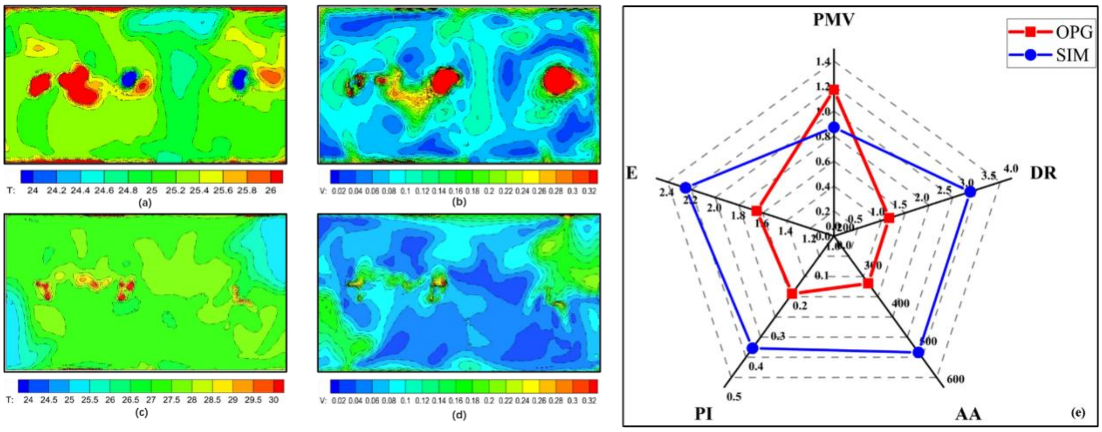


Figure 5 Comparison between the Optimized Group and the Control Group: (a) cloud diagram of temperature distribution at Z=1.4m in the control group; (b) cloud diagram of velocity distribution at Z=1.4m in the control group; (c) cloud diagram of temperature distribution at Z=1.4m in the optimized group; (d) cloud diagram of the velocity distribution at Z=1.4m in the optimized group; and (e) radar chart of comparison of the various indexes.

According to the numerical simulation results in Figure 5, the optimized group demonstrates significantly improved uniformity in both temperature and velocity fields at the Z=1.4m plane compared to the control group. At this height, the optimized group exhibits a mean temperature of 27.14°C (control: 25.41°C) and a slightly higher average airflow velocity (+0.02 m/s). Key performance metrics reveal that while the Predicted Mean Vote (PMV) of the optimized group increases by 17% to 1.17 (control: 0.87), indicating a marginally warmer thermal perception, critical improvements are achieved: Draft Rate (DR) decreases by 59.5% (from 3.06 to 1.24), Air Age (AA) shortens by 35.98% (from 507s to 325s), Infection Probability (PI) drops drastically by 98.05% (from 19% to 0.37%), and Energy Utilization Efficiency (E) improves by 36% (from 164.67% to 224.52%). The comprehensive evaluation score rises by 58.95%, demonstrating that the optimized design significantly enhances air quality, infection control efficacy, and energy efficiency while maintaining thermal comfort within acceptable limits.

5 CONCLUSION

This study investigates the impact of ventilation modes and air supply parameters on five thermal performance indicators in summer office buildings through orthogonal experimental design, computational fluid dynamics (CFD) simulations, and experimental validation. The main conclusions are as follows:

1) The standard k-ε, RNG k-ε, and Realizable k-ε models effectively predict indoor airflow characteristics. Among these, the Realizable k-ε model exhibits marginally superior accuracy in simulating air movement within summer office environments.

2) Range and variance analyses reveal distinct dominant factors across thermal performance metrics: PMV and *E* are predominantly governed by air supply mode and temperature. *DR* and infection probability are significantly influenced by air supply velocity and mode. Air age is synergistically controlled by *supply angle* and *temperature*.

3) By constructing a comprehensive evaluation model, the optimal parameter combination under displacement ventilation is determined: supply temperature = 24°C, velocity = 1 m/s, and angle = 0°. Compared to the baseline scenario, this optimized configuration achieves: PMV improvement by 17% (remaining within acceptable thermal comfort thresholds), DR reduction by 59.54%, Air age shortening by 35.89%, Infection probability decrease by 47.57%, Energy efficiency enhancement by 36.34%, Comprehensive evaluation score elevation by 58.95%.

These results demonstrate the synergistic optimization of thermal comfort, air quality, and energy efficiency in office environments.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Tsinghua University Building Energy Efficiency Research Center. China Building Energy Efficiency Annual Development Research Report 2022 Public Buildings. Beijing: China Building Industry Press, 2022.

[2] Zhong Wei. Current situation of energy consumption of central air conditioning in China's buildings and comprehensive energy saving measures. Chongqing University, 2004.

[3] Krajčík M, Simone A, Olesen B W. Air distribution and ventilation effectiveness in an occupied room heated by warm air. Energy and Buildings, 2012, 55: 94-101.

[4] Fisk W J, Faulkner D, Sullivan D, et al. Air change effectiveness and pollutant removal efficiency during adverse mixing conditions. Indoor Air, 1997, 7(1): 55-63.

[5] P F Linden. The fluid mechanics of natural ventilation, Annu. Rev. Fluid Mech. 1999, 31, 201-238.

[6] Sandberg M, Blomqvist C. Displacement ventilation systems in office rooms. 1989.
[7] Lin Z, Chow T T, Tsang C F, et al. Stratum ventilation–a potential solution to elevated indoor temperatures. Building and Environment, 2009, 44(11): 2256-2269.
[8] Cheng Y, Lin Z, Fong A M L. Effects of temperature and supply airflow rate on thermal comfort in a stratum-ventilated room. Building and Environment, 2015, 92: 269-277.
[9] Tian L, Lin Z, Wang Q. Comparison of gaseous contaminant diffusion under stratum ventilation and under displacement ventilation. Building and Environment, 2010, 45(9): 2035-2046.
[10] Cheng Y, Lin Z. Experimental study of airflow characteristics of stratum ventilation in a multi- occupant room with comparison to mixing ventilation and displacement ventilation. Indoor air, 2015, 25(6): 662-671
[11] Chen H, Feng Z, Cao S J. Quantitative investigations on setting parameters of air conditioning (air-supply speed and temperature) in ventilated cooling rooms. Indoor and Built Environment, 2021, 30(1): 99-113.
[12] Lin Z, Chow T T, Tsang C F, et al. Effect of air supply temperature on the performance of displacement ventilation (Part I)-thermal comfort. Indoor and Built Environment, 2005, 14(2): 103-115.
[13] Zhang Lin T T, Chow C F, Tsang K F, et al. Effect of air supply temperature on the performance of displacement ventilation (Part II)-Indoor air quality. Indoor and Built Environment, 2005, 14(2): 117-131.
[14] Liu Z, Yang H, Liu Z, et al. Experimental Study on the effects of layered air supply angle and air supply temperature on indoor environment. Procedia Engineering, 2017, 205: 213-218.
[15] Xu Jian. Optimization analysis of layer ventilation airflow organization in office buildings. Nanjing Normal University, 2016.
[16] Gao Shuzhen, Huang Mingming, Liu Yi. Calculation chart of human surface area. Qingdao Medicine and Health, 1995(04): 55-56.
[17] Di Gilio A, Palmisani J, Pulimeno M, et al. CO2 concentration monitoring inside educational buildings as a strategic tool to reduce the risk of Sars-CoV-2 airborne transmission. Environmental research, 2021, 202: 111560.
[18] Pless D, Keck T, Wiesmiller K, et al. Numerical simulation of air temperature and airflow patterns in the human nose during expiration. Clinical Otolaryngology & Allied Sciences, 2004, 29(6): 642-647.
[19] Mao Y, Ma J, Wang S, et al. A stratum ventilation system for pollutants and an improved prediction model for infection in subway cars. Atmospheric Pollution Research, 2022, 13(3): 101354.
[20] Liu F, Weng Miaocheng. Experimental design and data processing. Chongqing: Chongqing University Press, 2021.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3085

THE DECOUPLING OF SELF-ADJOINT SECOND-ORDER SYSTEMS WITH STRUCTURE-PRESERVING TRANSFORMATION

Nan Jiang, QiZhi Zhang*
Northeast Forestry University, Harbin 150040, Heilongjiang, China.
Corresponding Author: QiZhi Zhang, Email: zqz_math@nefu.edu.cn

Abstract: This paper focuses on the decoupling of self-adjoint second-order linear systems. It is proposed a method with congruence transformation matrix that preserving the Lancaster structure. This method requires fewer parameters during the decoupling while remaining the spectrum of the system before and after decoupling. Numerical simulation experiments demonstrate the implementation results achieved by applying this method to system decoupling.
Keywords: Self-adjoint; Second-order systems; Decoupling; Lancaster structure

1 INTRODUCTION

Typically, dynamic systems inherently possess highly intricate structures. It is defined as a coupled system when the system consists of interconnected subsystems. Coupled systems find extensive applications across diverse domains such as aerospace technology, economic development, and agricultural production[1-3]. However, the complex configurations of coupled system devices and apparatuses, combined with interactions and mutual influences among all or partial subsystems, introduce numerous uncertain, complex, and difficult-to-control factors during system analysis. These challenges ultimately hinder the effective optimization of system performance[4-8]. Currently, to address the challenges in system analysis caused by mutual interference among subsystems in coupled systems, engineers primarily employ system decoupling methods. These approaches preserve the intrinsic properties of the original coupled system while eliminating all or partial coupling relationships, thereby transforming a multi-degree-of-freedom system into multiple independent single-degree-of-freedom subsystems without mutual interactions. Given that higher-order linear coupled systems can be effectively reduced to second-order linear coupled systems through appropriate treatment, this study focuses on the decoupling of self-adjoint second-order linear dynamical systems. We present a congruence transformation of decoupling with preserving the Lancaster structure, accompanied by a series of numerical simulation experiments.

2 TRANSFORMATION FOR SELF-ADJOINT SECOND-ORDER LINEAR SYSTEMS

The second-order linear dynamical systems is represented by differential equations: $M_0\ddot{\boldsymbol{x}}(t)+C_0\dot{\boldsymbol{x}}(t)+K_0\boldsymbol{x}(t)=\boldsymbol{f}(t)$, where M_0, C_0 and K_0 are respectively the given initial coefficient matrices of an n-degree-of-freedom system, $\boldsymbol{x}(t)$ and $\boldsymbol{f}(t)$ are vectors with n degrees of freedom[9]. Such systems are extensively utilized in across critical disciplines such as applied mechanics, acoustics, circuit simulation, structural mechanics, fluid mechanics, and microelectronics design.

To perform decoupling on the coefficients of the second-order linear dynamical system, we first denote the Lancaster structure composed of the coefficient matrices as

$$\mathrm{L}(\lambda)=\mathrm{L}(\lambda;M_0,C_0,K_0)=B_0\lambda+A_0 \tag{1}$$

where

$$B_0=\begin{bmatrix} C_0 & M_0 \\ M_0 & 0 \end{bmatrix},\ A_0=\begin{bmatrix} K_0 & 0 \\ 0 & -M_0 \end{bmatrix}$$

the parameter λ represents spectrum of this system. The parameterized governing equations of the system are formulated as follow

$$M(t)\ddot{\boldsymbol{x}}(t)+C(t)\dot{\boldsymbol{x}}(t)+K(t)\boldsymbol{x}(t)=\boldsymbol{f}(t) \tag{2}$$

Then, denote the Lancaster structure[10] corresponding to the parameterized system (2) is as

$$\mathrm{L}(\lambda;M(t),C(t),K(t))=B(t)\lambda+A(t) \tag{3}$$

where

$$B(t)=\begin{bmatrix} C(t) & M(t) \\ M(t) & 0 \end{bmatrix},\ A(t)=\begin{bmatrix} K(t) & 0 \\ 0 & -M(t) \end{bmatrix} \tag{4}$$

Since the spectrum is critically important for determining the motion displacement and vibrational states of second-order linear systems, therefore, the spectrum of the parameterized system remains invariant. Additionally, all coefficient matrices are required to be differentiable with respect to the time parameter λ.

There must exist a set of $2n\times 2n$ invertible parameter-dependent real transformation $\Pi_L(t)$ and $\Pi_R(t)$, that ensure

$$\begin{cases} B(t)=\Pi_L^{\mathrm{T}}(t)B_0\Pi_R(t) \\ A(t)=\Pi_L^{\mathrm{T}}(t)A_0\Pi_R(t) \end{cases} \tag{5}$$

By differentiating both sides of Equations (5) with respect to the parameter λ, we obtain

$$\begin{cases} \dot{B}(t)=L^{\mathrm{T}}(t)B(t)+B(t)R(t) \\ \dot{A}(t)=L^{\mathrm{T}}(t)A(t)+A(t)R(t) \end{cases} \tag{6}$$

$L(t)$ and $R(t)$ are the introduced condition matrix. By reorganizing the aforementioned transformations, we derive the following system of equations comprising $5n^2$ equations with $8n^2$ unknowns

$$\begin{aligned} & L_{12}^{\mathrm{T}}(t)M(t)+M(t)R_{12}(t)=0 \\ & K(t)R_{12}(t)-L_{21}^{\mathrm{T}}(t)M(t)=0 \\ & L_{12}^{\mathrm{T}}(t)K(t)-M(t)R_{21}(t)=0 \\ & L_{11}^{\mathrm{T}}(t)M(t)-L_{22}^{\mathrm{T}}(t)M(t)+C(t)R_{12}(t)=0 \\ & M(t)R_{11}(t)-M(t)R_{22}(t)+L_{12}^{\mathrm{T}}(t)C(t)=0 \end{aligned} \tag{7}$$

A second-order linear system is defined as self-adjoint when all its coefficient matrices are symmetric. When the initial coefficient matrices are all symmetric matrices, we should require preserving the symmetry of the coefficient matrices during decoupling. Specifically, matrices $M(t)$, $C(t)$ and $K(t)$ must retain their symmetric properties for any time parameter t. Introducing three $n\times n$ parameter-dependent matrices $D(t)$, $N_L(t)$ and $N_R(t)$, where $D(t)$ is a skew-symmetric matrix. This ensures that the symmetry of all coefficient matrices is preserved. We let $N_L^{\mathrm{T}}(t)=N_R(t)=N(t)$, $\Pi_L(t)=\Pi_R(t)=T(t)$. At this stage, the number of parameter matrices is reduced to two, and the number of independent parameters is decreased to $2n^2$.

Impose a constraint on the self-adjoint second-order linear system by normalizing the initial mass matrix M_0 to the identity matrix. Simultaneously, to preserve the self-adjoint property, the following constraint is imposed $M(t)\equiv I$, i.e. $\dot{M}(t)=0$. Consequently, the system (7) can be simplified to

$$\begin{aligned} & K(t)R_{12}(t)-R_{21}^{\mathrm{T}}(t)M(t)=0 \\ & R_{12}^{\mathrm{T}}(t)M(t)+M(t)R_{12}(t)=0 \\ & M(t)R_{11}(t)-M(t)R_{22}(t)+R_{12}^{\mathrm{T}}(t)C(t)=0 \end{aligned} \tag{8}$$

where the coefficient matrix is

$$R(t)=\begin{bmatrix} D(t) & 0 \\ 0 & D(t) \end{bmatrix}\begin{bmatrix} -\dfrac{C(t)}{2} & -M(t) \\ K(t) & \dfrac{C(t)}{2} \end{bmatrix}+\begin{bmatrix} N(t) & 0 \\ 0 & N(t) \end{bmatrix} \tag{9}$$

Now the number of unknowns is reduced to $n(n-1)/2+n^2$. To further simplify the number of parameters, a skew-symmetric matrix is introduced that satisfied the following condition

$$N(t)-N^{\mathrm{T}}(t)=2S(t) \tag{10}$$

Matrix $S(t)$ serves as the new parameter matrix capable of replacing the original parameter matrix $N(t)$. After substitution, matrix $N(t)$ can be expressed as a combination of two skew-symmetric matrices $D(t)$ and $S(t)$

$$N(t)=\frac{1}{4}\big(C(t)D(t)-D(t)C(t)\big)+S(t) \tag{11}$$

Since the parameter matrices are now both skew-symmetric, the number of parameters has been simplified to $n(n-1)$. Furthermore, the derivative equation of the congruence transformation matrix $T(t)$ with respect to time t is derived as follows

$$\dot{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} \frac{1}{4}CD - \frac{3}{4}DC + S & -D \\ DK & \frac{1}{4}CD + \frac{1}{4}DC + S \end{bmatrix} \tag{12}$$

Herein, the description of parameter t is omitted, with $T(t)$ defined as the derivative of matrix $\dot{T}$ with respect to time parameter t. By appropriately selecting only $n(n-1)$ parameters from the skew-symmetric matrices $D(t)$ and $S(t)$, the transformation $T(t)$ can be obtained. Subsequently, a congruence transformation is applied to the Lancaster structure using the transformation $T(t)$. Through suitable optimal control methods, the trajectories of matrices $M(t)$, $C(t)$ and $K(t)$ are tracked at different time instants t, thereby driving these matrices progressively toward diagonal structures. A self-adjoint second-order linear system is said to achieve complete decoupling, at a specific time instant t, all transformed coefficient matrices are diagonal matrices.

3 NUMERICAL SIMULATION EXPERIMENTS

A numerical simulation verification is provided for the structure-preserving congruence transformation solving algorithm targeting self-adjoint second-order linear systems. The experiments were implemented using MATLAB programming, where the iterative integration method invokes MATLAB's standard ODE integrators. Both the absolute error tolerance (AbsTol) and relative error tolerance (RelTol) in the simulation program were set to 10^{-10}, with the output data retaining four significant decimal places.

Given a self-adjoint second-order linear system with 4 degrees of freedom, its initial mass matrix $M_0 = I$, initial damping C_0 and stiffness matrices K_0 are defined as follows

$$C_0 = \begin{bmatrix} 0.4108 & 0.0000 & -0.3529 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ -0.3529 & 0.0000 & 2.1661 & -1.8132 \\ 0.0000 & 0.0000 & -1.8132 & 1.8132 \end{bmatrix}$$
$$K_0 = \begin{bmatrix} 22.3480 & -9.3547 & -8.9365 & 0.0000 \\ -9.3547 & 18.5240 & -9.1690 & 0.0000 \\ -8.9365 & -9.1690 & 22.2080 & -4.1027 \\ 0.0000 & 0.0000 & -4.1027 & 4.1027 \end{bmatrix} \tag{13}$$

By employing the transformation $T(t)$ solving algorithm that maintains the Lancaster structure, appropriate numerical integration iterations are performed on the time parameter t. The integration process terminates at $t \approx 4.0\times10^2$, and the damping matrix and stiffness matrix of the decoupled system are output as follows

$$C = \begin{bmatrix} 0.2073 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 2.2029 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.9436 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0363 \end{bmatrix}$$
$$K = \begin{bmatrix} 29.5482 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 6.0921 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 27.5510 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.8435 \end{bmatrix} \tag{14}$$

In this case, the decoupled mass matrix M is the identity matrix. As evidenced by the output data, the damping matrix C and stiffness matrix K are perfectly diagonal matrices. The total decoupled system represents four independent single-degree-of-freedom subsystems, where the damping coefficient and stiffness coefficient of each subsystem are given by the diagonal elements of matrices C and K, respectively.

The variation in the sum of off-diagonal elements for matrices C and K during the complete system decoupling process is graphically illustrated in Figure 1.

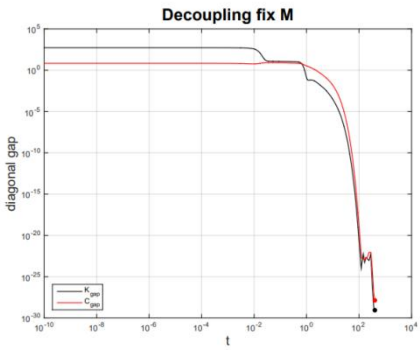


Figure 1 The Evolution of Off-Diagonal Portions of Matrices *C* and *K*

When the integration begins, the curves representing the sum of off-diagonal elements of the damping matrix and stiffness matrix start to descend at identical rates. Upon termination of the integration, both curves stabilize near the order of magnitude of 10^{-30}. Since the error tolerance in the program was set to 10^{-10}, this confirms that the system has been totally decoupled.

To validate spectrum invariance, the spectrum of the quadratic eigenvalue problem (QEP) associated with the total decoupled system is computed. The absolute error between the spectrum at each time parameter t and the initial spectrum is selected as the metric for spectral variation. The evolution of spectral variation during the decoupling is graphically illustrated in Figure 2.

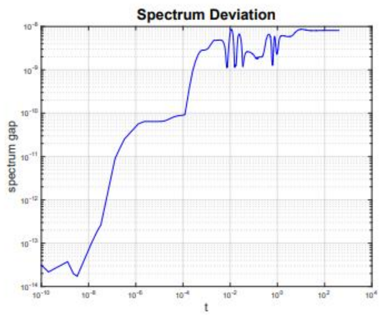


Figure 2 The Deviations of Spectrum

As illustrated in Figure 2, the spectrum of the totally decoupled system has changed. When the integration commences, spectral variations emerge, with their magnitude progressively increasing as the integration proceeds. Upon integration termination, the magnitude of spectral variations stabilizes near the order of 10^{-8}. Since the error tolerance in the numerical simulation was set to 10^{-10}, these spectral variations can be considered negligible.

4 CONCLUSION

Decoupling of second-order linear systems constitutes a pivotal challenge in system analysis. This paper focuses on self -adjoint second-order linear systems. We proposed a structure-preserving congruence transformation solving this problem. Leveraging the isospectral theory of Lancaster structure preservation for quadratic eigenvalue problems (QEPs) inherent to such systems, the algorithm constructs a derivative equation of the congruence transformation matrix with respect to the time parameter. Through appropriate numerical integration of this derivative equation, a congruence transformation matrix is obtained at a specific time, which nullifies the sum of non-prespecified off-diagonal elements in the coefficient matrices. Subsequent application of this matrix to transform the coefficient matrices achieves complete decoupling of the self-adjoint second-order linear system.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work is supported by the Fundamental Research Funds for the Central Universities (Grant No. 2572021BC08).

REFERENCES

[1] Zhang L P, Yang Z D, Qu Z Y, et al. Modeling and vibration decoupling control of multi-axial shaking table. Journal of Southwest Petroleum University, 2015, 409: 1878-1883.

[2] Zheng Y, Zhou Z C, Huang H. A multi-frequency MIMO control method for the 6DOF micro-vibration exciting system. Acta Astronautica, 2020, 170: 552-569.
[3] Madden B, Florin N, Mohr S, et al. Using the waste Kuznet's curve to explore regional variation in the decoupling of waste generation and socioeconomic indicators. Resources Conservation and Recycling, 2019, 149: 674-686.
[4] Chen W, Zhao H B, Li J F, et al. Land use transitions and the associated impacts on ecosystem services in the Middle Reaches of the Yangtze River Economic Belt in China based on the geo-informatic Tupu method. The Science of the Total Environment, 2020, 701: 134690. DOI:10.1016/j.scitotenv.2019.134690.
[5] Ravneet K, Omkar J, Rodney W E. The ecological and economic determinants of eastern redcedar (Juniperus virginiana) encroachment in grassland and forested ecosystems: A case study from Oklahoma. Journal of Environmental Management, 2020, 254: 109815. DOI:10.1016/j.jenvman.2019.109815.
[6] Sanye-Mengual E, Secchi M, Corrado S, et al. Assessing the decoupling of economic growth from environmental impacts in the European Union: A consumption-based approach. Journal of Cleaner Production, 2019, 236: 117535. DOI:10.1016/j.jclepro. 2019.07.010.
[7] Kumar S, Mahulikar S P. Aero-thermal analysis of lifting body configurations in hypersonic flow. Acta Astronautica, 2016, 126: 382-394.
[8] Li S B, Wang Z G, Barakos G N, et al. Research on the drag reduction performance induced by the counterflowing jet for waverider with variable blunt radii. Acta Astronautica, 2016, 127: 120-130.
[9] Li Anming. Modeling and simulation of dynamic systems. Beijing: National Defense Industry Press, 2012: 29-53.
[10] Lancaster P. Linearization of regular matrix polynomials. The Electronic Journal of Linear Algebra, 2008, 17: 21-27.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3088

ANALYSIS OF CROP PLANTING STRATEGIES USING IMPROVED SIMULATED ANNEALING OPTIMIZATION ALGORITHM

Bin Cheng[1], Xiang Gao[2], JiaWei Zhao[1*], KunLin Yu[1]
[1]School of Physics and Mechanics, Wuhan University of Technology, Wuhan 430070, Hubei, China.
[2]School of Management, Wuhan University of Technology, Wuhan 430070, Hubei, China.
Corresponding Author: JiaWei Zhao, Email: 342528@whut.edu.cn

Abstract: Given the actual needs of farmers and the need to protect the environment, selecting suitable crop optimization planting strategies is of great significance. Based on the analysis of the impact of different types of land parcels on crop yield, crop rotation demand, and market demand changes, this article optimizes the allocation of arable land resources through an improved simulated annealing algorithm. Determine the indicators based on the given data and use the Stackelberg model to calculate the selling price, establish a single objective optimization model, and solve for the optimal planting results through an improved simulated annealing algorithm. Firstly, visualize the data to understand the features and determine the indicators, and use the Stackelberg model to calculate the selling price. Secondly, construct a single objective optimization model based on indicators and constraints. Finally, in order to simplify the calculation and improve the accuracy of the optimal solution, the plot is divided into four parts, and a preliminary solution is obtained through simulated annealing algorithm. Then, an adaptive threshold is introduced to improve the optimal planting strategy. This indicates that the improved simulated annealing algorithm can obtain the optimal solution for multiple crop planting strategies under different unsold models, demonstrating that the improved simulated annealing algorithm has made good progress in solving related crop planting strategies.
Keywords: Optimization strategies for crop cultivation; Improve simulated annealing algorithm; Single objective optimization model; Optimal planting strategy

1 INTRODUCTION

As a large agricultural country with high cross regional, China has complex types and quantities of farmland plots. Under the background of the deepening of agricultural refinement and marketization, it is of great significance to realize the development of rural economy and revitalize the countryside how to use different types of farmland and multiple crops for rational distribution in order to obtain the maximum economic benefits.

A large number of studies have been carried out at home and abroad on the optimization of annual crop planting strategies. Not only the optimization methods are designed, but also the greedy heuristic algorithm is used to solve the problem of crop planting in the division of management areas in specific locations[1]; Some scholars use the improved neural network model to carry out a reasonable overall planting strategy for crops[2]; At the same time, the comparative advantage model is constructed for crop planting strategies in different geographical regions[3]; J. B.nixon and other scholars studied the optimal crop planting strategy in Western Kansas through the decision support system (dssat-csm) model. At the same time, some scholars have studied, such as corn[4], wheat[5], rice[6]. In addition, multi crop coupling research[7]. Rabbani M scholars use the entropy weight method-TOPSIS method combined with a two-stage optimization approach to optimize crop planting strategies[8]; Wu Hui scholars use ArcGIS spatial analysis and other methods to explore the changes in optimizing the relationship between water[9], land, energy, economy, environment, and food; simultaneously, Wang Chang and others scholars conduct research on the optimal planting strategy based on the NSGA-II algorithm[10]; Bellangue D emphasizes that the development of ranch economy is driven by integrating market prices to achieve the optimal crop planting strategy[11].

Through literature review and theoretical research, it is believed that the optimization of agricultural planting system forms a complex coupling network in the interaction of crops, cultivated land types and previous crops. Existing studies have confirmed that the scientific allocation of planting proportion of different crops can significantly improve the output value per unit area, but there are still research gaps in the spatial heterogeneity of cultivated land types, the dynamic matching mechanism of previous crops and the global optimization algorithm under multiple constraints. This paper constructs a multi-dimensional classification system based on multiple types of crops and heterogeneous cultivated land types, which breaks through the traditional single category classification model. By establishing a single objective optimization model with constraints of land carrying capacity and crop rotation, and innovatively combining neighborhood search strategy with annealing mechanism, an adaptive threshold algorithm with dynamic step size adjustment is designed.

2 MULTI CROP PLANTING PLAN AND INDICATOR CONSTRUCTION

2.1 Establishment of Farmland Types and Agricultural Product Production Indicators

The type of cultivated land, as a limiting factor for agricultural production and a major influencing factor for planting selection, results in different types of cultivated land cultivating different crops [7]. By preprocessing the data, we construct indicators for farmland types and agricultural product production:

2.1.1 Farmland type indicator (S_k)

The cultivated land type index classifies cultivated land into six types: flat dry land (A), terraced fields (B), hillside land (C), irrigated land (D), as well as ordinary greenhouses (E) and smart greenhouses (F). And different types of cultivated land produce different products, so cultivated land is represented as flat dry land S_1, terraced land S_2, hillside land S_3, irrigated land S_4, ordinary greenhouse S_5, smart greenhouse S_6, and the formula is given:

$$S_1 = [s_{1,1}, s_{1,2} \ldots s_{1,6}] \tag{1}$$

The above six types of cultivated land are represented by S_k; $s_{k,n}$ represents the nth plot of land in the k category.

2.1.2 Agricultural product planting rate index (X_k (t))

The planting rate of agricultural products is a numerical reflection of the value of intercropping. Referring to the influence of intercropping of agricultural products, the planting rate index of agricultural products is used to determine the quantity of agricultural products planted. Among them, flat dry land (A), terraced land (B), and hillside land (C) are planted once a year, so their crop planting indicators meet the following requirements:

$$\sum_{j=1}^{l_k} x_{k,i,j}(t) = 1, (k = 1,2,3) \tag{2}$$

Among them, l_k is the kth cultivated land type that can be used as the number of crops for planting; $X_{k,i,j}(t)$ are the kth cultivated land type and the jth agricultural product planting rate of the i land. The ordinary greenhouse (E) and the smart greenhouse (F) are planted twice a year, and there is a significant difference in the second quarter between the ordinary greenhouse and the smart greenhouse. Therefore, they respectively meet the following requirements:

$$\sum_{j=1}^{l_{k,1}} x_{k,i,j,1}(t) \leq 1 \quad (k = 5) \tag{3}$$

$$\sum_{j=1}^{l_{k,2}} x_{k,i,j,2}(t) \leq 1 \quad (k = 5) \tag{4}$$

$$\sum_{j=1}^{l_k} x_{k,i,j}(t) \leq 2 \quad (k = 6) \tag{5}$$

Among them, $l_{k,1}$ is the type of crop planted in the first season of Class E farmland; $l_{k,2}$ As a type of crop planted in the second season of Class E arable land.

$$\sum_{j=1}^{l_k} x_{k,i,j}(t) \leq 1 \quad (k = 4), \tag{6}$$

Among them, $X_k(t)=[x_{k,1}(t), x_{k,2}(t) \ldots x_{k,i}(t)]$ is used as the indicator of agricultural product planting rate.

2.1.3 Agricultural product yield indicator (S_kH_k (t))

The yield per mu of agricultural products reflects the production volume of agricultural products. Due to the consideration of intercropping, the yield per mu of a single agricultural product may not accurately reflect the actual planting capacity. Therefore, the agricultural product planting rate index is used to solve the problem:

$$H_k(t) = X_k(t)N_k(t) \tag{7}$$

Among them, N (k) is the crop yield per mu of the k type of land in the t year. So the total yield $S_kH_k(t)$ of each cultivated land type can be calculated as an indicator of agricultural product yield.

2.1.4 Cost index per mu of agricultural products (α_k(t))

The cost per acre of agricultural products is used as production cost, which is an additional expense incurred during the production of agricultural products. It has already been given in the question, so definition $\alpha_{k,i,j}(t)$. Using a single agricultural product cost index for matrix transformation as $\alpha_k(t)$ to obtain the cost index per mu of agricultural products.

2.2 Establishment of Sales Prices and Expected Sales Volume Indicators for Agricultural Products

The selling price and sales volume of agricultural products are the source of farmers' income and the only way in which agricultural products can provide value. As an important part of determining the sales price and sales volume of agricultural products and constructing crop planting strategies.

2.2.1 Expected sales volume indicator of agricultural products (D (t))

The expected sales volume of agricultural products is an important basis for producing agricultural products, and it is also a reasonable control of the production quantity of agricultural products during production; Using the quantity of agricultural products produced in 2023 as the expected sales volume for agricultural products for calculation:

$$d_j(t) = \sum_{i=1}^{81} h_{k,i,j} \tag{8}$$

Where $d_j(t)$ is the expected sales volume of the j crop planted in the t year; $H_{k,i,j}$ are the quantities of the *j* crop grown on the *i* land of the *k* type of cultivated land in 2023. Using $d_j(t)$ to form a new matrix function yields:

$$D(t) = [d_1(t), d_2(t), \ldots, d_j(t)] \tag{9}$$

As the final calculation function for the expected sales volume of agricultural products.

2.2.2 Agricultural product sales price index (P_k(t))

The selling price of agricultural products, as a floating price, is often difficult to measure intuitively. Using the Stackelberg game theory, a game analysis is conducted on consumers and producers of agricultural products to obtain a price function.

Considering that consumers desire lower prices for agricultural products; Being a leader in the agricultural product market. Using the buyer's price acceptance ability as the pricing standard, referring to the commonly used value λ=0.8 as the purchasing function, the sales price is obtained:

$$U(p_{k,j}(t)) = p_{k,j,max}(t) - 0.8(p_{k,j,max}(t) - p_{k,j,min}(t)) \quad (10)$$

Among them, $U(\cdot)$ is the utility function used as a price indicator for calculation; $p_{k,j,max}(t)$ is the highest price per kilogram of the *j* crop planted in the *k* type of cultivated land in the *t* year; $p_{k,j,min}(t)$ is the lowest price per kilogram of the *j* crop planted in the *k* type of cultivated land in the *t* year. Substitute the above formula into the price matrix to obtain $P_k(t)$ as the indicator of agricultural product sales price:

$$P_k(t) = [p_{k,1}(t), p_{k,2}(t), \dots p_{k,j}(t)] \quad (11)$$

3 CONSTRUCTION OF OBJECTIVE FUNCTION FOR CROP PLANTING STRATEGY

3.1 Construction of a Model for Unsold Waste

The type of cultivated land, as a limiting factor for agricultural production and a major influencing factor for planting selection, results in different types of cultivated land cultivating different crops. By preprocessing the data, we construct indicators for farmland types and agricultural product production:

The unsold waste model is a control strategy for agricultural product yield based on expected sales volume indicators. The following are the steps for building a function:

Step 1. Construction of crop profit function

Determine the planting rate and production cost of each crop to establish a simple crop profit function:

$$Q_k = S_k H_k(t) P_k(t) - S_k X_k(t) \alpha_k(t) \quad (12)$$

Where Q_k represents the profit amount of the k type of farmland; $S_k H_k(t)$calculates the crops produced on the k type of farmland. Further accumulation of the function gives the profit amount for each type of farmland:

$$Q = \sum_{k=1}^{6} [S_k H_k(t) P_k(t) - S_k X_k(t) \alpha_k(t)] \quad (13)$$

Among which$\sum_{k=1}^{6} S_k H_k(t)$refers to all crops produced from the cultivated land.

Step 2. Calculation of Unsold Waste

By calculating based on the number of unsold and wasted items:

$$u_j(t) = max[(\sum_{k=1}^{7} S_k H_k(t))_j - d_j(t), 0] \quad (14)$$

Among them, $u_j(t)$ is the excessive planting amount of the crop at time t; Max [f, 0] is the judgment function, and when f>0, the value of *f* is taken; On the contrary, take 0 as the numerical value; Substitute $U_j(t)$:

$$U(t) = [u_1(t), u_2(t) \dots u_{41}(t)] \quad (15)$$

U (t) is the stagnant sales volume of agricultural products in year t.

Step 3. Introduction of the unsold waste function

The profit function for crops will deduct the excess portion of the crops.

$$Q(t) = \sum_{k=1}^{6} (S_k H_k(t) P_k(t) - S_k H_k(t) \alpha_k(t)) - U(t) P(t). \quad (16)$$

Thus, the preliminary unsold waste function is constructed.

Step 4. Limiting Function Determination

The restrictions on crops are mainly divided into planting season restrictions and restrictions on crop rotation and the planting of leguminous crops. The planting season restrictions have been addressed in the agricultural planting rate indicators. Now, the restrictions on crop rotation and the planting of leguminous crops are being established:

$$\sum_{j=1}^{l_k} x_{k,i,j,1}(t) + x_{k,i,j,2}(t) \le 1 \quad (k = 6) \quad (17)$$

$$\sum_{j=1}^{l_k} x_{k,i,j}(t) + x_{k,i,j}(t+1) \le 1 \quad (k = 1,2,3)\,. \quad (18)$$

$$\sum_{t=n}^{n+2} \sum_{j=1}^{l_{k,1}} s_{k,i}\, x_{k,j}(t) \ge 1 \quad (19)$$

Among them, l_k represents the number of crops that can be planted on the *k* type of farmland; $x_{k,i,j,1}(t)$ is the yield of the *j* type of crop in the first season for the *i* plot of the *k* type of farmland at time *t*; $x_{k,i,j,2}(t)$ is the yield of the *j* type of crop in the second season for the *i* plot of the *k* type of farmland at time *t*. $l_{k,1}$ represents the number of the first *j* types of leguminous crops for the *k* type; $x_{k,j}(t)$ is the planting rate of crop *j* in the *k* type of farmland at year *t*. Considering that it is not advisable to disperse a crop too much within a single plot, a concentration index is used to manage the degree of dispersion:

$$\beta_{k,j}(t) = \frac{\sum_{i=1}^{n_k} (s_{k,i} x_{k,j}(t))^2}{(\sum_{i=1}^{n_k} s_{k,i} x_{k,j}(t))^2} \ge 0.6 \quad (20)$$

Among them, $\beta_{k,j}(t)$ is the dispersion of type j crops in class k cultivated land at time t. Finally, the model of unsold waste is obtained mould:

$$max \quad W = \sum_{t=1}^{7} Q(t) = \sum_{t=1}^{7}\sum_{k=1}^{6}[S_k H_k(t) P_k(t) - S_k X_k(t)\alpha_k(t)] - \lambda U(t)P(t).$$

$$s.t\begin{cases} \sum_{j=1}^{l_k} x_{k,i,j}(t) = 1 (k=1,2,3) \\ \sum_{j=1}^{l_{k,1}} x_{k,i,j,1}(t) \le 1 \quad (k=5) \\ \sum_{j=1}^{l_{k,2}} x_{k,i,j,2}(t) \le 1 \quad (k=5) \\ \sum_{j=1}^{l_k} x_{k,i,j}(t) \le 2 \quad (k=6) \\ \sum_{j=1}^{l_k} x_{k,i,j,1}(t) + x_{k,i,j,2}(t) \le 1 \quad (k=6) \\ \sum_{j=1}^{l_k} x_{k,i,j}(t) + x_{k,i,j}(t+1) \le 1 \quad (k=1,2,3) \\ \sum_{t=n}^{n+2}\sum_{j=1}^{l_{k,1}} s_{k,i}\, x_{k,j}(t) \ge 1 \\ \beta_{k,j}(t) = \frac{\sum_{i=1}^{n_k}(s_{k,i}x_{k,j}(t))^2}{(\sum_{i=1}^{n_k} s_{k,i}\, x_{k,j}(t))^2} \ge 0.6 \end{cases} \tag{21}$$

Here, n represents the seasonal indicator for a year of planting. When n = 1, it is considered the first season; when n = 2, it is considered the second season.season at that time; it was considered as the second season at that time.

3.1.1 Construction of the unsold inventory price reduction model

The unsold goods markdown model has undergone a partial change in its objective function compared to the unsold goods waste model. The original function $-U(t)P(t)$ has been modified to $-0.5U(t)P(t)$,resulting in a new optimal crop planting plan model. The parameter $\lambda=0.5$ is substituted as part of the unsold goods markdown model.

3.2 Simulated Annealing Algorithm and Improvements

For solving the linear strategy objective function, due to the large amount of variable data and numerous constraints, the simulated annealing algorithm can be used to solve the data. However, the simulated annealing algorithm struggles to eliminate the influence of many small plots, so new optimization constraints need to be added for algorithm improvement:

Step 1. Finding the initial solution

Use the raw data from 2023 as the initial solution for function value processing.

Step 2. Randomly generate a new solution that meets the constraint conditions

Calculate the fitness of the new solution and use the Metropolis criterion to accept worse solutions with a certain probability, in order to prevent getting stuck in a local optimal solution for agricultural product planting, and to find the global optimal solution for agricultural product planting. The function is set as:

$$p = \alpha_L e^{-|W_{great} - W_L|} \quad (W_{great} \ge W_L) \tag{22}$$

$$p = 1 \quad (W_{great} < W_L). \tag{23}$$

Among them, p represents the probability of accepting W_{j+1}; L is the number of iterations; Wgreat is the optimal solution value; W_L is the value at the L iteration. α_L is the decay function at the L iteration.

In order to prevent the local optimal solution from being unable to reach the global optimal solution due to the rapid decline of the decay function, after reviewing the literature and discussing with the team, it is concluded that a decay coefficient of α=0.95 is optimal.

Step 3. Improved Segmentation Calculation Using Simulated Annealing Algorithm

Due to the long runtime of the simulated annealing algorithm in practical data processing, further model simplification is required. From the problem analysis after constructing the aforementioned annealing algorithm, it can be understood that the types of crops planted differ according to different land types and planting frequencies. These can be divided into four main categories: grain crops planted on flat dry land (A), terraced fields (B), and hilly land (C); vegetable crops (excluding Chinese cabbage, white radish, and red radish) planted in the first season of irrigated land (D), ordinary greenhouses (E), and smart greenhouses (F); Chinese cabbage, white radish, and red radish planted in the second season of irrigated land (D); edible fungi crops planted in the second season of ordinary greenhouses (E). Additionally, the unsold waste model must account for rice production crops in irrigated land (D) with one season per year. Thus, the unsold waste model and the unsold price reduction model for agricultural products are divided into four main categories and solved separately.

Step 4. Improvement of Adaptive Threshold in Simulated Annealing Algorithm

The use of the dispersion function cannot perfectly control the spatial dispersion of crop planting. There still exists a situation where crops are planted in dispersed areas of the same type of farmland. It is necessary to further reduce the impact of crop planting dispersion and introduce a dispersion function optimization to minimize the planting dispersion caused by the simulated annealing algorithm.

Determine the adaptive threshold for the simulated annealing algorithm:

$$O(L) = 0.1 - 0.05e^{-L}. \tag{24}$$

Among them, $O(L)$ is the judgment function after the L iteration. When the planting rate of a single crop in the planting area is less than $O(L)$, the crop is excluded, and the planting rate is reallocated to the remaining crops:

$$x_{k,i,j}(t) = \frac{x_{k,i,j}(t)}{\sum_j x_{k,i,j}(t)}. \tag{25}$$

The simulated annealing algorithm resolves the issue of scattered planting in crop cultivation, while also providing the optimal crop planting plan.

4 MODEL SOLUTION RESULTS

4.1 Unsolved Overstock Waste Model Solution Results

By applying the improved simulated annealing algorithm to the unsold waste model, the optimal solutions for the segmented types were obtained. The final iteration results are shown in Figure 1:

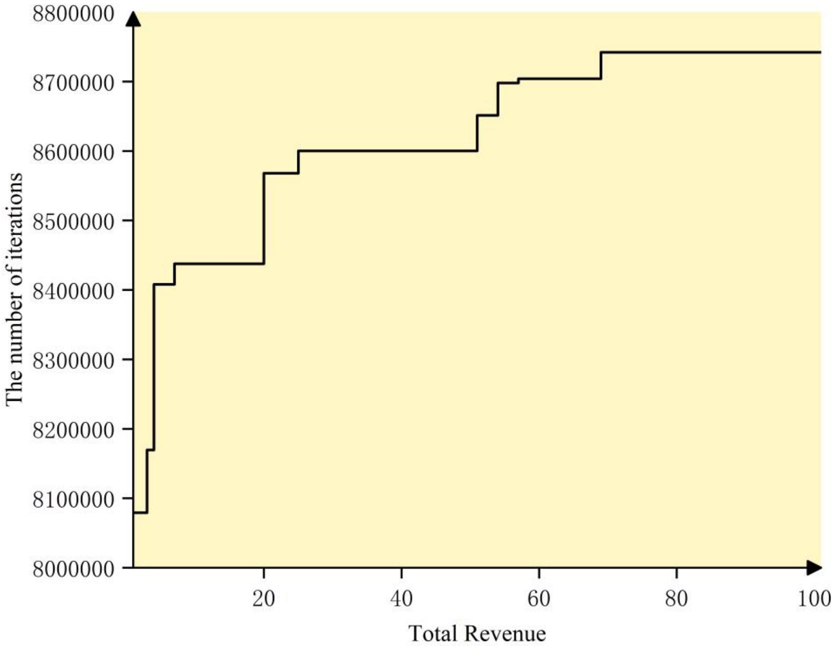


Figure 1 Simulated Annealing Algorithm for the Unsold Waste Model of Vegetable Crop Planting Iteration Revenue

The remaining iteration results; the final solution gives a total income of 26,473,326.43 yuan for the years 2024-2030.

4.2 Results of the Unsold Stock Price Reduction Model

By incorporating the improved simulated annealing algorithm into the unsold product markdown model, the optimal solution for each segmented type is obtained. The final iteration results are shown in Figure 2:

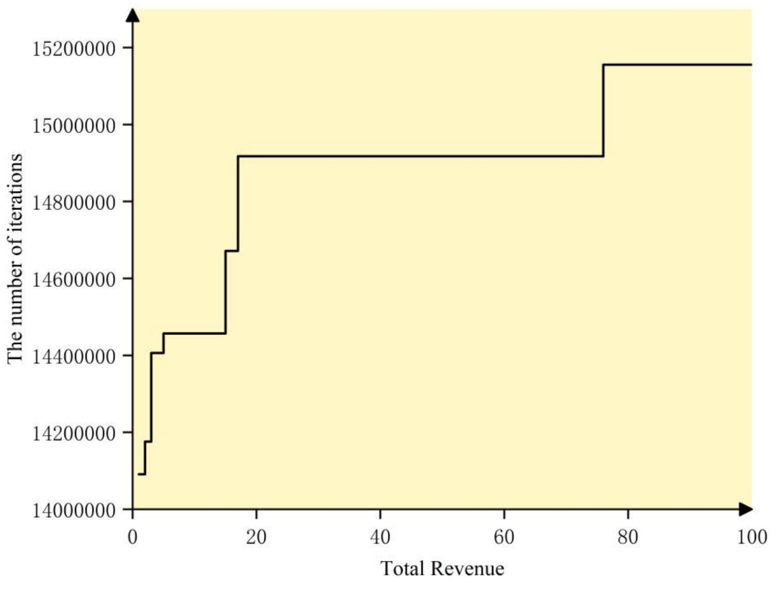


Figure 2 Simulated Annealing Algorithm for Slow-moving Price Reduction Model in Vegetable Crop Planting Iterative Profit

The remaining iterative results; the final solution yields a total income of 35,402,395.67 yuan for the years 2024-2030.

The final planting result is shown in Figure 3:

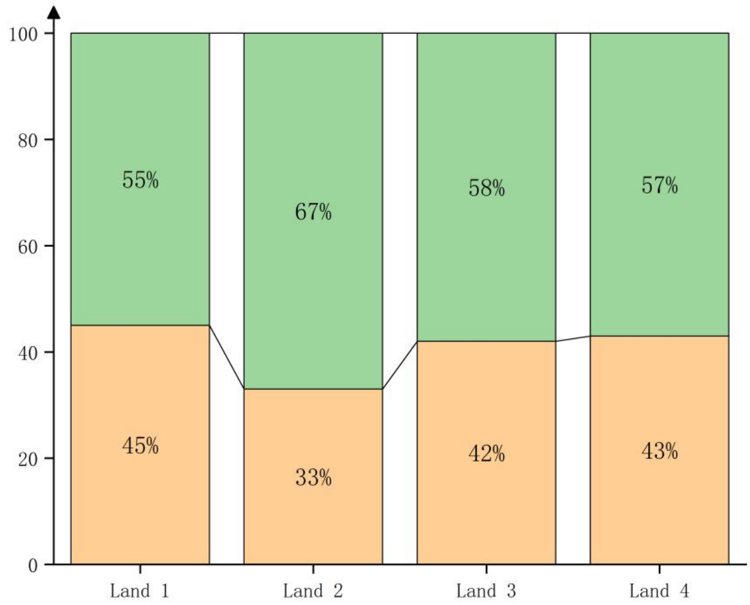


Figure 3 Comparison of Sales Volume between the Slow-Moving Discount Model and the Slow-Moving Waste Model

From Figure 3, it can be concluded that the unsold goods discount model enables the cultivation of more agricultural products to maximize income. In subsequent problem-solving, the unsold goods discount model will be used as the foundational model for improving crop cultivation strategies.

5 CONCLUSION

This article is based on the types of farmland for crop cultivation and the equations for crop yield and sales price indicators. It aims to construct an objective function for crop planting strategies and establishes a structural optimization model between multiple crops and multiple pieces of farmland. An improved simulated annealing algorithm with an adaptive threshold algorithm was designed to solve the model. The model was then applied to 1,201 acres of farmland, divided into 34 plots of varying sizes, for validation. The following conclusions were drawn:

a) The calculation results show that the unsold waste model indicates the actual agricultural output has already exceeded market demand. This result reflects that current agricultural production is sufficient to meet the general market demand. On the other hand, the unsold price reduction model proposes a solution that is more beneficial for agricultural profitability, increasing rural income, and promoting higher economic value for rural areas.

b) Traditional genetic algorithms, particle swarm algorithms, and simulated annealing algorithms have wide applications, but their actual computation time and accuracy are relatively low. By specifically adjusting and improving the simulated annealing algorithm, text integrate the complex constraint conditions with the process of generating new solutions in the algorithm, ultimately deriving the optimal planting strategy under the constraint conditions. This not only overcomes the issues of low accuracy, long convergence time, and difficulty in obtaining the optimal solution in the previous code, but also adds more practical value and significance to the proposed multi-crop planting plan.

c) The example calculation results indicate that under different types of unsold goods processing conditions, the improved simulated annealing algorithm can obtain the optimal solution for the overall model performance within the allowed error of the iteration, suggesting that the algorithm is widely applicable to crop planting strategy models.

Based on the budget provided by the above model, the following agricultural planting recommendations are given:

a) It is recommended to adopt a dynamic crop rotation mechanism, based on the planting rate index $X_k(t)$ generated by the simulated annealing algorithm, combined with the seasonal carrying capacity characteristics of the cultivated land type S_k. Priority should be given to deploying crops such as corn and wheat in flat dry lands (A) and terraced fields (B), utilizing their storage tolerance characteristics to balance the risk of unsold crops.

b) Implement the 'Rice + Fast-growing Vegetables' intercropping model in the irrigated area (D). In the first season, rice is planted to meet the basic food demand, and in the second season, a sales slump and price reduction model is used to guide the planting of cabbage-type crops with short growth cycles and high price elasticity, creating a differentiated planting sequence.

c) Establish an IoT-based agricultural product market forecasting system that updates the expected sales volume $D(t)$ and sales price $P_k(t)$ indicators in real time. Integrate sensor monitoring data with an improved simulated annealing algorithm to dynamically optimize the value range of the λ parameter ($0.5 \leq \lambda \leq 1$), and create a flexible mechanism for handling slow-moving goods.

d) Regarding the issue of scattered land plots for cultivation, it is recommended that local governments promote the construction of 'concentrated contiguous planting areas.' By facilitating land transfer, the concentration index $\beta_{k,j}(t)$ can be enhanced, ensuring that the dispersion of crops on similar types of arable land remains within the optimized range where $\beta_{k,j}(t) \geq 0.8$. This will help reduce the complexity of algorithmic solutions;

e) Strengthen training on algorithm applications for agricultural cooperatives, establish a digital twin model incorporating parameters from 34 plots, and regularly simulate planting strategy scenarios, with a particular focus on the dual-season planting constraints of k=6 types of cultivated land (smart greenhouses). This includes ensuring that $x_{6,i,j,1}(t) + x_{6,i,j,2}(t) \leq 1$ for continuous annual production. Subsequent research can delve deeper into the dynamic impact mechanisms of climate change factors on the yield indicator $H_k(t)$, as well as the corrective effects of different regional government subsidy policies on the cost indicator $\alpha_k(t)$.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Luis Eduardo Urbán Rivero, Jonás Velasco, Javier Ramírez Rodríguez. A Simple Greedy Heuristic for Site Specific Management ZoneProb-lem. Axioms, 2022, 11(7): 318.

[2] Araya A, Prasad P V V, Gowda P H, et al. Evaluating optimal irrigation strategies for maize in Western Kansas. Agricultural Water Management, 2021, 246: 106677.

[3] Chen Shang, He Liang, Cao Yinxuan, et al. Comparisons among four different upscaling strategies for cultivar genetic parameters in rainfed spring wheat phenology simulations with the DSSAT-CERES-Wheat model. Agricultural Water Management, 2021, 258: 107181.

[4] Zhang Z, Lin M, Han B, et al. Prediction of local scour depth around cylindrical piles: Using simulated annealing algorithm and ensemble learning. Ocean Engineering, 2025, 330, 121221.

[5] Zhang Q, Song L, Zeng Y, et al. Real-time power optimization strategy for fuel cell ships based on improved genetic simulated annealing algorithm. Electric Power Systems Research, 2025, 245, 111647.

[6] Kumar V, Gautam L, Dahiya R. Hybrid NSGA-III and simulated annealing approach for multi-objective time–cost -quality-sustainability optimization in wastewater treatment plant construction projects. Asian Journal of Civil Engineering, 2025, (prepublish): 1-16.

[7] Yingnian W, Hao W, Manhua L, et al. The adaptive two-stage ant colony Simulated Annealing Algorithm for solving the Traveling Salesman Problem. RAIRO - Operations Research, 2025, 59(2): 1199-1213

[8] Rabbani M, Ganjali A, Asl F H, et al. Using a hybrid genetic- simulated annealing algorithm for designing a recyclable waste collection system. OPSEARCH, 2024, (prepublish): 1-23.

[9] Wu H, Li Z, Deng X, et al. Enhancing agricultural sustainability: Optimizing crop planting structures and spatial layouts within the water-land-energy-economy-environment-food nexus. Geography and Sustainability, 2025, 6(3): 100258.

[10] Wang C, He X, Yan D, et al. Research on Optimal Crop Planting Strategy based on NSGA-II Algorithm. International Core Journal of Engineering, 2025, 11(4): 345-352.

[11] Bellangue D, Barney J, Flessner M, et al. Site Preparation and Planting Strategies to Improve Native Forb Establishment in Pasturelands. Agronomy, 2024, 14(11): 2676.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3092

SUSTAINABLE CROP PLANNING IN A MOUNTAINOUS VILLAGE USING LINEAR PROGRAMMING

Tong Chen
International Business College, South China Normal University, Guangzhou 510631, Guangdong, China.
Corresponding Email: 20223638032@m.scnu.edu.cn

Abstract: This study focuses on optimizing crop planning in a mountainous village in North China, where the cold climate permits only one growing season per year. The village's arable land is distributed across 34 plots of various types, including flat drylands, terraces, slopes, and irrigated fields. A linear programming model is developed to maximize planting revenue, considering crop rotation requirements, land suitability, planting costs, yields, and market prices. To simulate future conditions, crop prices in 2030 are modeled using normally distributed random numbers. The model sets two sales scenarios: (1) surplus production beyond expected sales cannot be sold, resulting in waste; and (2) surplus is sold at 50% of the 2023 market price. To address uncertainties in yield, cost, price, and demand, as well as potential planting risks, a particle swarm optimization algorithm is employed with random factors to simulate changing conditions. Results show that the village can achieve annual revenues between 8 to 9 million RMB under both deterministic and stochastic scenarios. The optimized strategies satisfy agronomic and environmental constraints and offer scientific support for sustainable agricultural development in resource-limited mountainous regions.
Keywords: Crop cultivation; Optimal planting strategy; Linear programming particle swarm algorithm; Multiple regression modeling

1 INTRODUCTION

With the in-depth promotion of the rural revitalization strategy, the sustainable development of the agricultural economy has gradually become a key issue in promoting the high-quality development of the social economy. How to realize the double enhancement of agricultural production efficiency and ecological efficiency by optimizing the crop planting structure under the conditions of limited land resources has become one of the core issues of modern agricultural development.

Du Neng first proposed the spatial pattern of crop planting structure in his agricultural location theory in 1826, and since then, scholars worldwide have explored its spatial distribution, migration and driving factors[1]. Taking the Greater Bay Area as the object, Taking the Greater Bay Area as the object, a study found that from 1990 to 2020, grain and oilseed planting declined, while vegetables and fruits increased, leading to a shift from a rice-based to a diversified vegetable-dominant planting pattern with evident spatial clustering[2]. Other scholars used GIS technology, time series trend and spatial agglomeration analysis methods to study crop structure evolution in Hunan province counties[3]. In terms of exploring the influencing factors, it has been pointed out that natural conditions and socio-economic factors jointly drive the change of cropping structure, and the influencing mechanism shows a complex development trend[4,5]. Some studies further showed that soil type, water resource status and input factors are the key constraints on cropping structure [6]. Other studies used remote sensing, fieldwork, and models to build planting databases and analyze structural evolution via indices and quantitative methods[7]. For optimization, studies generally start from the dimensions of ecological benefits, economic benefits and irrigation water consumption, and explore the sustainable models by adjusting crop sowing area[8]. Some built multi-objective models with agricultural yield, economic benefits, and water consumption as objectives, generating optimal solutions via improved weight combination methods and incorporating dialogue coordination mechanisms, thus offering a practical path to optimize planting structures in irrigation areas under water rights constraints [9].

However, the majority of existing studies focus on macro-scale regions such as provinces or counties, with limited attention to the optimization of crop planting structures at the village level under complex terrain and resource constraints. This study examines a mountainous village in North China, characterized by a cold climate, fragmented land, and a single annual growing season. A crop-plot matching optimization model based on linear programming is developed, incorporating crop rotation rules, land suitability, planting costs, yields, and market prices. To address planting risks and variability under uncertain conditions, a particle swarm optimization (PSO) algorithm is employed to solve the extended model and enhance the robustness of the planting strategy. And the aim of this research is to develop location-specific, profit-maximizing crop planning strategies that adhere to agronomic and ecological constraints, thus providing theoretical support and practical guidance for sustainable agricultural development in resource-limited mountainous regions.

2 MODELING AND SOLVING FOR OPTIMAL PLANTING SCHEMES

2.1 Data Preprocessing

Before modeling, the original planting data need to be preprocessed to ensure the accuracy of the analysis and the validity of the model inputs. This study utilizes 2023 data from a mountainous village in North China, focusing on plot information, crop adaptation and price range (Data source: https://www.mcm.edu.cn/). First of all, for the plot information and crop adaptation plot information, due to its relative simplicity and structure, the Excel tool is used to extract and organize, and the plot name, area and type are unified and summarized to ensure that each crop is accurately matched with its suitable plot. In order to more intuitively show the planting distribution and proportion of each type of 4 crop, corresponding visualization diagrams were drawn, as shown in figure 1 and figure 2, which assisted in assessing the adaptability of different plots to crops. Additionally, when dealing with the price range data of crops, a method of perturbation simulation based on the midpoint is adopted. Specifically, the midpoint value, which is the average of the maximum and minimum values, is used as the baseline. A small perturbation value that follows a normal distribution is then superimposed on this baseline to simulate the potential price fluctuations in 2023. The process is implemented in Python to ensure that the price data are both representative and reflect a certain degree of uncertainty, providing reasonable input parameters for the subsequent model.

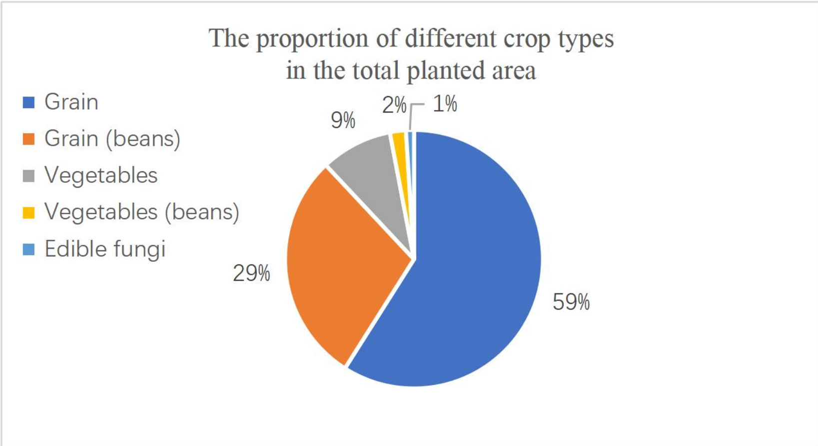


Figure 1 Proportion of Total Cultivated Area Represented by Different Crop Types

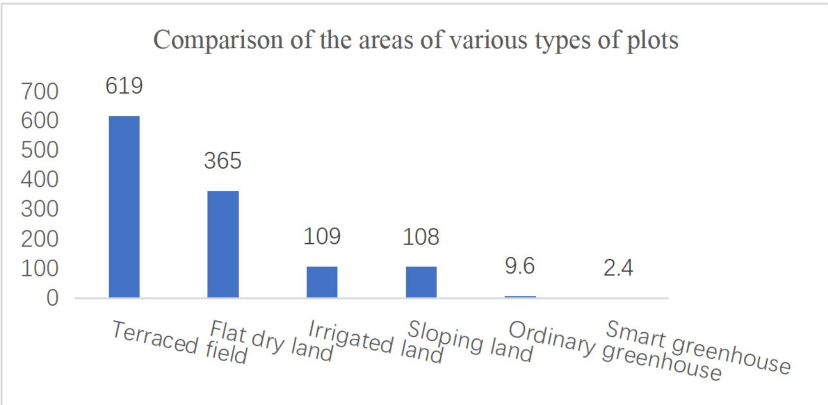


Figure 2 Ratio of Parcel Area by Type

2.2 Modeling and Solving Multi-Objective Linear Programming

2.2.1 Linear programming modeling

Linear programming is a typical mathematical optimization method for seeking optimal allocation strategies to maximize economic benefits under limited resources [10]. In crop planting decisions, linear programming can be used to determine the optimal acreage allocation scheme to maximize planting returns while satisfying various constraints.

For the crop planting problem, the model objective not only includes profit maximization, but also needs to comprehensively consider a variety of factors such as unit production, expected sales volume, planting costs, market prices and so on. In reality, when the total production of a crop exceeds its expected sales volume, there may be two ways to deal with it: one is to sell all of it, and the other is to sell the excess portion at a 50% discount. Therefore, two different yield objective functions are set in the model to reflect these two situations separately. At the same time, the model also needs to introduce planting type, plot size, seasonal arrangement, etc. as constraints, so as to construct a multi-objective linear planning model to enhance the scientific and adaptive planting decisions. The following are the multi-objective functions as well as the constraints:

$$\mathrm{MaxProfit1} = \mathrm{Max}\left(\sum_{j=2024}^{2030}\sum_{i}\sum_{k}\sum_{s}\left(\mathrm{Sales}(j,k,s) \times \mathrm{Price}(j,k) - \mathrm{Cost}(j,k) \times X(i,j,k,s)\right)\right) \quad (1)$$

$$\text{MaxProfit2} = \text{Max}(\sum_{j=2024}^{2030}\sum_{i}\sum_{k}\sum_{s}(\text{Sales}(j,k,s) \times \text{Price}(j,k) - \text{Cost}(j,k) \times X(i,j,k,s) + \text{Excess}(i,k,s) \times \text{Cost} \times 0.5) \tag{2}$$

$$\text{s.t.}\begin{cases} \sum_{k=1}^{41}\sum_{s=1}^{2} X(i,j,k,s) \le A_i \\ \sum_{k=1}^{41}\sum_{s=1}^{2} Z(i,j,k,s) \ge 1, \forall i,j,k,s(\mathit{legume\ crop}) \\ \sum_{j}^{j+2}\sum_{s=2}^{2} X(i,j,k,s) \ge 0.1 \times A_i \\ Z(i,j,k,s) + Z(i,j+1,k,s) \le 1 \\ X(i,j,k,s) \le A_i \times Z(i,j,k,s) \\ X(i,j,k,s) \le A_i \times \text{Suitable\_plot}_{i,k} \\ \sum_{s}^{2} X(i,j,k,s) \le A_i \times \text{Suitable\_season}_{i,k} \end{cases} \tag{3}$$

2.2.2 Solving linear programming models

Python is used to solve the above multi-objective linear programming model to obtain the optimal total return of the village from 2024 to 2030 under the two scenarios: (1) surplus production beyond expected sales cannot be sold and is treated as waste; (2) surplus production is sold at 50% of the 2023 market price. The simulation results for both scenarios are shown in figure 3. It presents the simulation results for both scenarios, showing that total revenue under Scenario 2 is generally higher than in Scenario 1, despite some fluctuations in both cases. The key difference lies in the handling of surplus production. In Scenario 1, unsold surplus leads to direct losses, while in Scenario 2, even selling at a discount allows partial revenue recovery. This highlights the importance of managing overproduction through strategies such as secondary markets or discount sales, which help stabilize income and enhance agricultural resilience.

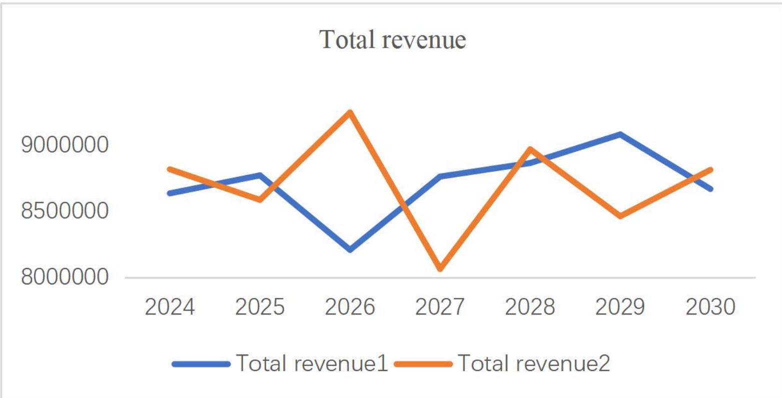


Figure 3 Total Returns from 2024 to 2030 for both Scenarios

On this basis, the optimal crop planting scheme in the corresponding time range is further obtained figure 4 and figure 5 show the distribution of planting area of different crops in each year under the two scenarios, respectively, reflecting the adjustment strategies of the model for crop types and plot configurations under different assumptions. In Scenario 1, where surplus production beyond expected sales cannot be sold and is therefore wasted, the model favors crops with stable yields and predictable market demand, such as wheat and mushrooms, to reduce financial risk. In contrast, Scenario 2 allows surplus to be sold at 50% of the 2023 market price, encouraging a more diversified planting strategy. The increase in vegetable and legume cultivation reflects an effort to extract value from potential overproduction and improve overall returns. These results demonstrate the model’s ability to adapt planting strategies to different economic environments.

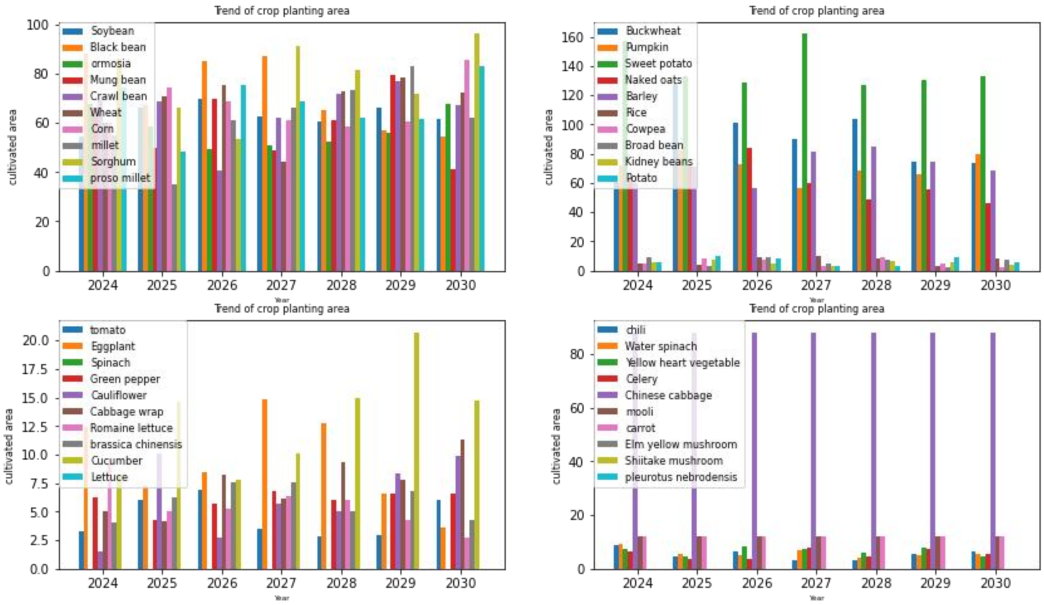


Figure 4 Scenario 1 Planting Distribution from 2024 to 2030

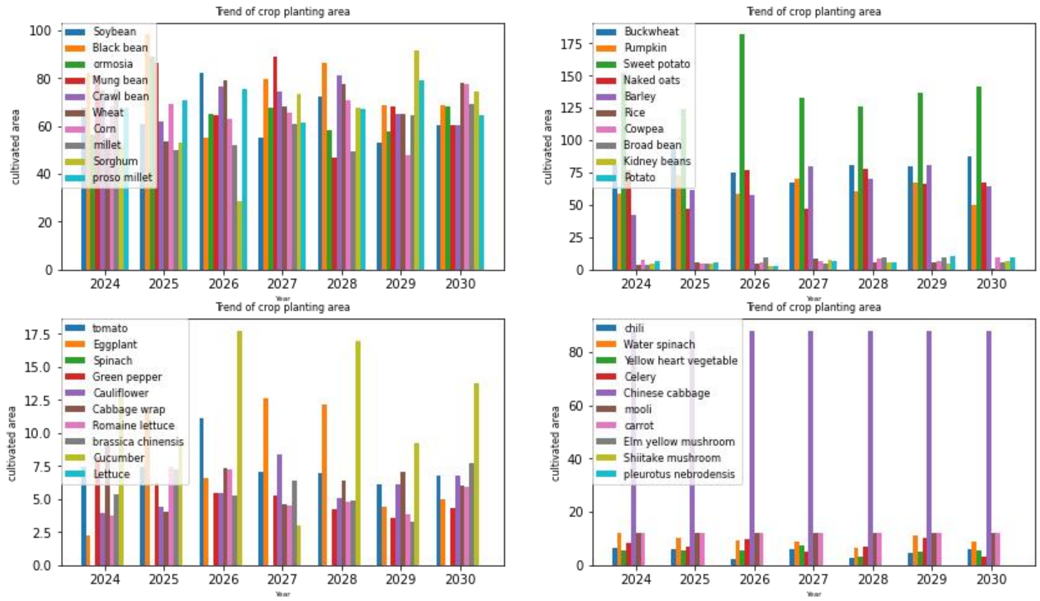


Figure 5 Scenario 2 Crop Planting Distribution from 2024 to 2030

3 MODELING AND SOLVING UNDER MULTIPLE RISK FACTORS

3.1 Optimization Modeling

While adhering to the restrictions of crop rotation and succession, the crop was modeled with different key factors using the adjust_parameters function in Python, taking into account the fluctuations in yield, cost of cultivation, sales price and expected sales volume over time, and the effects of uncertainty. The following model adjustments were made based on multi objective linear programming.

（1）Adjustments to sales volume:

For corn and wheat, sales are expected to grow at an average annual rate of 5 to 10 percent:

$$Sales_{j,k} = Sales_{0,k} \times (1 + r)^{t-2023}, r \in [0.05,0.1] \tag{4}$$

For other crops, sales are expected to be 5% up or down relative to 2023:

$$Sales_{j,k} = Sales_{0,k} \times (1 + r)^{t-2023}, r \in [-0.05,0.05] \tag{5}$$

（2）Adjustment of acreage: acreage fluctuates up and down by 10% per year due to climate and other factors:

$$Y_{j,k} = Y_{0,k} \times (1 + r)^{t-2023}, r \in [-0.1,0.1] \tag{6}$$

（3）Adjustments to planting costs: The average annual increase in planting costs is 5% due to market factors and other factors:

$$Costs_{j,k} = Costs_{0,k} \times (1 + 0.05)^{t-2023} \tag{7}$$

（4）Adjustments to the sales price:
For vegetable crops, sales prices have increased by an average of 5 to 10 percent annually:

$$Price_{j,k} = Price_{0,k} \times (1 + 0.05)^{t-2023} \tag{8}$$

For morel mushrooms, the sales price declined by an average of 5% per year:

$$Price_{j,k} = Price_{0,k} \times (1 - 0.05)^{t-2023} \tag{9}$$

（5）Determine the objective function:

$$\mathrm{MaxProfit} = \mathrm{Max}(\sum_{j=2024}^{2030} \sum_{i} \sum_{k} \sum_{s} (\mathrm{Price}_{j,k} \times \mathrm{Y}_{j,k} - \mathrm{Costs}_{j,k}) \times \mathrm{X}_{i,j,ks} \tag{10}$$

The constraints of this model are the same as those of the previous model.

3.2 Particle Swarm Algorithm for Solving Optimization Models

Particle Swarm Algorithm (PSO) is a population intelligent optimization algorithm that approximates the optimal solution in an iterative manner by simulating the search behavior of particles and continuously updating the individual extremes and global extremes[11].The following are the solution steps:

（1）Initialization:
Parameters such as particle swarm size, maximum iteration number, and learning factor are set. Each particle represents a possible planting strategy, its position represents the planting area of each crop on different plots, and its speed represents the adjustment amplitude of the planting strategy. Randomly generate the position and speed of the initial particles, calculate the fitness value of each particle, and record its individual optimal position, and determine the global optimal position in the current population(gbest).

（2）Calculation and updating of adaptation values
In each iteration, the fitness values of all particles are recalculated. If the current fitness of a particle is better than its historical individual optimum, its individual extreme value (Pbest) is updated; if the fitness of a particle is better than the current global optimum, the global extreme value and the corresponding position (gbest) are updated synchronously.

（3）Particle update
In each round of iteration, the position and velocity of the particles are updated with the following formula:

$$\mathrm{V}_{i,j}(t+1) = \mathrm{V}_{i,j}j(t) + s_1 r_{1,j}(\mathrm{P}_{dbj} - \mathrm{P}_{ij}(t)) + s_2 r_{2j}(\mathrm{P}_{dbj} - \mathrm{P}_{ij}(t)) \tag{11}$$

$$\mathrm{P}_{i,j}(t+1) = \mathrm{P}_{i,j}(t) + \mathrm{V}_{i,j}(t+1) \tag{12}$$

Where $i = 1, \ldots, N$, $j = 1, \ldots, D$, t is the number of iterations, S_1 and S_2 are non-negative learning factor constants, and r_{1j} and r_{2j} are independent random numbers uniformly distributed over the interval [0,1].

（4）Examination of termination conditions
The particle swarm algorithm checks after each round of iterations whether the termination conditions are satisfied, including reaching the maximum number of iterations (set to 100) or the particle swarm converges, i.e., the value of the global optimum no longer changes significantly. If the conditions are satisfied, the iteration is stopped and the current global optimal solution is output; otherwise, the execution continues to the next round.

In order to improve the convergence speed of the particle swarm algorithm, this paper introduces a priori knowledge, sets 35% of the initial particles as human-defined parameters, and the remaining 65% of the particles are randomly generated. In addition, in order to enhance the realism of the model, a constraint function containing sales volume, mu yield, planting cost and sales price is constructed. Finally, the optimization results with corresponding total returns and crop planting distributions were obtained through 9 rounds of 100 iterations each, as shown figure 6 and figure 7.

As shown in figure 6, the total revenue shows fluctuations during the period from 2024 to 2030. A peak occurs in 2026, which may be attributed to favorable market conditions, optimized planting strategies, and other positive influencing factors. Conversely, the significant decline in 2027 might be due to adverse external factors, such as natural disasters, market price drops, or increased planting costs. And figure 7 depicts the optimal planting distribution of different crops over the years. For example, staple crops like wheat and rice maintain relatively stable planting areas, highlighting their crucial significance in ensuring food security. Meanwhile, the planting areas of cash crops such as tomatoes and eggplants exhibit certain fluctuations, which are related to market demand changes, price fluctuations, and the impact of the optimization algorithm on maximizing returns. By comprehensively analyzing these distribution changes, the planting structure can be adjusted according to market demands and resource endowments, thereby improving the overall economic efficiency of crop planting and optimizing resource utilization.

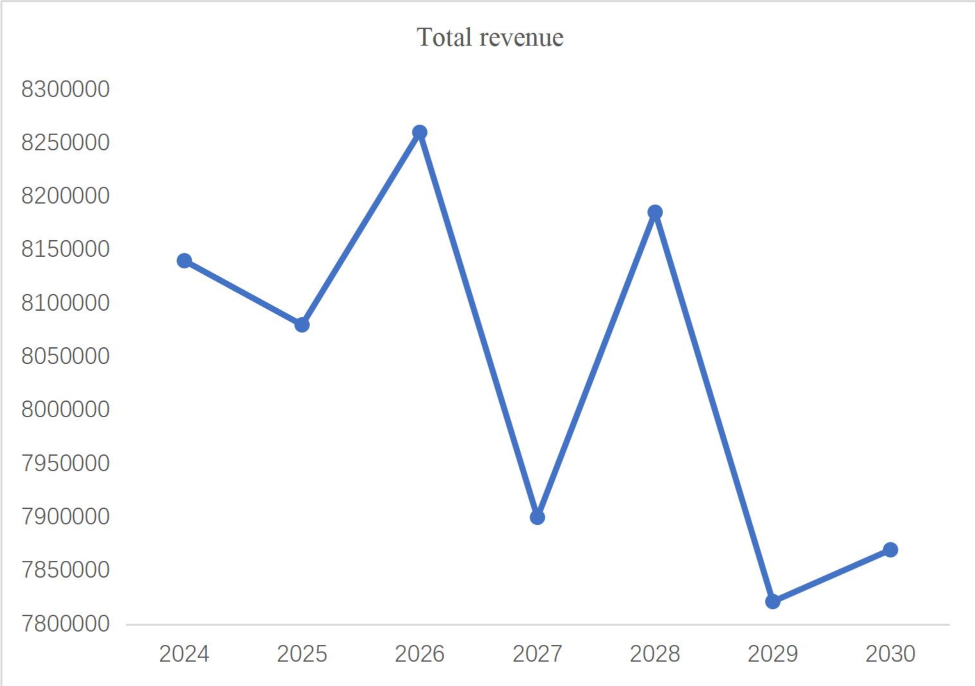


Figure 6 Total Revenue from 2024 to 2030

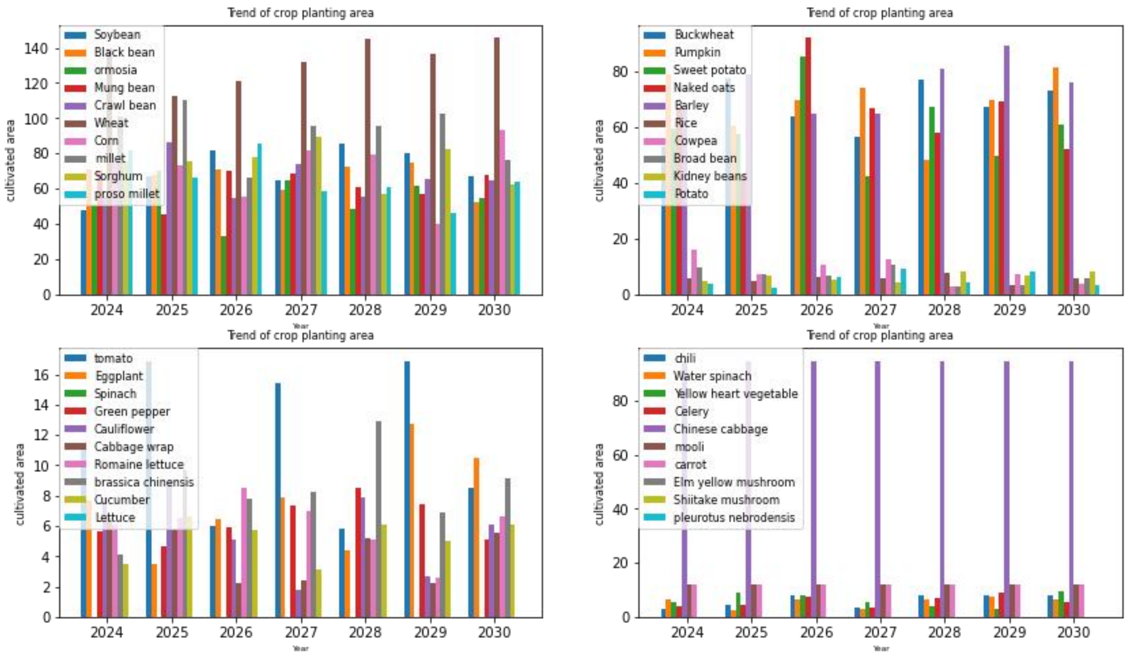


Figure 7 Optimal Planting Distribution Map of Crops

4 CONCLUSIONS

This paper focuses on the crop planting optimization problem, establishes a mathematical model based on linear programming, takes the maximization of revenue as the goal, combines the planting area, the market demand, the cost and revenue and other realistic factors, constructs a single-objective linear programming and multi-objective optimization model, and solves the optimal planting strategy under different conditions using Python.

The research results show that the linear programming model has good applicability and practical value in agricultural production decision-making, which can not only effectively improve the land use efficiency, but also provide a scientific basis for agricultural planting structure adjustment and resource allocation. However, this paper still has certain limitations. On the one hand, although a variety of influencing factors are considered in the modeling process, the actual planting environment is more complex, and there may be other factors that have not been taken into account to affect the planting efficiency. On the other hand, the particle swarm algorithm may not have optimal parameter settings in the solving process, which affects the accuracy of the final results. Future research can further expand the influencing factors of the model, optimize the parameters of the algorithm, and verify it with more actual data, so as to provide more accurate and effective decision support for the sustainable development of rural planting.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] LIN Dayan, ZHU Jing. Reasons for changes in the planting structure of major crops in China from the perspective of supply elasticity. Agricultural Technology and Economy, 2015(01): 33-41.
[2] HU Yunfei, LIU Xu, LIANG Junfen, et al. Spatial and temporal evolution of crop cropping structure in the Guangdong-Hong Kong-Macao Greater Bay Area from 1990 to 2020. Guangdong Agricultural Science, 2023, 50(01): 13-27.
[3] AN Yue, TAN Xuelan, TAN Jieyang, et al. Evolution of crop planting structure and influencing factors in Hunan Province. Economic Geography, 2021, 41(02): 156-166.
[4] Zhou XG. Analysis of the impact of agricultural planting factors on agricultural planting structure. Rural Practical Technology, 2023(10): 99-100.
[5] SONG Yu, LIU Weiping. Influencing factors and multivariate paths of the evolution of cultivated land planting structure - A qualitative comparative analysis of fuzzy sets based on case zones in 30 provinces. Geographic Research and Development, 2024, 43(04): 133-138.
[6] Bao Shuzong. Influence of agricultural planting factors on planting structure and optimization strategy. Agriculture and Technology, 2020, 40(13): 90-91.
[7] LUO Qiancheng,WANG Yuexiang,CHI Wenfeng,et al. Factors affecting the evolution of crop cropping structure in the black soil area of Inner Mongolia. Resource Science, 2023, 45(05): 994-1005.
[8] WU Menghan,WANG Yi. Optimized adjustment of multi-objective planting of crops in Shache Irrigation District of Xinjiang. People's Yellow River, 2024, 46(01): 120 125+131.
[9] ZHANG Hao Yang,CHEN Xing. Optimal planting structure preference method for irrigation areas under water right constraints. Modern Agricultural Science and Technology, 2024(24): 161-164.
[10] An Jinping. Application of linear programming in economic analysis. Journal of Central University of Finance and Economics, 2005(01): 44-47.
[11] Deng Chan,Li Chun,Li Mengqi, et al. Progress of particle swarm algorithm in agricultural hydrology. Anhui Agricultural Science,2021, 49(08): 16-20+29.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3093

THE LOCATION OF COLD CHAIN LOGISTICS FOR AGRICULTURAL PRODUCTS FOR THE"FIRST KILOMETRE"

YiTao Yin[1*], YiFei Liao[2], LeTong Ji[3], ZiFan Wang[2]
[1]*School of mathematics and statistics, Nanjing University of Information Science and Technology, Nanjing 210000, Jiangsu, China.*
[2]*School of Business, Nanjing University of Information Science and Technology, Nanjing 210000, Jiangsu, China.*
[3]*School of Computer Science, School of Cyber Security, Nanjing University of Information Science and Technology, Nanjing 210000, Jiangsu, China.*
Corresponding Author: YiTao Yin, Email: allenyin2004@163.com

Abstract: With the accelerated development of China's agricultural modernisation, the importance of cold chain logistics for agricultural products has become increasingly prominent. However, problems such as insufficient cold chain coverage and lack of pre-cooling links have led to high spoilage rates and high logistics costs for fresh agricultural products. Taking Jiangsu Province as an example, this study constructs a two-stage cold chain logistics and distribution centre site selection model that takes economic, environmental and social factors into account. In the first stage, this study selects 12 indicators in four dimensions, namely, economic development, traffic congestion, total logistics demand, and green and low-carbon level, and uses the entropy weight method to quantitatively assess the logistics level of 13 cities in Jiangsu Province. By calculating the logistics level score of each city, the spatial economic distance between cities is defined, and the K-means clustering algorithm is combined to divide Jiangsu Province into regions. The clustering results show that the clustering method based on spatial economic distance can more scientifically reflect the economic connection and logistics demand among cities, which provides an important basis for the subsequent site selection decision. In the second stage, the study uses the centre of gravity method to select the optimal distribution centre location within each clustering region. The centre of gravity method determines the optimal location of the distribution centre by calculating the weighted average location of each demand point within the region with the objective of minimising transport costs. Ultimately, by comparing the two methods based on spatial economic distance clustering and traditional geographic distance clustering, the results show that the clustering method based on spatial economic distance significantly reduces the transport cost. The total transport cost is reduced from $16,356.33 to $14,156.00, which is about 13.45%. This result not only verifies the validity and practicability of the constructed model, but also provides a scientific basis for the location of cold chain logistics distribution centre in Jiangsu Province, which helps to improve the operational efficiency and economic benefits of cold chain logistics.
Keywords: Cold chain logistics; Site selection model; Entropy weight method; K-means clustering; Centre of gravity method

1 INTRODUCTION

With the acceleration of China's agricultural modernisation process, the production, circulation and consumption patterns of agricultural products have undergone significant changes. As a key link in guaranteeing the quality of agricultural products, reducing losses and enhancing market competitiveness, cold chain logistics for agricultural products has become an indispensable part of the modern agricultural industrial system. In recent years, the state attaches great importance to the development of cold chain logistics as an important hand in promoting rural revitalisation and high-quality development of agriculture. 2020, the Ministry of Agriculture and Rural Development issued the "Implementation Opinions on Accelerating the Construction of Agricultural Products Warehousing and Freshness Preservation Cold Chain Facilities", which explicitly puts forward that we should focus on the main production areas of fresh and live agricultural products and poverty-stricken regions, accelerate the construction of warehousing and freshness preservation cold chain facilities, and solve the "first kilometre" problem of agricultural products going out of villages and entering cities. "The No.1 Document of the Central Government in 2024 further emphasised the need to optimise the construction of cold chain logistics system for agricultural products, accelerate the layout and construction of backbone cold chain logistics bases and public cold chain logistics facilities in county areas, and promote the deep integration of cold chain logistics with modern agriculture and food processing.
Cold chain logistics of agricultural products refers to the maintenance of a suitable low-temperature environment at all times during the whole process of production, processing, transport, storage and marketing, so as to guarantee the quality and safety of agricultural products. At present, some progress has been made in the infrastructure construction of cold chain logistics, and the pre-cooling and cold storage functions of cold storage and the use of refrigerated trucks in the transport process can effectively prolong the freshness period of agricultural products and reduce losses [1,2]. However, the degree of standardisation and marketisation of cold chain logistics still needs to be improved. Relevant studies have pointed out that the current cold chain logistics exists problems such as high logistics costs and serious losses, and put forward countermeasures such as modernisation of hardware and facilities, standardisation of management, and resource intensification. With the continuous development of the Internet of Things, blockchain and

other new technologies, scholars have begun to explore their application in cold chain logistics, optimize the layout of the cold chain network through the model, and provide a scientific basis for the construction of infrastructure [3]. Aiming at the current situation and problems of the cold chain logistics of special agricultural products, some studies have also proposed optimisation strategies based on IoT technology, including cost management, supervision and control, and the cultivation of third-party enterprises, in order to improve the quality and efficiency of services [4]. In addition, there are also studies that analyse the development opportunities and challenges of cold chain logistics from the perspective of rural revitalization strategy, and put forward development paths such as strengthening top-level design, optimizing layout, and innovation drive [5].

The site selection of cold chain logistics distribution centre is the core link of optimizing the network layout, which is crucial for reducing logistics costs, improving distribution efficiency and guaranteeing the quality of fresh produce products. Early researches mainly focused on the basic siting model, with the goal of minimising transport costs, using traditional methods such as the centre of gravity method and linear programming, which are applicable to the siting of a single distribution centre. With the development of the logistics industry, research has gradually shifted to multi-objective optimisation, taking into account multiple objectives such as cost, service quality and distribution efficiency [6]. In recent years, the research pays more attention to the comprehensive consideration of economic, environmental and social factors [7], and from the perspective of green logistics, the improved algorithm solves the siting problem considering the time window, carbon emission and cargo damage cost [8], and constructs the siting model with integrated carbon emission and freshness, which is solved by using multiple optimisation algorithms [9-11].

In summary, the cold chain logistics of agricultural products in China is facing challenges such as high loss, high cost, insufficient infrastructure and lagging standardisation. Existing research focuses on infrastructure construction, distribution centre location and technology application, and is shifting from single-objective optimization to multi-objective optimization, with the introduction of algorithms and models to enhance decision-making science. New technologies such as the Internet of Things (IoT) provide new ideas for industry development. Future research needs to comprehensively consider economic, environmental and social factors to promote the sustainable development of cold chain logistics. This study focuses on systematically analysing the key factors of agricultural cold chain logistics site selection, constructing a scientific site selection model and verifying its application. It provides a theoretical basis for the governmental departments to formulate the layout planning of cold chain logistics facilities, and provides cold chain logistics enterprises with a scientific decision-making method for site selection, so as to promote the efficient development of cold chain logistics of agricultural products.

2 A TWO-STAGE COLD CHAIN LOGISTICS AND DISTRIBUTION CENTRE SITE SELECTION MODEL BASED ON K-MEANS CLUSTERING AND CENTRE OF GRAVITY APPROACH

Based on the existing research, this study takes 13 cities in Jiangsu Province as examples, selects 12 indicators from four dimensions of economic development status, traffic congestion degree, total logistics demand and green low-carbon level, uses entropy weighting method to carry out the calculation of the weight of each indicator, constructs the spatial economic distance and combines with the K-means clustering algorithm and centre of gravity method to construct the two-phase siting model, which takes into account the reasonableness of the region delineation and the site selection scientific, providing a new solution for the cold chain logistics distribution centre site selection problem [12].

2.1 Calculation of Spatial Economic Distance

2.1.1 Selection of logistics level score indicators

The logistics level score was calculated based on 12 indicators for each city and used to redefine the distance between the 13 cities. The logistics level score we used mainly considered four dimensions: economic development, traffic congestion, total logistics demand, and green and low-carbon level. Specifically, the green and low-carbon level was measured by total energy consumption, CO2emissions, and carbon intensity [13]; and the economic development was measured by the per capita disposable income of the urban resident population, total investment in fixed assets, and growth rate of the tertiary industry; Vehicle density (motor vehicle ownership divided by urban construction area), population density (permanent urban population divided by urban construction area), and road mileage are used to measure the level of traffic congestion; permanent urban population, freight volume, and area of land used for logistics warehousing are used to measure the demand for logistics, and the relevant data are obtained from the China Statistical Yearbook.

2.1.2 Entropy weighting method to determine indicator weights

This paper adopts the entropy weight method to quantitatively measure the logistics level score indicators, and the calculation steps of the entropy weight method are mainly as follows:

(1) Data normalisation. The indicators have different units of measurement, so they need to be normalised, and the specific normalisation methods are as follows. Taking into account the need for logarithmic operations on the data in the subsequent calculations, a non-negative translation is carried out in advance, with a translation of 0.001 units:

Positive indicators:

$$X'_{\theta\,ij} = \frac{x_{\theta\,ij} - x_{min}}{x_{max} - x_{min}} + 0.001 \tag{1}$$

Negative indicators:

$$X'_{\theta ij} = \frac{x_{max} - x_{\theta ij}}{x_{max} - x_{min}} + 0.001 \tag{2}$$

where $x_{\theta ij}$ is the value of the original indicator j for the i evaluation object in the year θ ,x_{max} ,x_{min} are the maximum and minimum values of the indicator, i =1,2,..., n , j =1,2,..., n , θ =1,2,..., n . $X'_{\theta ij}$ are the new values after normalisation and non-negative leveling. In this paper, three indicators, namely total energy consumption, CO2emission and carbon intensity, are negatively normalised, while the rest are positively normalised.

(2) Calculate the weight of the i evaluator under the j indicator in the year θ for that indicator:

$$Y_{\theta ij} = \frac{X'_{\theta ij}}{\sum_{\theta=1}^{r} \sum_{i=1}^{n} X'_{\theta ij}} \tag{3}$$

(3) Calculate the information entropy for the j th indicator:

$$S_j = -k \sum_{\theta=1}^{r} \sum_{i=1}^{n} (Y_{\theta ij} \ln \quad (Y_{\theta ij})) \tag{4}$$

included among these $k = \frac{1}{\ln \quad rn}$

(4) Calculate the coefficient of variation:

$$E_j = 1 - S_j \tag{5}$$

(5) Determine the weights, which reflect the importance of the indicator, and evaluate the importance of the indicator in direct proportion to the value of the weights. Calculate the weight of the indicator j :

$$W_j = \frac{E_j}{\sum_{j=1}^{m} E_j} \tag{6}$$

The results of calculating the weight coefficients of the 12 indicators using the entropy weighting method are shown in Table 1:

Table 1 Entropy Weighting Method of Logistics Level Score Indicators

dimension	norm	weights
Economic development	Per capita disposable income of the urban resident population	0.0855
	Total investment in fixed assets	0.1015
	Tertiary growth rate	0.0724
Level of traffic congestion	Vehicle density	0.1107
	population density	0.1132
	Number of road miles	0.0632
Total logistics demand	Municipal resident population	0.0538
	volume of freight	0.0356
	Land area for logistics and warehousing	0.0626
Green and low-carbon levels	Total energy consumption	0.1254
	Carbon dioxide emissions	0.0852
	carbon intensity	0.0909

2.1.3 Calculation of spatial economic distance

Based on the normalised data of the 12 indicators of the 13 cities and the weighting coefficients of each indicator, the logistics level score of each city is calculated Z_i with the following formula:

$$Z_i = \sum_{j=1}^{n} Y_{ij} W_j \tag{7}$$

After deriving the logistics level score, the improved distance formula is as follows:

$$D(X_i, X_j) = \frac{D_{ij}^2}{(Zi\ Zj)^u} \tag{8}$$

$D(X_i, X_j)$ measures the spatial economic distance between two cities, influenced by two types of distance factors: spatial distance factor and economic distance factor. D_{ij}^2 indicates the spatial distance factor between two cities. Z_iZ_j denotes the logistics economic distance factor between two cities. Where Z_i ,Z_j represent the logistics level scores of city i and city j respectively. u represents the degree of influence of the logistics economic distance factor in the measurement of distance between two cities.The larger the value of u ,the greater the impact of the logistics economic distance factor.When u is 0,the distance formula is equivalent to the Euclidean distance.In this paper,we assume that the logistics economic distance factor $u = 0.1$,which best scientifically balances the influence of economic gravity and distance[14].

2.2 Regional Division of Jiangsu Province based on K-means Clustering

Based on the calculated spatial economic distance between the 13 cities, clustering was performed using the K-means algorithm. The Euclidean distance is used as a measure of the similarity between the data objects, and the similarity is inversely proportional to the distance between the data objects, the greater the similarity, the smaller the distance. The algorithm needs to pre-specify the number of initial clusters k and k initial clustering centres, according to the similarity between the data objects and the clustering centres, the position of the clustering centres is continuously updated, and the sum of squares of errors (SSE) of the class clusters is continuously reduced, when the SSE is no longer changing or the objective function is converging, the clustering is finished, and the final result is obtained.
The Euclidean distance between a data object and a clustering centre in space is calculated as.

$$d(x, C_i) = \sqrt{\sum_{j=1}^{m} \left(x_j - C_{ij}\right)^2} \tag{9}$$

where, x is the data object,C_i is the i th clustering centre,m is the dimension of the data object, ,$x_j C_{ij}$ is the j th attribute value of x and C_i .
The sum of squared errors SSE for the entire dataset is calculated as.

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} |d(x, C_i)|^2 \tag{10}$$

Before performing K-means clustering, based on the relationship between the number of clusters k and the error squared and SSE, the elbow method was used to select the k value corresponding to the elbow point as the optimal number of clusters based on the trend of SSE at different k values [15].
In order to make a comparison, clustering was performed based on the spatial economic distance and the actual geographic distance calculated from the latitude and longitude coordinates of the 13 cities, respectively, and the optimal number of clusters based on the spatial economic distance was determined to be 3 based on the elbow method first, while the optimal number of clusters based on the actual geographic distance for clustering was 2. The results of the clustering are shown in Table 2 and Table 3.

Table 2 Spatial Economic Distance Clustering

Region	Cities covered
Region I	Nanjing, Suzhou, Huai'an, Yancheng
Region II	Wuxi City, Xuzhou City, Nantong City, Yangzhou City, Zhenjiang City, Taizhou City
Region III	Lianyungang City, Suqian City, Changzhou City

Table 3 Geographic Clustering

Region	Cities covered
Region I	Nanjing, Suzhou City, Wuxi City, Changzhou City, Nantong City, Yancheng City, Yangzhou City, Zhenjiang City, Taizhou City
Region II	Xuzhou, Lianyungang, Huai'an, Suqian

2.3 Selection of Optimal Distribution Centre Locations within each Clustered Area Based on the Centre of Gravity Method

Since the clustering algorithm part of the integrated consideration of both spatial and non-spatial factors for the division of the region and site selection, so this study to minimize the transportation cost as the sole basis for decision-making, first of all, according to the actual situation to determine the demand point of the coordinates of$j(x_i, y_i)(i, j = 1,2\ldots, n)$, set the coordinates of the logistics centre P for (x_0, y_0) , which can be obtained as follows.

$$F = \sum_{j=1}^{n} c_j \tag{11}$$

$$c_j = f_j w_j d_j \tag{12}$$

$$d_j = \sqrt{(x_0 - x_j)^2 + (y_0 - y_j)^2} \tag{13}$$

where F is the total transport and distribution cost from centre P to each demand point:c_j is the cost from P to a single pointj :f_j is the unit transport and distribution cost from P to pointj ;w_j is the distribution volume from P to pointj ;d_j is the straight line distance between logistics centre P and pointj .
Bringing equation (12) into equation (11) yields.

$$F = \sum_{j=1}^{n} f_j w_j d_j \tag{14}$$

virtuous$\frac{\partial F}{\partial x_0} = \frac{\sum_{j=1}^{n} f_j w_j (x_0 - x_j)}{d_j} = 0,\ \frac{\partial F}{\partial y_0} = \frac{\sum_{j=1}^{n} f_j w_j (y_0 - y_j)}{d_j} = 0$
Equations (15) and (16).

$$x_0 = \frac{\sum_{j=1}^{n} f_j w_j \frac{x_j}{d_j}}{\sum_{j=1}^{n} f_j \frac{w_j}{d_j}} \tag{15}$$

$$y_0 = \frac{\sum_{j=1}^{n} f_j w_j \frac{y_j}{d_j}}{\sum_{j=1}^{n} f_j \frac{w_j}{d_j}} \tag{16}$$

At this point, the obtained $P(x_0, y_0)$ is the extreme point of the total distribution costF expression: if it is not the optimal solution, then continue to iterate, the iteration formula is as follows.

$$x^{(q+1)} = \frac{\dfrac{\sum_{i\in I} f_i w_i x_i}{\left[\left(x^{(q+1)} - x_i\right)^2 + \left(y^{(q)} - y_i\right)^2\right]^{1/2}}}{\dfrac{\sum_{i\in I} f_i w_i}{\left[\left(x^{(q+1)} - x_i\right)^2 + \left(y^{(q)} - y_i\right)^2\right]^{1/2}}} \tag{17}$$

$$y^{(q+1)} = \frac{\dfrac{\sum_{i\in I} f_i w_i y_i}{\left[\left(x^{(q+1)} - x_i\right)^2 + \left(y^{(q)} - y_i\right)^2\right]^{1/2}}}{\dfrac{\sum_{i\in I} f_i w_i}{\left[\left(x^{(q+1)} - x_i\right)^2 + \left(y^{(q)} - y_i\right)^2\right]^{1/2}}} \tag{18}$$

In the above two equations,q is the number of iterations,i is the demand point number, andI is the cluster class where the demand point number is located. When the iteration ofF no longer changes, the iteration stops. At this time,F obtains the minimum value, and the corresponding solution is the optimal solution, i.e., the centre of gravity of logistics centre P site selection and construction.

2.4 Analysis of Model Results

On the basis of using the clustering algorithm to determine each distribution area, each city in the divided area is taken as a logistics demand point, and the specific situation is shown in Figure 1. The per capita disposable income of each point is taken as the demand for logistics resources, and the centre of gravity method is used to make corrections to the cold chain logistics centre.

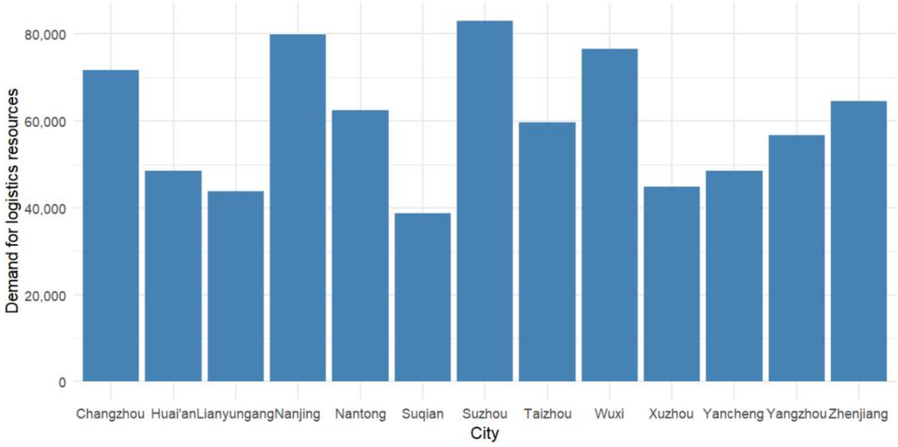


Figure 1 Demand for Logistics Resources by City

Not taking into account the changes in freight rates in each region of Jiangsu Province, assuming that the unit freight rate is RMB 1/kg*km, and taking into account that the establishment of a cold chain logistics centre requires supporting infrastructure, and it cannot be arbitrarily established in a location outside of the city, the logistics centre will be set up in the city closest to the centre of gravity.

Calculate the transport costs and site selection cities in Jiangsu Province under the 2 modelling methods, and the results are shown in Table 4 and Table 5.

Table 4 Results of Spatial Economic Distance Clustering Centre of Gravity Method

Region	City of centre of gravity	Transport costs
Region I	Taizhou	8829.00
Region II	Yangzhou	4330.67
Region III	Suqian	996.33

Table 5 Results of Geographic Clustering Centre of Gravity Method

Region	City of centre of gravity	Transport costs
Region I	Changzhou	14779.33
Region II	Suqian	1577.00

According to the data in the table, the model with clustering based on spatial economic distance and combined with the centre of gravity method for site selection has a total transport cost of $14,156.00, compared to the model with clustering based on geographic distances calculated directly from the latitude and longitude coordinates of the 13 cities, which has a total transport cost of $16,356.33. This indicates a significant reduction in the total transport cost from $16,356.33 to $14,156.00, which is about 13.45 per cent, after taking into account the economics. This result fully proves the effective improvement of the cold chain logistics site selection model in this study in the three aspects of indicator selection, the improvement of distance formula driven by the economicity factor, and the application of the centre of gravity method based on the clustering results. In addition, by comparing the two clustering methods, the clustering method based on spatial economic distance is more reasonable in terms of regional division, and can more accurately reflect the economic linkages and logistics demand among cities.

In summary, the two-stage model proposed in this study is not only innovative in theory, but also shows high practical value in practical application. By comprehensively considering economic, environmental and social factors, the model is able to more scientifically determine the location of cold chain logistics and distribution centres, effectively reduce the transportation cost and improve the logistics efficiency, thus providing a strong support for the sustainable development of cold chain logistics of agricultural products.

3 CONCLUSION

This study focuses on the siting of cold chain logistics and distribution centres for agricultural products in Jiangsu Province, aiming to construct a two-stage siting model that integrates economic, environmental and social factors in order to optimise the layout of distribution centres and improve the efficiency and sustainability of cold chain logistics. In the first stage of the study, this study uses the entropy weight method to quantitatively assess the logistics level of each city in Jiangsu Province. By calculating the logistics level scores of each city, this study provides important data support for the subsequent regional division and distribution centre location. Subsequently, the study further combines the K-means clustering algorithm to divide Jiangsu Province into regions based on spatial economic distance.In the second stage, this study uses the centre of gravity method to select the optimal distribution centre location within each region, and by calculating the weighted average location of each demand point within the region, the optimal location of the distribution centre can be effectively determined so as to minimize the transportation cost.

The research results show that the clustering method based on spatial economic distance can divide the area more scientifically and significantly reduce the transport cost. By optimizing the site layout of the distribution centre, the total transport cost is reduced from 16356.33 yuan to 14156.00 yuan, which is about 13.45% lower. This result not only verifies the validity and practicability of the constructed model, but also provides a scientific basis for the location of cold chain logistics distribution centres in Jiangsu Province, which helps to improve the operational efficiency and economic benefits of cold chain logistics.

In the future, intelligent optimisation algorithms such as machine learning can be further introduced to construct a dynamic siting model to adapt to the dynamic changes in demand and improve the optimisation accuracy. In addition, the scope of the study can be expanded from Jiangsu Province to other regions or more application scenarios to explore the applicability and effectiveness of the model under different geographic, economic and policy conditions, so as to provide a broader and more practical scientific basis for the location decision-making in the cold chain logistics industry.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] WANG Ailing, CHEN Ruyi, CHEN Guoqing. Research on the development problems and countermeasures of cold chain logistics of agricultural products in the context of rural revitalisation. New Agriculture, 2022(18): 90-91.

[2] Cheng Juan. Analysis of the Role of Rural E-commerce Logistics Service Quality on Consumers' Purchase Intention--Taking Fresh Agricultural Products as an Example. Research on Business Economy, 2022(13): 85-88.

[3] ZHU Kun-Ping, JIANG Lin-Lin, WANG He-Nan. Analysis and countermeasures of cold chain logistics market of agricultural products in Hebei Province. Price Monthly, 2016(12): 64-68.

[4] Hao Hongqiang. Research on the construction of cold chain logistics of speciality agricultural products in Henan Province under the vision of Internet of Things. Food Research and Development, 2021, 42(23): 233-234.

[5] Liu Ying, Zhang Yuan. Research on the development path of cold chain logistics of agricultural products under the strategy of rural revitalisation. Modernisation of Shopping Malls, 2025(10): 57-59.

[6] WANG Yong, HUANG Siqi, LIU Yong, et al. Optimisation of logistics multi-distribution centre site selection based on K-means clustering method. Highway Traffic Science and Technology, 2020, 37(01): 141-148 .

[7] QIN Binbin,LI Shizhen,CAO Yu. Research on site selection of cold chain logistics centre for fresh agricultural products combined with economic development factors//Chinese Society of Logistics,China Federation of Logistics and Purchasing.Proceedings of the 2024 (23rd) China Logistics Academic Annual Conference. College of Economics and Management, Changjiang University, 2024: 526-540.

[8] Gao Xuejie, Zhang Shoujing, Liu Yueqiang. Research on low-carbon cold chain logistics path optimisation with multi-objective synergy. Manufacturing Automation,2025, 47(04): 127-135.

[9] Su Jianing. Research on site selection and path optimisation of cold chain logistics and distribution centre considering carbon emission and freshness. East China Jiaotong University,2023.

[10] Wang Dan. Research on site selection of cold chain logistics and distribution centre under carbon emission constraint. Anhui University of Technology, 2024.

[11] LIU Yi, HU Enthalpy, WU Jiang. Research on comprehensive optimisation of cold chain logistics network under carbon emission reduction policy. Journal of Management Engineering, 2025: 1-14.

[12] Baile. Research on emergency distribution, coordination mechanism and recovery strategy of fresh produce cold chain logistics under uncertainty environment. Beijing University of Chemical Technology, 2022.

[13] Guo Pei, Liang Dong. Whether low-carbon pilot policies improve urban carbon emission efficiency - A quasi-natural experimental study based on low-carbon pilot cities. Journal of Natural Resources, 2022, 37(07): 1876-1892.

[14] YANG Junbang, ZHAO Chao. A review of research on K-Means clustering algorithm. Computer Engineering and Applications, 2019, 55(23): 7-14+63.

[15] YAO Suo, WU Xiurong, LI Hao, et al. Research on site selection of logistics distribution centre based on improved K-means algorithm. Logistics Science and Technology, 2024, 47(05): 10-13+19.

Eurasia Journal of Science and Technology
ISSN: 2663-1016
DOI: https://doi.org/10.61784/ejst3091

EFFECT OF THE LIGHT QUANTUM EFFICIENCY ON THE SYNCHRONOUS PHOTO CATALYTIC ABATEMENT OF SO_2–NO_X BINARY GAS

Kuo Yao[1#], JingLi Gao[1,2 #], Yasser M. A. Mohamed[3*], YingHui Han[1*]
[1]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China.
[2]School of Mathematics & Physics, North China Electric Power University, Baoding 071003, China.
[3]Photochemistry Department, National Research Centre, Dokki, Giza, P. O. 12622, Egypt.
[#]Kuo Yao JingLi Gao are both the first authors.
[*]Yasser M. A. Mohamed and YingHui Han are both the corresponding authors.
Corresponding Author: YingHui Han, Email: hanyinghui@ucas.ac.cn; Yasser M. A. Mohamed, Email: y.m.a.mohamed@outlook.com

Abstract: Photocatalytic decontamination has emerged as a highly promising technology for air pollutant control and energy conversion. In this work, a semiconductor photocatalyst was synthesized via a hydrothermal method, featuring nano-sized TiO_2 doped with nickel (Ni), zirconium (Zr), and nonmetallic nitrogen (N). The kinetic process of synchronous abatement of the SO_2-NO_x binary gas system was systematically investigated in a self-designed photocatalytic reactor. From the perspective of quantum optics, a light quantum efficiency model was established to elaborate the photocatalytic degradation mechanism, accompanied by a detailed analysis of the synergistic effects of multiple influencing factors on the interaction between incident light and pollutants. Experimental results revealed that flue gas concentrations exert significant impacts on contaminant removal efficiencies. Notably, the experimental data showed excellent consistency with the calculated results across a wide range of operational conditions.
Keywords: Photocatalysis; Simultaneous desulfurization and denitrification; Quantum efficiency; Reaction mechanism

1 INTRODUCTION

Air pollutants, including particulate matter, sulfur dioxide (SO_2), nitrogen oxides (NO_x), and mercury (Hg)—predominantly emitted from coal-fired flue gases—have triggered escalating environmental challenges, most notably the intensification of haze episodes in recent years[1–4]. This has galvanized global efforts toward air pollution mitigation. While conventional treatment technologies—such as physical adsorption, chemical oxidation, microbial processes, and high-temperature calcination—have contributed to environmental purification, they are hindered by inherent limitations: low removal efficiency, potential for secondary pollution, and restricted applicability across diverse contaminant profiles[4,5].

Against this technological backdrop, photocatalytic technology has emerged as a promising solution. Since Fujishima's seminal 1972 publication on water electrolysis for hydrogen production over TiO_2 single crystals, photocatalytic reactions have garnered interdisciplinary attention[6]. Titanium dioxide (TiO_2) stands out as the most widely studied photocatalyst due to its non-toxicity, chemical stability, strong surface hydrophilicity, recyclability, and unique ability to completely mineralize organic pollutants into harmless end products (H_2O and inorganic ions) without secondary contamination[6,7].

Current research on TiO_2-based systems in environmental remediation—particularly for flue gas treatment—focuses on leveraging its intrinsic properties and enhancing performance through structural/chemical modifications. Notable investigations include: Yuan et al.[8] prepared nanometer TiO_2 composite material using the aluminum silicate fiber as carrier by sol-gel method, under UV irradiation(wavelength was 253.7nm), examining the efficiency of desulfurization and denitrification and mercury removal of the material, the results show that denitrification efficiency can reach 40%, and found that a low concentration of SO_2 in the flue gas could play a role in promoting the removal of NO, and the temperature has an inhibition effect on the process of photocatalytic oxidation. Meng et al.[9] reported 37.4% NO and 25.84% NO_2 removal over activated carbon-TiO_2 composites, highlighting light-intensity-dependent enhancements and temperature-related deactivation mechanisms.

Despite these advancements, practical implementation of TiO_2 photocatalysis remains constrained by fundamental challenges. As a wide-bandgap semiconductor (3.2 eV, $\lambda \approx 387$ nm), TiO_2 responds only to ultraviolet (UV) light, which constitutes merely 3–4% of solar irradiance. The underutilization of the visible spectrum (45% of solar energy) presents a critical bottleneck, necessitating strategies to extend its optical response to longer wavelengths[10]. Additionally, mechanistic understanding of photocatalytic reactions—including charge carrier dynamics and surface redox processes—remains incomplete, urging deeper exploration of reaction pathways[11].

To address these knowledge gaps, this study establishes a rational photocatalytic reaction model. By synthesizing N, Ni, and Zr-codoped TiO_2 photocatalysts, we investigate the quantum efficiency of photocatalytic processes from an optical physics perspective. Experimental validation and dynamic kinetic simulations are employed to corroborate the model's accuracy, providing a foundational framework for advancing TiO_2-based photocatalysis toward real-world applications.

2 EXPERIMENTAL

2.1 Materials

All reagents were of analytical grade and used without further purification. TiO_2 and $Zr(NO_3)_4$ were purchased from BASF Chemical Co., Ltd., Tianjin, China. $Ni(NO_3)_2$ was supplied by Founder Reagent Factory, Beichen, Tianjin, China. All other reagents were purchased from Huaxin Reagents Co., Baoding, China. Samples were prepared via a hydrothermal reaction, with the nonmetallic element N introduced for photocatalyst modification.

2.2 Experimental Apparatus and Methods

The primary objective in designing a photocatalytic reactor is to provide a stable reaction space and environment, thereby enhancing reaction efficiency. Two critical theoretical bases underpin reactor design: light transport dynamics and catalytic reaction efficiency. The transmission and distribution of light within the reactor directly influence the interaction between the catalyst and incident photons. Substantially uniform light exposure minimizes optical loss during transmission, maximizing the catalyst's light utilization efficiency. Conversely, catalyst regions with insufficient or negligible light exposure cannot undergo adequate photocatalytic reactions. The self-designed experimental setup for NO and SO_2 oxidation is depicted in Figure 1.

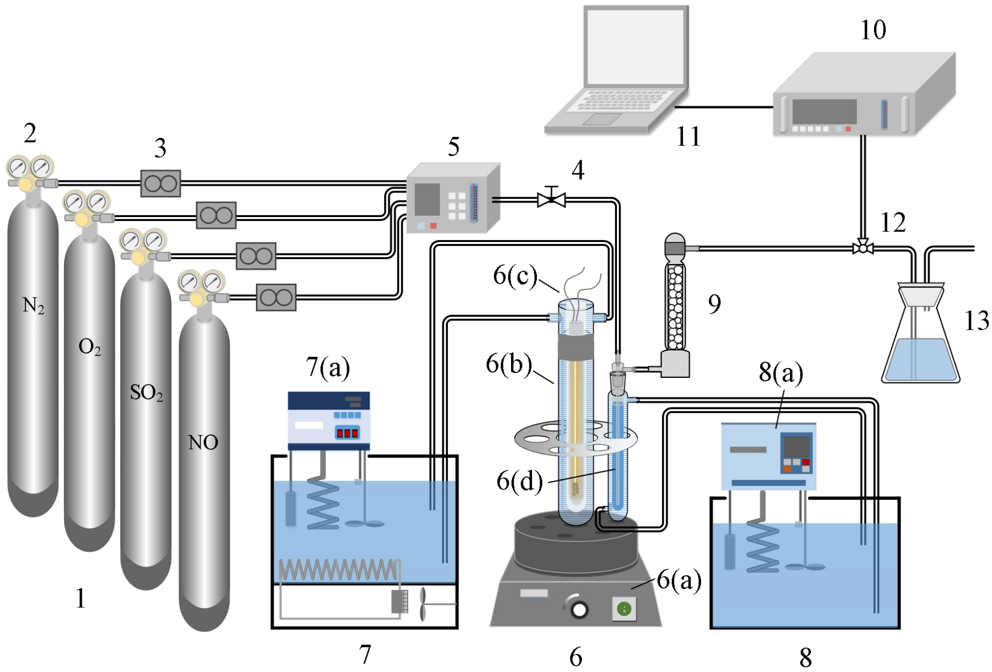


Figure 1 Schematic Diagram of the Polluting Gases Removal Apparatus

1:Cylinders of N_2、SO_2、NO and O_2;2: Regulator; 3: Gas rotameter;4: Valve;5: Gas mixer;6: Photochemical reaction apparatus (6(a): Magnetic stirrer;6(b): Cold trap;6(c): UV lamp;6(d): Reaction tube self-made);7: Cryostat device(7(a) Controller);8: High temperature thermostat(8(a) Controller);9: Drying tower;10: Flue gas analyzer;11: Computer;12: Three-way valve;13: Waste gas scrubber.

The experimental apparatus (Figure 1) consists of three integrated units: inlet flue gas simulation (Components 1–5), removal reactor (6–9), and outlet gas analysis (10–13). N_2 from cylinder (1) flows at 1 L/min through a regulator (2) and rotameter (3, LZB-30, Tianjin) into gas mixer (5), where SO_2, NO, or O_2 from separate cylinders are introduced via identical flow control systems with balance N_2 adjusted by a secondary flow meter, and the mixed gas passes through valve (4) into the reactor (6). The reactor features a UV-A light source (6(c), 800 W, λ=365 nm) with cooling jacket (6(b)), catalyst suspension in reaction tube (6(d)), a magnetic stirrer (6(a)) for uniform irradiance, cryostat (7) to maintain lamp temperature, and high-temperature thermostat (8) for stable reaction conditions, with gas velocity kept at 4.5–6 cm/s to minimize back-mixing and solution pH monitored by a waterproof meter (PHS-3G, Shanghai LEICI). Post-reaction gases are dried and analyzed via flue gas analyzer (10, Beijing Yunenghuitong) interfaced with computer (11), while residual gases pass through three-way valve (12) to scrubber (13) for complete removal before venting. All experiments were replicated thrice with results averaged for statistical rigor.

The output response is the removal efficiency of SO2 and NO (%), which were calculated as below:

$$\mathrm{Re}_{(NO/SO_2)} = (1-\frac{C_{out}}{C_{in}})\times 100\% \tag{1}$$

where C_{in} and C_{out} stand for the inlet and outlet concentrations of NO or SO2, respectively.

In industrial applications, numerous factors influence flue gas removal efficiency. This study investigates the effects of NO, SO_2, and O_2 concentrations, with initial NO levels ranging from 200 to 800 mg/m³ and O_2 from 3% to 21%. To explore reaction mechanisms, a quantum efficiency model was developed from an optical perspective using a photocatalytic reactor. Experimental data validation confirmed the model's accuracy and reliability.

3 3 RESULTS AND DISCUSSION

3.1 Analysis on the Photocatalytic Reaction Mechanism

Broadband semiconductors serve as photocatalysts primarily due to their photoelectric properties. Take TiO_2, an n-type oxide semiconductor, as an example: the bandgap energy (E_g) of its anatase phase is 3.2 eV. A photocatalyst functions only when the energy of incident photons is not less than its E_g , i.e.:

$$E = \hbar c / \lambda > E_g \tag{2}$$

Where E stand for the intensity of the incident light, $\hbar$ is Planck's constant, c is the speed of light, λ is the wavelength of light.

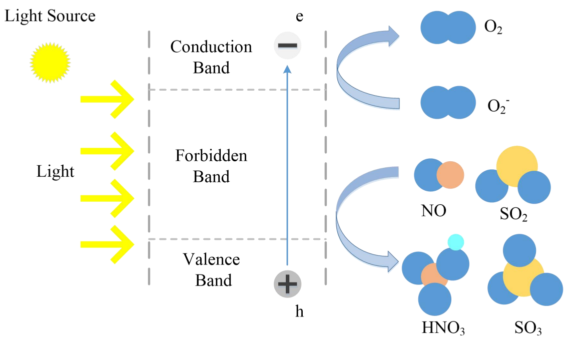


Figure 2 Schematic Diagram of the Mechanism of Photocatalytic

Figure 2 illustrates the photocatalytic electron-hole transport dynamics in TiO_2: upon irradiation with light whose photon energy (E) exceeds the anatase phase bandgap (Eg = 3.2 eV), valence band electrons are excited to the conduction band, generating photo-induced electron-hole pairs (e^- and h^+). Two dominant reaction pathways emerge: first, photoexcited electrons (e^-) reduce adsorbed oxygen to form superoxide anions (O_2^-), while photogenerated holes (h^+) react with surface-adsorbed water or OH^- to produce highly oxidative hydroxyl radicals (•OH), enabling NO decomposition via non-selective oxidation; on the other hand, after specific adsorption of pollutant molecules onto the TiO_2 surface, photo-induced carriers migrate to the interface and react directly with adsorbed species, driven by the redox potential difference between the carriers and contaminants.

3.2 The Quantum Efficiency Model of Photocatalytic Synchronous Abatement of SO_2–NO_x Binary Gas

3.2.1 Process of the photocatalytic reaction

At a UV irradiation, whose wavelength is less than 387nm, the catalytic process can be expressed as follows:

$$TiO_2 + hv(\mathrm{E > Eg}) \rightarrow e^- + h^+$$

$$O_2 + e^- \rightarrow O_2^-$$

$$H_2O + h^+ \rightarrow OH + H^+$$

$$O_2^- + H^+ \rightarrow HO_2$$

For NO, there are:

$$NO + HO_2 \rightarrow HNO_3$$

There may also be another reaction pathway:

$$O_2(\mathrm{ads}) + 2\,\mathrm{e}^- \rightarrow O_2^{2-}(\mathrm{ads}) \rightarrow 2\,\mathrm{O}^-(\mathrm{ads})$$

$$O_2(\mathrm{ads}) + \mathrm{h}^+ \rightarrow O^*$$

$$NO + O^* \rightarrow NO_2$$

$$NO_2 + OH \rightarrow HNO_3$$

SO_2 is oxidized into SO_3 by living radical:

$$SO_2 + 2OH \rightarrow SO_3 + H_2O$$

$$SO_2 + O_2^- + H_2O \rightarrow SO_3 + 2OH^-$$

3.2.2 The quantum efficiency of photocatalytic reaction process

Quantum efficiency (QY) quantifies the ratio of photo-generated carriers participating in catalysis to the number of absorbed photons. In photocatalytic kinetics, photoexcited carriers (e^-/h^+) undergo two competitive processes:recombinationandtrapping. Carriers are only effective when trapped and reacted with electron donors / acceptors.

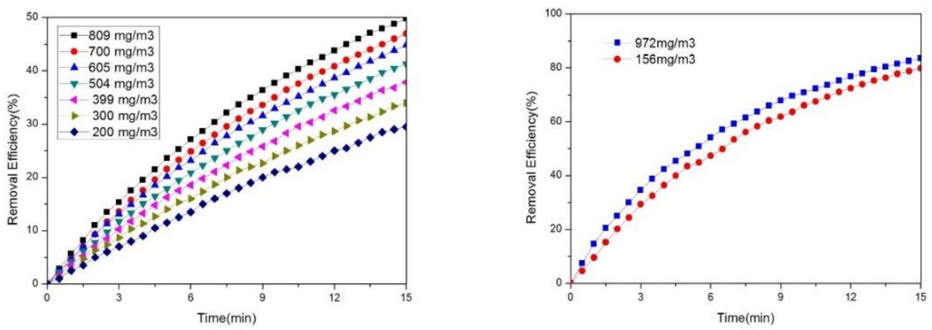
Formally: QY = Photo-generated carriers participating in catalysis / Absorbed photons. When photon flux is constant, enhancing photocatalytic efficiency fundamentally requires improvingcarrier trapping efficiency(e.g., reducing recombination), which is governed by material electronic structure, carrier migration dynamics, and carrier . For NO degradation, integrating QY with photocatalytic reaction kinetics yields:

$$QY = [(\frac{k_1[O_2][\mathrm{NO}]}{4k_2\phi})^2 + \frac{k_1[\mathrm{O}_2][\mathrm{NO}]}{2k_2\phi}]^{\frac{1}{2}} - \frac{k_1[\mathrm{O}_2][\mathrm{NO}]}{4k_2\phi} \tag{3}$$

Where, k is a constant, ϕ is the amount of photons. As equation shows, QY is inversely proportional to the amount of photons. When the light is intensity, the equation can be simplified to:

$$QY = [\frac{k_1[\mathrm{NO}]}{2k_2}]^{1/2}\phi^{-1/2} \tag{4}$$

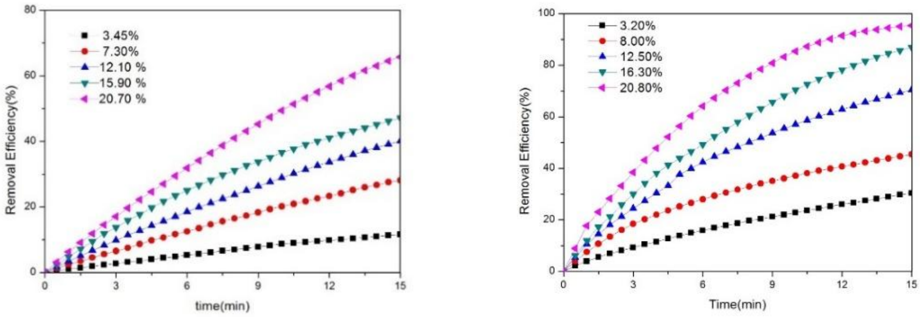
Evidently, quantum efficiency (QY) exhibits a linear relationship with the reciprocal of the square root of incident photon flux under certain kinetic regimes (e.g., diffusion-limited surface reactions). By correlating QY with the photon intensity profiles in designed experiments, a quantitative relationship between QY and pollutant concentration is established. Initially, QY increases with NO concentration due to enhanced surface adsorption of contaminants, which facilitates more efficient photon utilization by promoting carrier trapping at the catalyst-pollutant interface. As NO concentration rises, QY approaches unity when adsorption sites are fully occupied; however, excessive concentrations induce surface adsorption saturation, beyond which QY plateaus or declines due to reduced active site availability. Practically, optimizing NO concentration within the sub-saturation range can enhance QY by maximizing carrier-pollutant interaction efficiency.



(a) (b)

Figure 3 Relationship between Gas Concentration and Removal Efficiency (a)NO Removal Efficiency Changes with the Initial Concentration;(b) SO_2 Removal Efficiency Changes with the Initial Concentration

At 60°C and 20.6% O_2, NO and SO_2 concentrations were varied individually, with gas concentrations monitored over 15 minutes. Figure 3(a) and 3(b) depict the relationship between initial gas concentration and removal efficiency. For NO, removal efficiency increases with concentration up to a threshold, whereas SO_2 exhibits a weaker concentration dependence. To enhance clarity, SO_2 data are presented for two initial concentrations with a large disparity (e.g., low vs. high). This divergence may arise from distinct reaction pathways: NO preferentially undergoes (specific mechanism, e.g., radical-driven oxidation), while SO_2 adsorption/oxidation is limited by (e.g., surface site saturation or slower electron transfer kinetics).



(a) (b)

Figure 4 Relationship between Removal Efficiency and O_2 Concentration. (a) NO Removal Efficiency Changes with the Concentration of O_2;(b) SO_2 Removal Efficiency Changes with the Concentration of O_2

At 60°C, with initial NO and SO_2 concentrations of 700 mg/m³ and 500 mg/m³, respectively, the effect of O_2 concentration on pollutant removal efficiency was investigated. As shown in Figure 4(a) and (b), removal efficiency increased with rising O_2 concentration, particularly for SO_2: at 20.80% O_2 (v/v), SO_2 removal efficiency reached 95.41%. This behavior aligns with O_2's role as an electron acceptor in photocatalysis, facilitating superoxide (O_2^-) and hydroxyl radical (•OH) generation for enhanced oxidation.

4 KINETICS OF CATALYTIC REACTION

4.1 Effect of the Concentration of Pollutant on the Photocatalytic Reaction Kinetics

The kinetic study of desulfurization and denitrification is shown in Figure 5. When the content of oxygen was 20.6% and the temperature was 60°C, we using Langmuir-Hinshelwood (L-H) kinetic model to describe the process of photocatalytic degradation of NO and SO_2, which is:

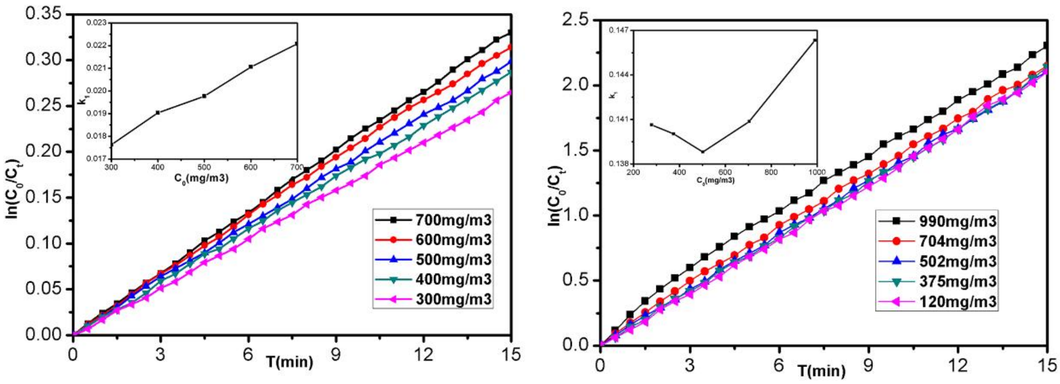
$$-\frac{dC}{dt} = v = \frac{kK[C]}{1+K[C]} \tag{5}$$

Where K is the adsorption constant, which is equal to the ratio of adsorption rate constant and desorption rate constant, and C is the concentration of the pollutant, v is the reaction rate, k is Langmuir rate constant. By integrating, the formula is rewritten as:

$$t = \frac{1}{kK}\ln(\frac{C_0}{C_t}) + \frac{1}{k}(C_0 - C_t) \tag{6}$$

Where C_0 and C_t are represent the initial concentration and instantaneous concentration respectively. Using k_1 was the product of K and k .

By calculating and fitting, we can get a good linear relationship between $\ln(C_0/C_t)$ and t , the degradation process is in line with the first-order kinetics model. At the same time, the k_1 is increasing with the initial concentration of the pollute, which further demonstrates the correctness of the optical quantum model.



(a) (b)

Figure 5 The Effect of the Concentration of Pollute on the Photocatalytic Degradation Rate Constant

The inset figure is the relationship between k_1 and the concentration of pollutants. Where (a) is used to show the relationship between the concentration of NO and the parameters of kinetic, while (b) is the relationship between the concentration of SO_2 and the parameters of kinetic.

4.2 Effect of the concentration of O_2 on the photocatalytic reaction kinetics

In order to further demonstrate the correctness of the optical quantum model, we study the effect of the concentration of O_2 on the photocatalytic reaction kinetics, which are shown in Figure 6 below.

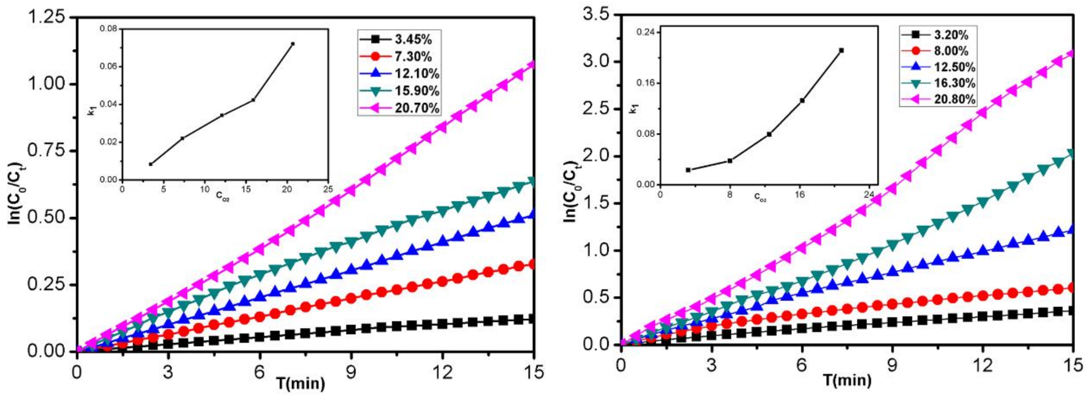


Figure 6 The Effect of the Concentration of O_2 on the Photocatalytic Degradation Rate Constant

The inset figure is the relationship between k_1 and the concentration of O_2. Where (a) is used to show the degradation process of NO, while (b) is the degradation process of SO_2.

A good linear relationship could be got between $\ln(\frac{C_0}{C_t})$ and t by calculating and fitting, so the degradation process of difference concentration of O_2 follows the first-order kinetics model. From the change rules of k_1, as shown in Figure 6, we can conclusion that the catalytic efficiency is proportional to the oxygen content, which is consistent with the above optical quantum model.

5 CONCLUSION

Ti-based composite photocatalysts were developed for efficient simultaneous removal of SO_2 and NO. Through systematic investigations on the impacts of pollutant concentrations and oxygen content on removal efficiency, a quantum efficiency model was established. This model reveals a strong correlation among photon utilization, carrier dynamics, and catalytic performance. Notably, experimental data exhibit good consistency with the model predictions, providing in-depth mechanistic insights into photon-driven deep oxidation for simultaneous gas purification.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

The study was supported by CAS-ANSO Co-funding Research Project (No. CAS-ANSO-CFRP-2024-04), National Natural Science Foundation of China (No. 52320105003), and the Fundamental Research Funds for the Central Universities (Project No.: E3ET1803).

REFERENCES

[1] WU Y, ZOU L, ZHU H. Improved combustion of NH_3/C_2H_4 with Ni modified Fe-based catalyst. Chemical Engineering Journal, 2023, 472: 145187.

[2] YANG L, WANG S, CHEN C. Monitoring and Leak Diagnostics of Sulfur Hexafluoride and Decomposition Gases from Power Equipment for the Reliability and Safety of Power Grid Operation. Applied Sciences, 2024, 14(9): 3844.

[3] RAJAGOPALAN S, BROOK R D, SALERNO P R V O. Air pollution exposure and cardiometabolic risk. The Lancet Diabetes & Endocrinology, 2024, 12(3): 196-208.

[4] BOYJOO Y, SUN H, LIU J. A review on photocatalysis for air treatment: From catalyst development to reactor design. Chemical Engineering Journal, 2017, 310: 537-559.
[5] WANG X, DUAN R, LI Z. The Critical Role of Oxygen Vacancies in N_2O Decomposition Over Cobalt-Doped CeO2 Catalysts. Environmental Science & Technology, 2025, 59(11): 5839-5847.
[6] MARTIN S T, LEE A T, HOFFMANN M R. Chemical mechanism of inorganic oxidants in the TiO2/UV process: increased rates of degradation of chlorinated hydrocarbons. Environmental Science & Technology, 1995, 29(10): 2567-2573.
[7] ALAHIANE S, SENNAOUI A, SAKR F. Synchronous role of coupled adsorption-photocatalytic degradation of Direct Red 80 with nanocrystalline TiO_2-coated non-woven fibres materials in a static batch photoreactor. Groundwater for Sustainable Development, 2020, 11: 100396.
[8] YUAN Y, ZHANG J, LI H. Simultaneous removal of SO_2, NO and mercury using TiO_2-aluminum silicate fiber by photocatalysis. Chemical Engineering Journal, 2012, 192: 21-28.
[9] CHEN M, CHU J W. NOx photocatalytic degradation on active concrete road surface — from experiment to real-scale application. Journal of Cleaner Production, 2011, 19(11): 1266-1272.
[10] WANG S, WANG S, WANG J. Achieving 20% Efficiency in Organic Solar Cells Through Conformationally Locked Solid Additives. Advanced Energy Materials, 2025: 2405205.
[11] M S A, HYAM R S. Au nanoparticles decorated titanium dioxide nanotube arrays with enhanced photocatalytic dye degradation under ultraviolet and sunlight irradiation. Results in Surfaces and Interfaces, 2023, 12: 100140.