



Mapping the Landscape of Soil Organic Carbon Research in China from 1995 to 2024: A Computational Analysis Using CiteSpace and CNKI Database

Jiashuo Xu^{1,2}; Zhouli Liu^{1,2,*}; Xin Wang^{1,3}; Shuyan Zhang^{1,3} and Mingran Xin^{1,2}

¹ College of Life Science and Engineering, Shenyang University, Shenyang, 110044, China

² Northeast Geological S&T Innovation Center of China Geological Survey, Shenyang, 110000, China

³ Key Laboratory of Black Soil Evolution and Ecological Effect, Ministry of Natural Resources, Shenyang, 110000, China

* Corresponding author: sydx2021@163.com

Abstract. Soil organic carbon (SOC) is critical for soil structure improvement, climate change mitigation, and ecological balance maintenance. This study leverages advanced computational tools to map the landscape of SOC research in China over the past three decades (1995–2024). Utilizing CiteSpace 6.3.R1 software, we conducted a comprehensive visual analysis of 17,020 relevant papers sourced from the China National Knowledge Infrastructure (CNKI) database. Our computational approach involved data mining, keyword extraction, and network visualization to uncover trends and hotspots in the field. The results indicate: (1) A significant increase in the number of publications, with average annual publications reaching 9,762, 1,654, and 2,038 in four distinct developmental periods, reflecting a growing research emphasis in China; (2) The University of Chinese Academy of Sciences emerged as the leading institution with 691 publications, though interdisciplinary collaboration remains limited; (3) Keyword frequency analysis identified biochar, microorganisms, community structure, and organic carbon components as current research hotspots. This study demonstrates the power of computational methods in analyzing large-scale scientific literature and provides valuable insights for future research directions in SOC.

Keywords: soil organic carbon; Citespace analysis; CNKI database; research landscape

1 Introduction

The terrestrial surface of our planet is enveloped by soil, a complex matrix composed of minerals, organic matter, water, and air. This dynamic medium is the product of prolonged interactions between rock weathering and various factors such as biology, climate, and topography. In China, soil serves not only as the foundation for plant growth, anchoring root systems and supplying essential nutrients and moisture, but also as a significant carbon reservoir. For instance, humus in China's forest soils can emit greenhouse gases in response to changes in temperature or land use patterns [1]. Soil

© The Author(s) 2026

H. Li et al. (eds.), *Proceedings of the 2026 5th International Conference on Big Data Economy and Digital Management (BDEDM 2026)*, Advances in Economics, Business and Management Research 383,

https://doi.org/10.2991/978-94-6239-640-1_8

fertility, which includes nutrient status and physicochemical properties, is crucial for agricultural productivity. In China, crop rotation involving legumes, rice, and maize has been shown to enhance soil fertility, thereby increasing crop yields [2]. The availability of key nutrients such as carbon, nitrogen, phosphorus, and potassium is vital for soil health. A well-structured soil with good aeration and permeability supports robust plant root development [3]. Soil organic carbon (SOC), derived from plant residues and microbial metabolites, plays a pivotal role in the global carbon cycle through its accumulation and decomposition processes. In China, SOC serves multiple functions: it releases nutrients during microbial decomposition to support plant growth, improves soil structure through the incorporation of organic carbon into inorganic forms (for example, coccolithophores in China's coastal regions [4]), helps reduce atmospheric CO₂ levels to mitigate climate change [5], and maintains the dynamic balance of soil carbon pools and ecosystem functions [6]. Over the past few decades, research on SOC in China has been on the rise. Studies have explored various aspects, such as the impact of fertilization frequency and cutting intensity on SOC storage in grassland ecosystems [7], the sensitivity of organic carbon decomposition to temperature changes in subtropical forests [8], organic carbon use efficiency at different soil depths in forest ecosystems [9], and the mechanisms by which artificial mangroves influence SOC dynamics in wetlands [10]. Despite the growing body of research, the field lacks comprehensive reviews and in-depth discussions on future research directions. CiteSpace, a powerful bibliometric visualization tool, has been widely used in diverse fields including social and environmental governance, artificial intelligence, botany, and microbiology. It offers a unique approach to analyzing citation relationships and revealing the underlying knowledge structure, evolutionary trajectories, and research hotspots through visual maps. By employing advanced algorithms for cluster analysis and representing documents or keywords as nodes connected by citations or co-occurrence relationships, CiteSpace can depict complex literature networks. Additionally, it identifies emerging research hotspots and frontiers through keyword burst analysis, thereby supporting discipline evaluation and the formulation of research policies. This study aims to leverage CiteSpace and the China National Knowledge Infrastructure (CNKI) database to map the landscape of SOC research in China from 1995 to 2024. By analyzing the relevant literature, we seek to identify cutting-edge trends and developmental directions in the field, provide forward-looking suggestions for future research, assist researchers in pinpointing potential research topics, transcend disciplinary boundaries, foster multidisciplinary communication and collaboration, and drive interdisciplinary development.

2 Materials

2.1 Data Source

The data of this study is derived from the China National Knowledge Infrastructure (CNKI) database, which brings together high-level academic research results across China, thus ensuring the authenticity and authority of the data. The time range for data retrieval is from January 1995 to December 2024. The topics searched were academic

research related to "soil organic carbon" and "SOC". After careful reading and screening, 11,321 academic journal articles that were not related to the topic or were repeated, and finally 17,020 academic journal articles closely related to the research topic were identified for subsequent analysis.

Analysis Methods.

CiteSpace is an advanced tool for visualizing and analyzing academic literature, which generates graphical representations through quantitative analysis methods [11,12]. The tool was originally developed by Dr. Chen and his research team in 2004 and has been continuously refined and disseminated since then [13].

At present, CiteSpace is widely used in research fields such as soil research and plant science. In this study, we performed a visual bibliometric analysis of relevant research records in the CNKI database with the help of CiteSpace 6.3.R1 (<https://citespace.podia.com/>). When analyzing literature growth trends and major publishing channels, we used data processed by CiteSpace. We then further refined and organized the data to construct visual graphics for bibliometric analysis. Since 1999, CiteSpace has been used to visually present the distribution of authors, research institutions and keywords. The software focuses on using connection lines to highlight the importance of each discipline and analyzes clustered connections between nodes [14-17]. Therefore, CiteSpace can accurately reveal the core content of research topics related to "soil organic carbon". In addition, we also analyzed the overall publication trends using Microsoft Office Excel 2021 and Origin 2022 and drew corresponding charts based on this.

3 Results and Discussion

3.1 Time Analysis of Literature Publication Volume

Figure 1, generated in Origin based on Excel statistics, presents the annual publication counts related to "soil organic carbon" and "SOC" from 1995 to 2024, thereby illustrating the research trajectory in this field. The x-axis represents the years (1995–2024), and the y-axis shows the number of publications (0–3000). The bar chart color transitions from green to red, reflecting the trend in publication volume. The data indicate a slow increase from 1995 to 2001, averaging 9 publications per year; a rapid growth from 2002 to 2014, averaging 762 per year; a slower increase from 2015 to 2021, averaging 1654 per year; and a sharp rise from 2022 to 2024, averaging 2038 per year. This steady upward trend shows growing attention and importance placed on this research area [18].

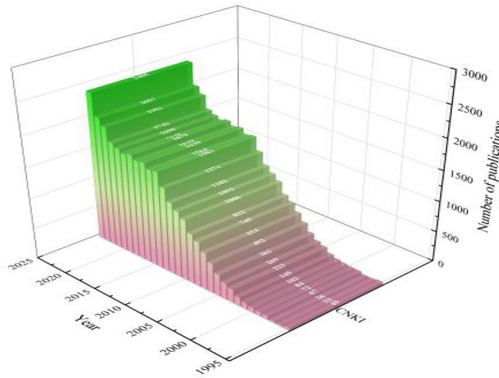


Fig. 1. The number of papers published in CNKI each year from 1995 to 2024.

3.2 Institutional Analysis

By constructing an institutional cooperation network knowledge graph, the scientific research level of various institutions in this field and their collaborative relationships can be clearly presented. Using CiteSpace software and data from the top ten domestic institutions ranked by the number of published papers, relevant charts were drawn (Table 1). Among them, the University of Chinese Academy of Sciences ranked first with 691 published papers; the College of Natural Resources and Environment ranked second with 335 papers; and another institution, University of Chinese Academy of Sciences, ranked third with 332 papers. This result indicates that the University of Chinese Academy of Sciences has a significantly higher number of published papers than the second-ranked College of Natural Resources and Environment, reflecting University of Chinese Academy of Sciences's high activity, abundant research resources, and strong innovation and output capacity in the research areas of "soil organic carbon" and "SOC."

Furthermore, Figure 2 presents the institutional co-occurrence network, containing $N=151$ nodes and $E=109$ connections, indicating that 151 institutions have published related literature on this research topic, with 109 collaborative relationships between institutions. In the figure, the low network density (0.0096) indicates that domestic research collaborations are fragmented and concentrated among a handful of core institutions rather than spanning the entire research community. Furthermore, the largest connected component encompasses only 65 % of the network's nodes; the remaining 35 % of institutions remain functionally isolated, underscoring the presently limited scope of collaboration. CiteSpace analysis identifies the Institute of soil and water conservation Chinese Academy of Sciences (betweenness centrality = 0.55), University of Chinese Academy of Sciences (0.48), and Graduate School, University of Chinese Academy of Sciences (0.47) as the principal hubs that function as "bridges" among disparate research communities. Conversely, numerous institutions exhibit zero or near-zero betweenness, coupled with a very low network density, indicating that most

3.3 Keyword Analysis

Research Hot Spots Analysis.

Research hot spots refer to a type of literature that shares a certain specific connection. The topics studied in these literatures are called research hot spots [19]. Keywords are the subject of the literature, and research hotspots are determined by the frequency of keywords. The higher the frequency of keywords, the more attention it receives from researchers [20]. Centrality is used to evaluate the importance and influence of nodes in facilitating connections between other nodes. Centrality is independent of the size of the node, but is based on the number of connections through a specific node. When the centrality is greater than or equal to 0.1, it indicates that this keyword is an important topic in the research field. By using the CiteSpace software, a keyword co-occurrence graph for soil science and agronomic research (Figure 3) and a high-frequency mediator central word list (Table 2) were created. The central word list is used to explain the keyword co-occurrence graph.

Subsequent keyword co-occurrence analysis revealed 148 nodes ($N = 148$) and 561 edges ($E = 561$) in Figure 3, underscoring the broad scope of the research domain and the strong interconnections among keywords. In the figure, each node represents a keyword, the size of the node is proportional to its occurrence frequency, and a larger node means a higher occurrence frequency; while the depth of the color indicates the focus of the keyword in recent research. The darker the color means the more attention it has received in the recent past [21]. It can be seen from Figure 3 that the research mainly focuses on core topics such as soil organic carbon, organic carbon and soil.

In summary, according to the data in Table 2 and Figure 3, "soil organic carbon" is the most frequent keyword, mainly because "soil organic carbon" is the core topic of the research, leading to its frequent appearance and high centrality in the literature. Although "organic carbon" ranks second in frequency, "organic carbon" has the highest centrality, indicating its more critical connecting role in the research network and that it is a key focus in the field. Besides "soil organic carbon" and "organic carbon," other high-frequency keywords include "soil," "soil nutrients," "soil aggregates," and "straw return." Among them, "soil" ranks third in both frequency and betweenness centrality, reflecting its importance as a research subject in this field. "Soil" mainly relates to the study of organic carbon storage locations and influencing factors. Additionally, most other keywords are related to soil science and agronomy, indicating domestic researchers' strong focus on soil health, sustainable soil use, and environmental management.

Table 2. Table of high-frequency keywords.

Rank	Keywords	Count	Central
1	Soil Organic Carbon	2481	0.27
2	Organic Carbon	1541	0.76
3	Soil	1034	0.23
4	Soil Nutrients	911	0.14
5	Soil Aggregation	561	0.39

6	Straw Incorporation	450	0.12
7	Soil Physicochemical Properties	439	0.06
8	Biochar	410	0.06
9	Soil Enzyme Activity	388	0.12
10	Soil Respiration	351	0.05

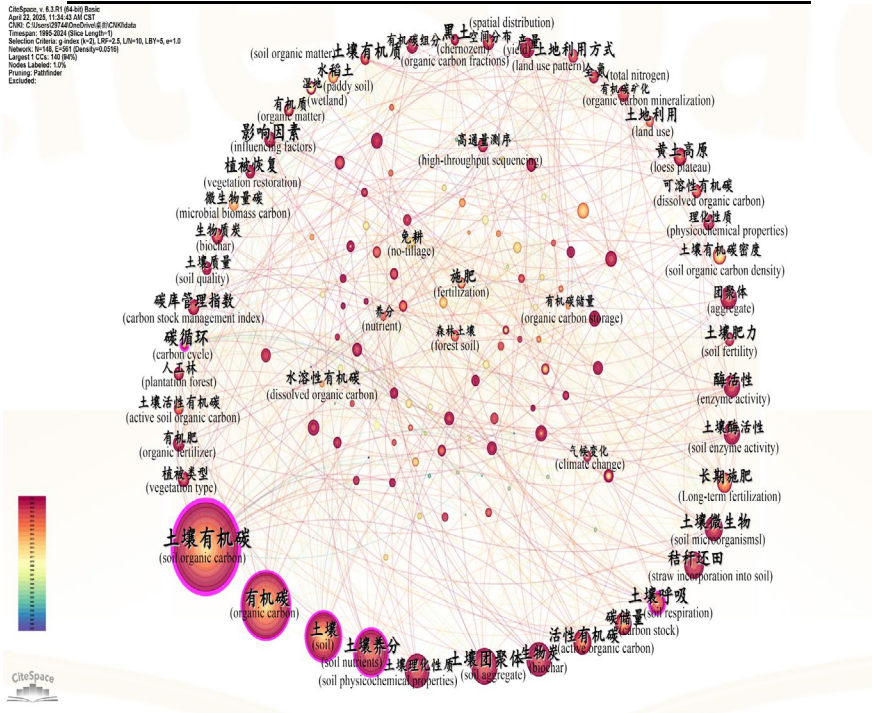


Fig. 3. Keyword co-occurrence network map.

3.4 Research Trend Analysis

Using the clustering function of CiteSpace software, all keywords were categorized and their evolutionary trends over time were displayed. Clustering groups closely related keywords, with each cluster number corresponding to a different cluster value. In Figure 4, the keywords are divided into 11 categories. Detailed information is provided in the upper left corner of the figure, where the Q value measures clustering quality ($Q > 0.3$ indicates significant modularity) and the S value assesses clustering consistency ($S > 0.5$ indicates more reasonable clustering) [22]. Figure 4 shows a Q value of 0.825 and an S value of 0.9469, indicating good clustering performance that clearly reflects the categories and characteristics of the keywords.

Subsequently, Figure 4 uncovers several key clusters and their temporal evolution. Based on the nature of each cluster, it can be observed that Cluster group #2 (Soil Organic Carbon), Cluster group #3 (Soil Physical and Chemical Properties), and Cluster

group #5 (Soil Aggregation) are all related to soil structure; whereas Cluster group #1 (Soil Respiration), Cluster group #9 (Carbon Cycling), and Cluster group #10 (Enzyme Activity) are all associated with the dynamic movement of carbon within the soil.

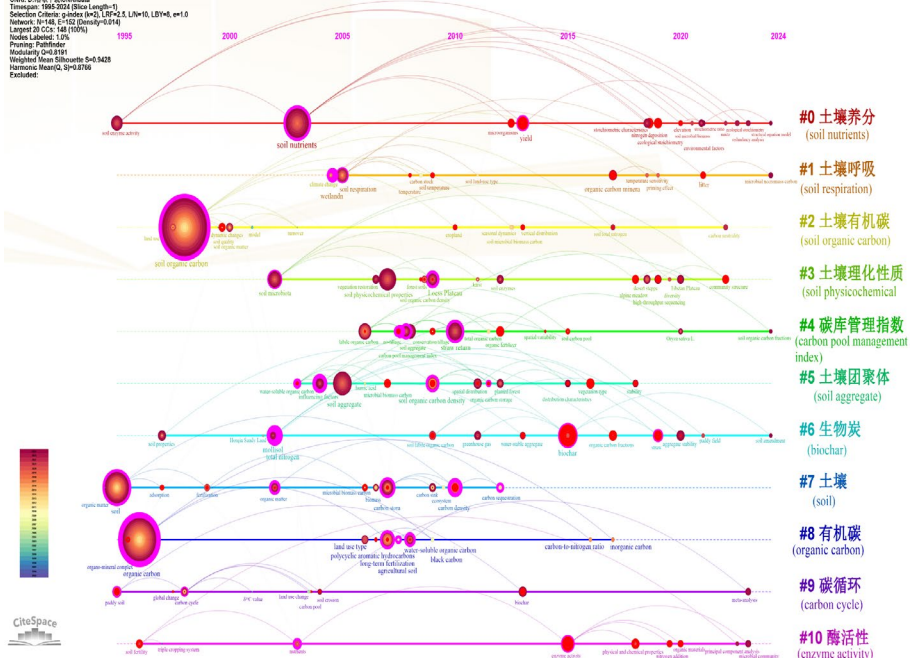


Fig. 4. Chronological keyword map.

Keywords	Year	Strength	Begin	End	1995 - 2024
soil	1995	31.11	1995	2005	-----
adsorption	1997	35.53	1997	2011	-----
carbon cycle	1998	43.14	1998	2012	-----
land use	1998	33.33	2004	2013	-----
wetland	2005	23.65	2005	2011	-----
soil respiration	2005	18.80	2005	2012	-----
microbial biomass carbon	2007	42.36	2007	2017	-----
long-term fertilization	2007	20.57	2007	2015	-----
no-tillage	2008	24.69	2008	2015	-----
soil organic carbon density	2009	19.73	2009	2013	-----
carbon density	2010	30.94	2010	2016	-----
carbon stock	2007	22.10	2010	2016	-----
organic carbon mineralization	2017	23.00	2017	2022	-----
ecological stoichiometry	2019	20.63	2019	2024	-----
high-throughput sequencing	2019	19.60	2019	2022	-----
alpine meadow	2018	18.77	2020	2024	-----
biochar	2015	21.28	2021	2024	-----
microorganism	2013	19.32	2021	2024	-----
community structure	2022	18.91	2022	2024	-----
organic carbon fractions	2017	17.61	2022	2024	-----

Fig. 5. Keywords exhibiting the most pronounced citation bursts ranked by intensity.

The clustering relationships can be primarily categorized into two major types: The first type of soil structure: Cluster group #2 focuses on soil organic carbon, with keywords such as soil organic matter, soil quality, croplands, seasonal dynamics, total soil nitrogen, and carbon sequestration. This cluster highlights the primary forms of carbon storage, which underpin soil fertility and carbon sink functions. Cluster group #3 involves soil physicochemical properties, including soil microbes, vegetation restoration, organic carbon density, alpine meadows, soil enzymes, and community structure. This cluster illustrates the factors that directly influence the stability and decomposition rates of organic carbon. Cluster group #5 centers on soil aggregates, with keywords including water-soluble organic carbon, humic acid, soil organic carbon density, spatial distribution, and speculative carbon storage. This cluster shows how aggregate structure protects organic carbon from microbial decomposition, a key mechanism for carbon sequestration. The second type of carbon cycle in soil: Cluster group #1 centers on soil respiration, encompassing terms related to climate change, temperature, land use models, organic carbon mineralization, and litter. Soil respiration, which first appeared in 2005, is the most representative keyword, indicating carbon output and soil metabolic activity. Cluster group #9 centers on the carbon cycle, with keywords such as paddy soils, soil erosion, land use change, biochar, and meta-analysis. Cluster groups #10 centers on enzyme activity, with keywords including soil fertility, nutrients, nitrogen addition, and microbial communities.

These 10 clustering groups can be divided into three stages: The first stage focuses on basic static properties, including soil physical and chemical properties, the content and distribution of soil organic carbon, and soil aggregates, addressing the core questions of "where is carbon located?" and "how much carbon exists?". The second stage takes soil respiration and enzyme activity as key indicators, and combines the carbon cycle framework to analyze the processes of carbon transformation and loss as well as their microscale driving mechanisms, responding to the question of "how is carbon transformed and lost?". The third stage uses biochar as a representative regulatory measure to intervene in the core links of the first two stages, and relies on the carbon pool management index (CPMI) to evaluate the size, activity, and stability of carbon pools, solving the applied questions of "how to manage and improve soil carbon pools?" and "what is the effect of such management?".

The forefront of research refers to the large number of emerging research-related topics that appear during specific stages of study. Identifying these frontier domains is pivotal for grasping the latest disciplinary advances, forecasting future research trajectories, and uncovering latent hotspots. [23]. The frequency of keyword appearance, that is, the number of occurrences of a certain word in the abstract, keyword list and title, can reflect the hot issues in the field and thus reveal its cutting-edge development direction [24]

Figure 5 presents the burst term graph of the soil organic carbon field from 1995 to 2024 generated using CiteSpace software, analyzing the top 20 keywords by impact. Research in this field can be divided into two phases: The first phase (1995–2016) primarily focused on terms such as "soil," "carbon cycle," "land use," "soil respiration," "microbial biomass carbon," "long-term fertilization," "carbon density," and "carbon storage." This indicates that since 1995, research in the field has been exploring soil

carbon-related issues, with a core focus on “soil” as the medium, gradually extending to key terms like “adsorption,” “carbon cycle,” and “soil respiration,” forming a research framework centered on “soil carbon cycling and its maintenance mechanisms.” The second phase (2016–2024) shifted focus to “organic carbon mineralization,” “ecological stoichiometry,” “high-throughput sequencing,” “biochar,” “microbes,” and “organic carbon components,” with 2016 marking a key transition point in the field’s research. The focus transitioned from the basic mechanisms of soil carbon cycling in the first phase to more application-oriented and interdisciplinary directions. In this second phase, emerging keywords such as “biochar,” “microbes,” “community structure,” and “organic carbon components” reveal the field’s cutting-edge directions and the rise of interdisciplinary research.

4 Conclusions and Prospects

This study employs bibliometric analysis and visualization techniques to examine soil organic carbon-related literature from the China National Knowledge Infrastructure (CNKI) database over nearly 30 years. It reviews publication dates, annual output, achievements of domestic research institutions, research hotspots, and future trends.

The number of research papers in this field has steadily increased over the past three decades, with research interest gradually rising in three phases: a stable growth period from 1995 to 2014, with an average of 762 papers per year; a slow growth period from 2014 to 2021, averaging 1,654 papers annually; and a rapid growth phase from 2022 to 2024, with an average of 2,038 papers per year. This trend reflects researchers’ growing attention to the field.

Among domestic research institutions, the University of Chinese Academy of Sciences has published the most papers, with a total of 691 articles from 1995 to 2024, demonstrating its emphasis and technical leadership in this domain. Currently, cooperation among research institutions is relatively close, but interdisciplinary collaboration remains limited, mostly confined within respective fields. Promoting effective interdisciplinary cooperation could stimulate innovation potential in the future.

Over time, soil organic carbon research has deepened, evolving from preliminary exploration to focusing on areas such as biochar, community structure, and organic carbon components. Emerging keywords in recent years indicate current major research directions and will likely be key focuses going forward.

In China, soil organic carbon research is highly valued for its important role in maintaining the global carbon cycle, promoting sustainable agriculture, protecting biodiversity, and improving soil quality. Continued research in this area holds significant scientific and practical value for addressing global challenges.

Funding

This study is financially supported by Science and Technology Plan Joint Project Natural Science Foundation-General Program of Liaoning Province (2024-MSLH-506), LiaoNing Revitalization Talents Program (XLYC2203070), the funding project of

Northeast Geological S&T Innovation Center of China Geological Survey (QCJJ2022-44).

References

1. Dhillon GS; Van Rees KCJ. Soil organic carbon sequestration by shelterbelt agroforestry systems in Saskatchewan. 2017; *Can. J. Soil Sci.* 97(3): 394-409.
2. AU Rahman MH; Islam MR; Jahiruddin M; et al. Integrated nutrient management in maize-legume-rice cropping pattern and its impact on soil fertility. 2013; *J Food Agric Environ*, 1111(11): 648-652.
3. M.R Carter. Researching structural complexity in agricultural soils, 2004; *Soil Tillage Res*, 79(1): 1-6.
4. Godrijan J; Drapeau DT; Balch WM. Osmotrophy of dissolved organic carbon by coccolithophores in darkness. *New Phytol.* 2022; 233(2): 781-794.
5. Wang, MY; Shi, XZ; Yu, DS; et al. Regional differences in the effect of climate and soil texture on soil organic carbon. 2013, *Pedosphere*, 23(6): 799-807.
6. L. Yang; J. Pan; Y. Shao; et al. Soil organic carbon decomposition and carbon pools in temperate and sub-tropical forests in China. *J. Environ. Manage.* 2007, 85(3): 690-695.
7. Zhang, J.; Meng, Q.; Tang, Y.; et al. Study on the response mechanism of biogas slurry and chemical fertilizer combination application on the physiological characteristics of summer maize leaves. *Arch. Agron. Soil Sci.*, 2025, 71(1), 1–15.
8. Di Ma; Yu Sun; Min Liu; et al. Higher endogenous labile organic carbon decreases the temperature sensitivity of soil organic matter decomposition in two subtropical forests. *Appl. Soil Ecol.*, 2025, 26(105816): 0929-1393.
9. Pei, J.; Li, J.; Luo, Y.; et al. Patterns and drivers of soil microbial carbon use efficiency across soil depths in forest ecosystems. *Nat. Commun.*, 2025, 16, 5218.
10. Liao Haoyu, Hu Minjie, Wang Jingtao, et al. The impact of mangrove restoration on soil organic carbon storage in wetlands and its controlling factors. *Acta Scientiae Circumstantiae*, 1-9 [2025-06-15].
11. Che, S.; Kamphuis, P.; Zhang, S.; et al. A visualization analysis of crisis and risk communication research using CiteSpace. 2023; *Int. J. Environ. Res. Public Health*, 19, 2923.
12. Yang, X.; Liu, Q. Research foundation and hotspot analysis of urban road ecology—a bibliometric study based on CiteSpace. 2024; *Sustainability*, 16, 5135.
13. Qiu, Y.; Liao, K.; Zou, Y.; et al. A bibliometric analysis on research regarding residential segregation and health based on CiteSpace. *Int. J. Environ. Res. Public Health*, 2022, 19, 10069.
14. Chen, B.; Shin, S.; Wu, M.; et al. Visualizing the knowledge domain in health education: A scientometric analysis based on CiteSpace. *Int. J. Environ. Res. Public Health*, 2022, 19, 6440.
15. Pang; Xue Meng; et al. Analysis of research hotspots in COVID-19 genomics based on citespace software: Bibliometric analysis. *Front. Cell. Infect. Microbiol.* 2022, 12, 1060031
16. Liu, Y.; Li, Q.; Li, W.; et al. Progress in urban resilience research and hotspot analysis: a global scientometric visualization analysis using CiteSpace. *Environ Sci Pollut Res*, 2022, 29, 63674–63691.
17. Xu, C.; Yang, T.; Wang, K; et al. Knowledge domain and hotspot trends in coal and gas outburst: a scientometric review based on Cite Space analysis. *Environ Sci Pollut Res*, 2023, 30, 29086–29099.

18. Yang, X.; Liu, Q. Research foundation and hotspot analysis of urban road ecology—a bibliometric study based on CiteSpace. *Sustain.* 2024, 16(12): 5135.
19. Li, W.; Sun, B.D. The knowledge structure and research hotspots of western economic geography—a quantitative analysis based on CiteSpace. *Econ. Geogr.*, 2014, 34(4): 7-12+45.
20. Gao, C.; Wang, R.; Zhang, L.; et al. Visualization analysis of CRISPR gene-editing knowledge map based on Citespace. *Biol. Bull.*, 2021, 48: 705-720.
21. Wang, W.; Lu, C. Visualization analysis of big data research based on Citespace. *Soft Comput.*, 2020, 24(11): 8173-8186.
22. Seyler, B. C.; Luo, H.; Wang, X.; et al. Assessing the impact of urban greenspace on physical health: An empirical study from Southwest China. *Front. Public Health.*, 2023, 11: 1148582.
23. Geng, Y.; Zhu, R.; Maimaituerxun, M. Bibliometric review of carbon neutrality with CiteSpace: evolution, trends, and framework. *Environ. Sci. Pollut. Res.*, 2022, 29(51): 76668.
24. Su, Y.X.; Huang, G.Q.; Chen, X.Z.; et al. Research progress on the ecological environmental effects of urban green spaces. *Acta Ecol. Sin.*, 2011, 31(23): 302-315.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

