Review

Not peer-reviewed version

Artificial Intelligence in Natural Carbon Sink Research: A Scientometric Review and Evolutionary Analysis (2001-2025)

Jianhua Ma , Yongzhang Zhou * , Luhao He *

Posted Date: 1 September 2025

doi: 10.20944/preprints202509.0050.v1

Keywords: artificial intelligence (AI); natural carbon sink; carbon estimation; remote sensing modeling; keyword evolution; scientometric analysis

Review

Artificial Intelligence in Natural Carbon Sink Research: A Scientometric Review and Evolutionary Analysis (2001–2025)

Jianhua Ma [1,2,3,4], Yongzhang Zhou [1,2,3,4,*] and Luhao He [1,2,3,4,*]

[1] Sun Yat-sen University, Center for Earth Environment and Earth Resources, Zhuhai 519000, Guangdong Province, China

[2] Sun Yat-sen University, School of Earth Sciences and Engineering, Zhuhai 519000, Guangdong Province, China

[3] Key Laboratory of Geological Processes and Mineral Resources Exploration, Zhuhai 519000, Guangdong Province, China

[4] Sun Yat-sen University, Institute of Carbon Neutrality and Green Development, Zhuhai 519000, Guangdong Province, China

* Correspondence: zhouyz@mail.sysu.edu.cn (Y.Z.); helh8@mail2.sysu.edu.cn (L.H.)

Abstract

Under the global imperative of carbon neutrality, artificial intelligence (AI) has emerged as a transformative force in enhancing the monitoring, assessment, and governance of natural carbon sinks. This study presents a comprehensive scientometric analysis of AI-enabled research on natural carbon sinks from 2001 to 2025, based on data from the Web of Science Core Collection. By applying co-word network construction, clustering analysis, and evolutionary trajectory mapping, we characterize the methodological progression, thematic structure, and temporal dynamics of this fast-evolving field. The analysis reveals four developmental phases: Emergence (2001-2010), Initial Growth (2011-2017), Acceleration (2018-2021), and Expansion (2022-2025). We observe a paradigm shift from early machine learning methods—such as support vector regression and basic neural networks—toward ensemble algorithms and deep learning architectures. Keyword evolution highlights the prominence of terms like "machine learning," "soil organic carbon," and "forest biomass," reflecting a methodological loop of remote sensing, ecological modeling, and predictive simulation. Geographically, the field exhibits a China-led research trend with increasing international collaboration. This work outlines a structural and technological roadmap for the application of AI in carbon sink research, while also addressing key challenges such as algorithmic adaptability, data heterogeneity, and multi-scale model integration. The findings offer strategic guidance for future studies and contribute to intelligent carbon sink governance in the era of climate transition.

Keywords: artificial intelligence (AI); natural carbon sink; carbon estimation; remote sensing modeling; keyword evolution; scientometric analysis

---

1. Introduction

Climate change has emerged as one of the most pressing environmental challenges of the 21st century, with increasingly frequent and severe extreme weather events, ecosystem degradation, and socio-economic impacts (Global CCS Institute, 2023, Intergovernmental Panel On Climate Change, 2022). To mitigate global warming, the Paris Agreement established a clear objective of limiting the global average temperature increase to well below 2°C, while pursuing efforts to cap it at 1.5°C (Paris Agreement, 2015). In response, numerous countries and regions have proposed "carbon peaking" and "carbon neutrality" strategies, advancing coordinated emission reductions through energy transitions, industrial decarbonization, and natural ecosystem enhancement (Halkos and Aslanidis, 2024, He et al., 2022). Within the carbon neutrality framework, a two-pronged approach that balances carbon sources and sinks has become critical, with carbon sequestration technologies playing a pivotal role in bridging the carbon budget gap in hard-to-abate sectors (Bartell, 2024, Fuss et al., 2014). Consequently, advancing research on carbon sink systems—particularly their sequestration mechanisms and storage stability—and optimizing their integration into global climate governance have become key priorities in contemporary climate science and environmental policy research.

In contrast to technological approaches such as Carbon Capture and Storage (CCS), natural carbon sinks are increasingly recognized for their ecological synergies and sustainability benefits. Forests, grasslands, soils, wetlands, and blue carbon ecosystems—including mangroves, seagrass beds, and salt marshes—sequester atmospheric CO_2 through mechanisms such as photosynthesis, soil carbon fixation, and oceanic carbon pumping, representing one of nature's most vital negative emission pathways (Friedlingstein et al., 2024, Huang et al., 2024, Lal, 2004). Terrestrial ecosystems are estimated to absorb approximately 30% of annual fossil fuel CO_2 emissions (Friedlingstein et al., 2024), while also providing essential ecosystem services such as climate regulation, water conservation, and biodiversity protection (Piao et al., 2022). Marine ecosystems, as the planet's largest active carbon reservoir, have absorbed about one-third of anthropogenic CO_2 emissions since the Industrial Revolution (Doney et al., 2009). However, natural carbon sinks exhibit significant spatiotemporal variability in carbon fluxes due to complex interactions among climate change, land use, soil properties, and anthropogenic management. These dynamics pose substantial challenges for accurate quantification and long-term prediction (Green and Keenan, 2022, Kim et al., 2024, Searchinger et al., 2018). Traditional ecological monitoring methods are limited in scalability, timeliness and data consistency, restricting their ability to meet the precision requirements for carbon sink monitoring under carbon peak and neutrality strategies (Piao et al., 2022, Wang et al., 2020). Accordingly, enhancing our understanding of carbon cycling processes and developing modeling approaches capable of integrating heterogeneous data sources while accounting for spatiotemporal complexities has become a research imperative.

Recent advances in Artificial Intelligence (AI) have opened new avenues for natural carbon sink research, providing powerful analytical tools. Machine Learning (ML) and Deep Learning (DL) techniques—capable of extracting complex patterns and nonlinear relationships from heterogeneous, multi-source data—have found wide application in various aspects of carbon sink studies (Bracarense et al., 2022, Reichstein et al., 2019a). Convolutional Neural Network (CNN) has proven particularly effective in processing remote sensing imagery for accurate estimation of forest carbon stocks and biomass distribution (Li et al., 2020). Ensemble learning methods, including Support Vector Machine (SVM) and Random Forest (RF), have been used to predict soil organic carbon content, land use changes, and carbon emission fluxes (Padarian et al., 2019, Soman and Indu, 2022). More recently, advanced architectures such as Graph Neural Network (GNN) have been applied to model spatiotemporal variables in complex ecosystems, enabling the analysis of spatial dependencies and propagation mechanisms in carbon cycling (Mei et al., 2022). Emerging approaches, including reinforcement learning and digital twin, show great potential for ecological simulation, decision-making optimization, and system feedback analysis (Rolnick et al., 2022, Sahu and Upadhyay, 2024). These AI technologies are rapidly becoming indispensable for enhancing monitoring accuracy, improving process simulations, and forecasting carbon sink dynamics, thereby laying the groundwork for intelligent ecosystem carbon management.

Despite the growing application of AI in natural carbon sink research, the field remains fragmented and lacks a unified research framework. Most existing studies have focused on evaluating the applicability of specific AI algorithms (e.g., CNN, RF) within single ecosystem types (e.g., forests or soils), without establishing a systematic framework for multi-ecosystem and multi-model comparison and integration (Wang et al., 2025). Additionally, the geographical distribution of existing studies is highly uneven, with research largely dominated by developed countries. In contrast, developing regions—particularly tropical and coastal areas—remain significantly underrepresented, which limits the generalizability and scalability of global carbon sink models

(Velastegui-Montoya et al., 2022, Zhong et al., 2023). Moreover, from the perspective of literature evolution, there remains a lack of systematic mapping of the research trajectory, clustering analysis of hotspots, and identification of emerging trends in the cross-disciplinary field of AI-enabled natural carbon sink, resulting in unclear research directions and repeated efforts in subsequent studies (Ma et al., 2025, Peng et al., 2023, Ruiz-Sánchez et al., 2024, Wang et al., 2023). Most existing reviews focus either on methodological evaluations or discussions of ecological processes, but lack comprehensive analyses based on knowledge graphs and bibliometric networks, which hampers the efficient allocation of research resources, the development of academic collaboration networks, and the understanding of technological evolution (Masria and Abouelsaad, 2025, Pimenow et al., 2025, Wong et al., 2021). Therefore, conducting systematic bibliometric analyses to trace research topic evolution, identify technological hotspots, uncover institutional collaboration patterns, and map interdisciplinary linkages has become essential for advancing the integration of AI technologies with natural carbon sink.

Building upon the aforementioned research background and challenges, this study adopts a bibliometric approach to systematically analyze and visualize the global research landscape of AI-enabled natural carbon sink studies. Drawing on the Web of Science (WoS) Core Collection, we construct a high-precision keyword framework encompassing "AI technologies" and "natural carbon sink systems," while excluding confounding terms related to CCUS and other unrelated technical pathways. Our final dataset includes 3,885 publications as of June 10, 2025. Using three major bibliometric tools (VOSviewer, CiteSpace, and Bibliometrix), we conduct multidimensional quantitative analyses to examine: (1) Temporal evolution of publication trends; (2) Collaborative networks among high-impact countries, institutions, and authors; (3) Co-occurrence relationships and clustering patterns of core keywords; (4) Temporal progression and frontier identification of thematic areas. Methodologically, we integrate co-word analysis, co-citation networks, burst detection, and research trajectory mapping to construct a comprehensive knowledge graph and research evolution framework. At the application level, this study systematically examines the development of AI applications in key subfields of natural carbon sink research, including forest carbon monitoring, soil carbon modeling, and blue carbon estimation. It further identifies interdisciplinary convergence points with strong integration potential. By doing so, the study sheds light on the foundational landscape and evolving trajectories of the interdisciplinary field of AI and natural carbon sinks, uncovers key academic communities and emerging technological pathways, and provides a theoretical foundation and knowledge base to support intelligent monitoring and management of natural carbon sinks under future carbon neutrality strategies.

2. Research Methodology and Data Collection

2.1. Data Collection and Search Strategy

To delineate the research landscape, thematic evolution, and methodological convergence of AI applications in natural carbon sink studies, this study developed a structured literature retrieval strategy, which serves as the data foundation for the subsequent bibliometric analysis. All bibliographic data were extracted from the WoS Core Collection as of June 10, 2025. To ensure comprehensive coverage of this inherently interdisciplinary domain, we included publications categorized as Article, Review, and Proceedings Paper, without restrictions on publication year or language.

A dual-axis thematic framework was adopted to guide the query construction, targeting the intersection of AI technologies and natural carbon sink systems. For the AI dimension, the keyword set encompassed major algorithmic paradigms and representative models, including: AI, ML, DL, neural networks, reinforcement learning, transfer learning, SVM, GNN, transformer, digital twin, large language model (LLM), pretrained models, and ChatGPT, among others. The natural carbon sink axis incorporated terms reflecting carbon accumulation mechanisms and modeling processes in terrestrial and marine ecosystems, such as: forest carbon sequestration, soil organic carbon, wetland carbon, blue carbon, carbon flux, biomass mapping, carbon stock, net primary productivity, and ecosystem simulation. This ensured the inclusion of both aboveground and belowground components across forest, soil, wetland, and coastal systems.

To address the potential semantic noise stemming from the wide application of AI in unrelated disciplines (e.g., medicine, social sciences, chemical engineering), we implemented a three-level semantic filtering protocol: Logical Exclusion (Boolean NOT): (1) Logical exclusion mechanism (NOT operation): Systematically exclude non-target terms frequently found in interdisciplinary fields, such as CO_2 capture, membrane separation, CCS, and sentiment analysis, which are related to CCUS pathways, the medical field, social sciences, and other boundary disciplines; (2) Discipline category restriction mechanism: Using the subject classification system of WoS, actively exclude fields like medicine and social sciences, which are not directly related to natural carbon sink ecological processes, thereby improving the ecological relevance and knowledge structure consistency; (3) Manual iteration and keyword co-occurrence feedback mechanism: Conduct manual checks and co-occurrence analysis on a small sample of the initial search results, restricting irrelevant content and ambiguous terms (e.g., "SOC" which can refer to both soil organic carbon and state of charge) through contextual restrictions, ensuring the representativeness of the samples and accuracy of clustering.

Given the insufficient standardization of terminologies in both AI and ecological domains, we used dual-form expressions combining full terms and acronyms (e.g., "support vector machine" or "SVM"), and introduced representative model names as semantic probes to identify the actual penetration of AI techniques in ecological modeling literature. Following semantic refinement and multiple rounds of quality control, a final dataset of 3,885 qualified publications was obtained. These records were exported in BibTeX format for Bibliometrix, and in Plain Text format for VOSviewer and CiteSpace, thus providing a robust dataset for constructing knowledge graphs and analyzing evolutionary research trajectories.

2.2. Data Analysis Tools and Methodological Framework

To systematically explore the knowledge structure, research hotspots, and temporal evolution of AI applications in natural carbon sink research, we employed a multi-tool bibliometric framework composed of VOSviewer, CiteSpace, and Bibliometrix. Each tool offers distinct analytical advantages, forming a complementary triad aligned with the analytical triad of "structure—evolution—statistics"(Figure 1). This hybrid approach is widely validated in bibliometric studies across complex interdisciplinary domains, such as AI in climate research and ecological remote sensing, supported by solid theoretical foundations and proven operational effectiveness.

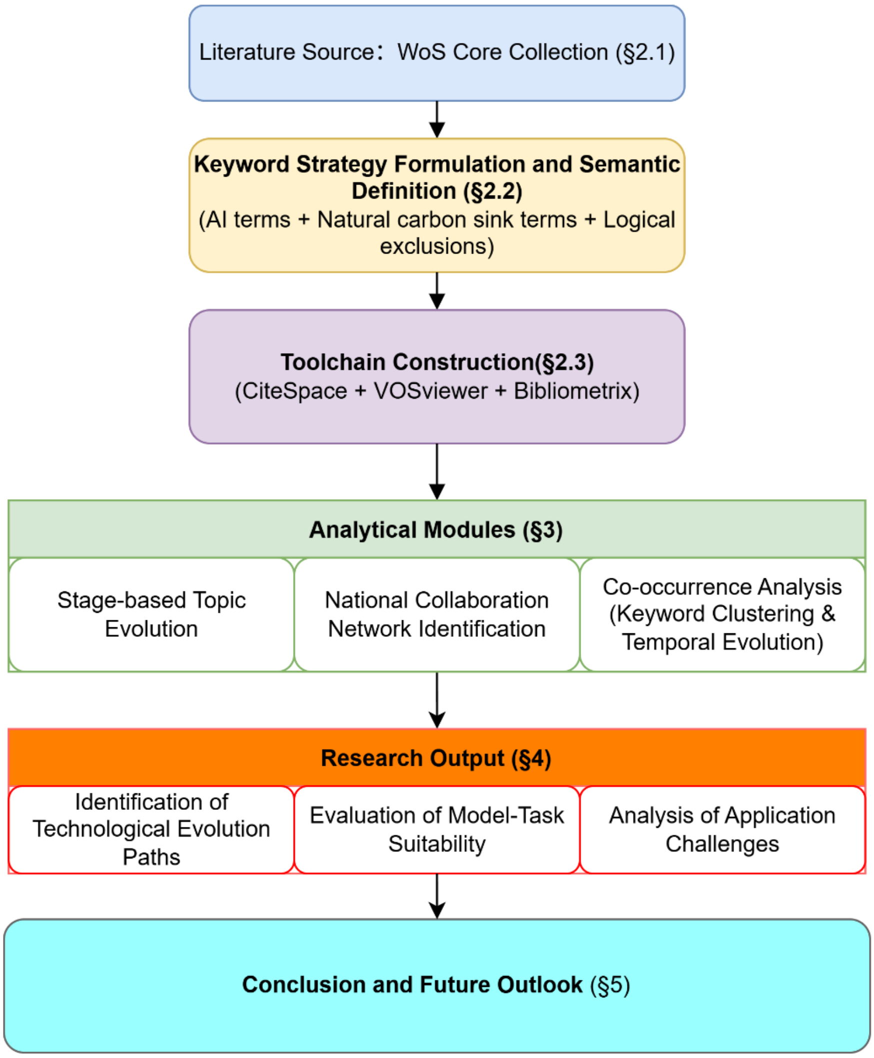

Figure 1. Methodological Framework for AI-enabled Natural Carbon Sink Knowledge Graph Construction.

First, VOSviewer was used to construct keyword co-occurrence networks, institutional and national collaboration maps, and research topic clustering (Wong, 2018). Based on a distance-based clustering algorithm for visualization, the tool automatically identifies co-occurrence relationships and the degree of aggregation between terms, making it suitable for revealing the coupling structure of AI methods and research clusters in natural carbon sink systems. By setting thresholds for co-occurrence frequency and minimum connection strength, high-frequency terms in the research can be effectively extracted, and clustering visualizations can be generated to identify core research topics and key technological paths.

Next, CiteSpace is primarily used for co-citation analysis, emerging term detection, and the temporal evolution analysis of research topics (Chen, 2016). The tool offers functions such as an emergence detection algorithm, time slicing, and evolution path recognition, making it suitable for uncovering the evolutionary stages and technological trends in a research field. In this study, CiteSpace is used to identify the structure of the knowledge base, build a literature co-citation network, and explore the evolution of AI methods within the intersection of AI and natural carbon sink research by detecting emerging terms and analyzing keyword evolution. Key parameters include time slicing, node type settings (e.g., keywords or cited papers), and the emergence detection algorithm (Kleinberg algorithm), all executed according to standard practices.

Finally, Bibliometrix, an R-based bibliometric analysis tool, is primarily used for macro statistical analysis and visualization, including annual publication trends, journal distribution, country/institution output comparisons, and the generation of topic evolution maps (Aria and Cuccurullo, 2017). This tool supports comprehensive data cleaning and statistical modeling for literature in BibTeX format, making it effective for handling large datasets and multi-dimensional analyses. Using core functions such as biblioAnalysis(), thematicMap(), and threeFieldPlot(), this study presents research activity and structure distribution from a statistical perspective, providing a quantitative foundation for the structural and evolutionary maps.

Regarding the operational process, Bibliometrix is first used for preliminary annual trend analysis and sample structure statistics. Next, data in Plain Text format is imported into VOSviewer to build keyword co-occurrence and research networks. Finally, CiteSpace processes the sample with time slicing, performing evolution path and emergence analysis. All tools follow official documentation and established methodologies to ensure the standardization, repeatability, and quality of the visualization results.

3. Results and Analysis

3.1. Phase Division

Based on trends in annual publication volume and average citation frequency, the integration of AI in natural carbon sink research exhibits clear temporal evolution patterns. From 2001 to 2025, related research transitioned from early theoretical exploration to large-scale application and the integration of multiple technologies. The number of publications increased rapidly from single digits to nearly a thousand per year, while citation indicators show periodic fluctuations in research impact and structural changes. By combining technological evolution with changes in ecological policy, this study divides the field's development into four distinct stages: the Emergence Stage (2001–2010), the Initial Growth Stage (2011–2017), the Acceleration Stage (2018–2021), and the Expansion Stage (2022–2025). This division reflects both the depth of AI's integration into natural carbon sink systems and the shift in research topics from "technology introduction" to "system modeling" (Figure 2).

During the Emergence Stage (2001–2010), the publication volume was very low, with an average of fewer than 12 articles per year, yet the average citation rate was relatively high, indicating that, despite the scarcity of early research, it had significant academic impact and exploratory value. Research mainly focused on the application of early ML methods, such as the traditional Artificial Neural Network (ANN) and Stochastic Gradient Boosting (SGB), in applications such as carbon stock prediction and soil parameter fitting. Data sources primarily relied on experimental sites or local monitoring systems, with limited support from large-scale remote sensing (Bricklemyer et al., 2006, Moffat et al., 2007, Rossel and Behrens, 2010). Externally, the formal adoption of the Kyoto Protocol in 2005 marked the incorporation of carbon sink mechanisms into the global climate governance framework, explicitly recognizing the role of natural ecosystems, such as forests, agriculture, and land-use change, in mitigating climate change (Canadell and Raupach, 2008, Protocol, 1997). This provided both institutional support and guidance for the future integration of AI methods into ecological carbon modeling. Overall, this phase was interdisciplinary and exploratory, with AI not yet central to modeling, but its technical feasibility was initially validated, and several highly-cited papers established the groundwork for future research.

During the Initial Growth Stage (2011–2017), with advancements in ecological remote sensing, geostatistical modeling, and DL algorithms, AI's application in natural carbon sinks gradually expanded, and publications steadily increased. The number of publications rose from 16 in 2011 to 83 in 2017, with an average citation rate of approximately 5, indicating a transition towards a more systematic approach. Regarding algorithmic applications, ensemble learning models, such as RF, were widely used for forest aboveground carbon estimation, significantly enhancing the accuracy of remote sensing data modeling in low-biomass forests (Safari et al., 2017, Wiesmeier et al., 2011). The ANN outperformed traditional statistical models in Soil Organic Carbon (SOC) modeling, particularly demonstrating strong nonlinear modeling capabilities in various soil layers of forest ecosystems (Vahedi, 2017). In terms of data, the use of remote sensing platforms like Landsat-8, MODIS, and Sentinel greatly improved the spatial coverage and data quality of carbon sink system modeling (Dąbrowska-Zielińska et al., 2016, Kumar et al., 2017, Safari et al., 2017). From a policy perspective, the 2014 IPCC Fifth Assessment Report (AR5) clearly emphasized the crucial role of terrestrial ecosystem carbon sinks in controlling atmospheric greenhouse gas concentrations and mitigating global warming, highlighting the need for strengthened long-term monitoring and modeling capabilities for carbon stocks and fluxes in natural systems such as forests, soils, and land-use change (Change, 2014). This consensus further encouraged the academic community to focus on technological integration and improved modeling accuracy in ecosystem carbon cycle modeling (Griscom et al., 2017). This stage saw the establishment of the basic technical framework and collaborative network for AI and natural carbon sink research, laying the foundation for large-scale expansion.

During the Acceleration Stage (2018–2021), research entered a period of rapid growth, with the number of publications doubling over four years, from 137 in 2018 to 325 in 2021, and citation frequency remaining high. Mainstream AI algorithms, such as CNN, Long Short-Term Memory (LSTM), and RF, were systematically applied to subfields like carbon flux simulation, carbon stock estimation, and blue carbon ecosystem remote sensing inversion (Besnard et al., 2019, Khan et al., 2021). Technically, the integration of heterogeneous data from multiple sources, including remote sensing, meteorology, and land use, became standard, facilitating the transition of carbon sink systems from static spatial evaluation to dynamic spatio-temporal modeling (Sun et al., 2019, Zhang et al., 2021). From a policy perspective, since the Paris Agreement officially came into effect in 2016, major global economies have successively introduced timelines and roadmaps for carbon peaking and carbon neutrality. As an integral part of achieving net-zero emissions, the demand for accurate quantitative evaluation and dynamic simulation of natural carbon sinks has increased significantly (Delbeke et al., 2019, Paris Agreement, 2015, Streck et al., 2016). With the growing demand for modeling accuracy and an improved understanding of complex systems in natural carbon sink research, the integrated application of AI methods has accelerated. In particular, in fields like soil carbon and terrestrial biomass, researchers have progressively advanced interdisciplinary integration and collaborative development across ecological science, remote sensing technology, and computational intelligence through AI modeling platforms driven by multi-source data (Padarian et al., 2019). In this phase, AI technologies are no longer just tools but have become core components of natural carbon sink modeling, leading to a significant shift in research paradigms.

During the Expansion Stage (2022–2025), the annual number of publications rapidly surpassed 600, peaking at 921 in 2024, though the average annual citation frequency significantly declined. From a methodological perspective, new-generation AI methods such as LLM (e.g., ChatGPT, LLaMA), GNN, and reinforcement learning have been widely applied in carbon sink modeling, carbon footprint estimation, and green AI design (Fu et al., 2024, Nie et al., 2025). In terms of application, carbon sink research has gradually expanded from forests and soils to blue carbon systems like mangroves and salt marshes, with estimation methods evolving from static carbon stock estimation to dynamic carbon flux modeling, incorporating multi-factor regulatory mechanisms involving climate, vegetation, and soil (Alongi, 2023, Wang et al., 2023, Wang et al., 2024). On the policy side, the rise of ESG (Environmental, Social, and Governance) investment and “Nature-based Solutions (NbS)” has become a global hot topic, driving a surge in demand for AI tools in carbon accounting and ecological monitoring (La Notte, 2024, Ogunyemi, 2023a, Sangha et al., 2024). However, this rapid expansion has also exposed issues such as insufficient model generalization and weak theoretical explanations, leading to significant variation in overall literature quality (Liu et al., 2025). This period is characterized by a “high-density technological penetration—fluctuating research quality—the coexistence of theory and application” composite feature, marking a critical turning point in the evolution of the field from engineering to intelligent systems.

 doi:10.20944/preprints202509.0050.v1

In summary, the AI-enabled natural carbon sink research has undergone a threefold progression from the introduction of methods, through system construction, to the explosive expansion of application scenarios. The four-phase evolutionary pathway not only clearly reflects the changing technological position of AI methods in natural carbon sink research, but also maps the multiple influences of global carbon neutrality policies, ecological remote sensing technologies, and data infrastructure evolution on the shaping of research paradigms. This phase division lays the groundwork for the subsequent analysis of research topic evolution, hotspot technology clustering, and academic collaboration structures. It also provides historical and logical support for understanding how AI technology has achieved rapid penetration in natural carbon sink research.

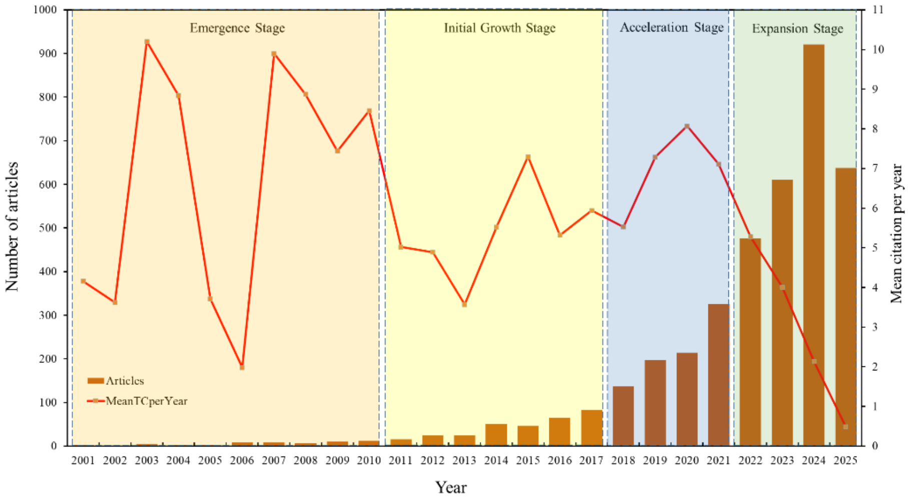

Figure 2. Annual Publication Volume and Average Citation Frequency.

3.2. National Research Landscape and Collaboration Network

In the ongoing evolution of research on AI-enabled natural carbon sink, the national research landscape not only reflects the heterogeneity of technological diffusion and policy mobilization but also illuminates the participation pathways and influence mechanisms of the global scientific community in this interdisciplinary domain. To accurately characterize the development trends in this field, Chapter 3 focuses on three key dimensions: publication output, research productivity, and collaboration networks. Based on bibliometric and author contribution data, it analyzes the output structure, stage evolution, and collaboration patterns of leading research countries.

From a national perspective, global research on AI-enabled natural carbon sink displays significant spatial concentration and hierarchical differences in influence, with research primarily concentrated in a few major scientific powers, such as China, the United States, Germany, India, and Australia. As shown in Table 1, China leads with a total of 1,772 publications, accounting for 45.6% of the global total, reflecting its dominant output capacity in this field. However, its average citation frequency is only 14.90, which is relatively low compared to its publication volume, indicating a "high output, low citation" pattern. This may be due to the focus of some publications on applied research or the uneven influence of uneven influence of publications in English. In contrast, the United States, with 487 publications, ranks second but boasts a significantly higher average citation frequency of 29.60, suggesting that its research has a broader global influence and a higher proportion of highly cited papers. While Germany, Australia, and France have relatively limited publication volumes (none exceeding 200 papers), their average citation frequencies are 51.40, 55.50, and 31.10, respectively, all significantly higher than the global average. This highlights their strong research capabilities in areas such as AI modeling, ecological remote sensing, and blue carbon systems, particularly in core issues like ecosystem predictive modeling and carbon flux estimation. Additionally, Multi-Country Publications reflect the tendency for collaboration in different countries:

Germany (55.5%), Australia (46.7%), and France (54.9%) have relatively high levels of international collaboration, leading to greater cross-border citations and the dissemination of methodologies. Conversely, China (27.7%) and India (28.5%) have lower collaboration rates, suggesting that their research is more domestically focused, posing a risk of knowledge isolation. Overall, at the national level, the field exhibits a typical structure: countries with high publication output are primarily concentrated in East and South Asia, whereas countries with high research influence are mainly located in Europe, North America, and Oceania. This reflects a complex global research ecosystem marked by the coexistence of "quantity-quality separation" and "core-periphery" dynamics. Such a pattern highlights the importance of fostering cross-regional collaboration and strengthening international research exchange—particularly in advanced modeling approaches and data-sharing mechanisms—that demand enhanced global cooperation.

Table 1. National Publication and Citation Statistics.

Rank	Country	Articles	Frequency	%	MCP[1]%	Total Citations	Avg Citations
1	CHINA	1772	8901	45.6	27.7	26333	14.90
2	USA	487	2765	12.5	32.9	14395	29.60
3	GERMANY	164	916	4.2	55.5	8422	51.40
4	INDIA	144	599	3.7	28.5	2058	14.30
5	AUSTRALIA	122	677	3.1	46.7	6772	55.50
6	CANADA	97	527	2.5	34	1719	17.70
7	BRAZIL	92	484	2.4	42.4	2090	22.70
8	IRAN	87	318	2.2	63.2	2417	27.80
9	FRANCE	71	476	1.8	54.9	2205	31.10
10	ITALY	66	329	1.7	45.5	2168	32.80

[1]MCP=Multi-Country Publications/Total。.

To gain deeper insight into the national-level evolution of research on AI-enabled natural carbon sink, this study selects five representative countries: China, the United States, Germany, Australia, and India. These nations include high-output countries (China and the United States), high-impact contributors (Germany and Australia), and rapidly emerging participants (India). By integrating the growth trajectories of cumulative author appearances from 2001 to 2025 (Figure 3) with total citation counts and collaboration rates (Table 1), this study systematically identifies heterogeneous patterns in research participation intensity and scholarly quality across countries. In the Emergence Stage (2001–2010), the United States was the only country to exhibit continuous growth, with cumulative author appearances increasing from 2 to 68. This trend reflects the early maturity of AI technologies and the country's leading capacity to access remote sensing data (e.g., MODIS products), which facilitated the early integration of ML methods into ecosystem modeling (Justice et al., 2002, Reichstein et al., 2019b). In contrast, the cumulative author frequency in the other four countries remained near zero, with China's first contributions appearing in 2006. This delay was primarily attributed to underdeveloped remote sensing infrastructure and the lack of ecosystem carbon sink data in China at the time (Schimel et al., 2015). It also mirrors the international trend following the enactment of the Kyoto Protocol in 2005, wherein natural carbon sink mechanisms began to be gradually embedded into national science and technology agendas. In the Initial Growth Stage (2011–2017), all five countries experienced rising author frequencies, with China showing particularly notable growth—from 12 to 290. This surge reflects the rapid convergence of AI and ecological remote sensing applications during the late stages of China's 12th Five-Year Plan. Germany and Australia also demonstrated steady progress, with cumulative author appearances reaching 109 and 89, respectively. Although their absolute contributions remained modest, their average citation frequencies were 51.4 and 55.5, and their international collaboration rates were 55.5% and 46.7% (Table 1), significantly surpassing those of the United States (29.6%), China (14.9%), and India (14.3%).

These figures suggest a research paradigm oriented toward high-quality international collaborations, representing a model of "low quantity, high quality." Meanwhile, India's author frequency began at zero and grew slowly, gradually establishing its foundational research network. In the Acceleration Stage (2018–2021), the widespread application of DL, image recognition, and remote sensing-integrated modeling in carbon stock estimation and carbon flux prediction drove a dramatic rise in China's author frequency, which surged to 1,667—surpassing the United States (1,149) for the first time and positioning China as the most active contributor in the field. This trend was fueled by China's "carbon peak and carbon neutrality" policy (proposed in 2020) and significant advances in remote sensing capabilities. The United States maintained steady growth, while Germany and Australia, though contributing at a lower volume, continued to exert substantial research influence. In the Expansion Stage (2022–2025), disparities in research expansion became even more pronounced. Within three years, China added over 7,000 author appearances, reaching a total of 8,901—far exceeding the United States (2,765) and other countries, forming a "high-density participation" research cluster. However, as indicated in Table 1, China's average citation frequency was only 14.9, lagging behind the United States and most European nations, indicating room for improvement in the international dissemination and impact of its research. Germany and Australia maintained their steady growth trajectories, while India increased sharply to 599 author appearances, emerging as one of the fastest-growing developing nations in terms of research participation. In conclusion, based on the composite indicators of cumulative author appearances and citation impact, the research trajectories of the five countries in the field of AI-enabled natural carbon sink exhibit a typical pattern of temporal misalignment and heterogeneous participation structures. China exemplifies research density expansion driven by policy initiatives; the United States reflects technological maturity and strong citation conversion; Germany and Australia represent high international collaboration and quality-driven models; and India is entering a phase of rapid ascent. This structural pattern underscores the varied national positions in the integration of AI and ecological carbon sinks and highlights the need to enhance research influence and promote deeper international collaboration moving forward.

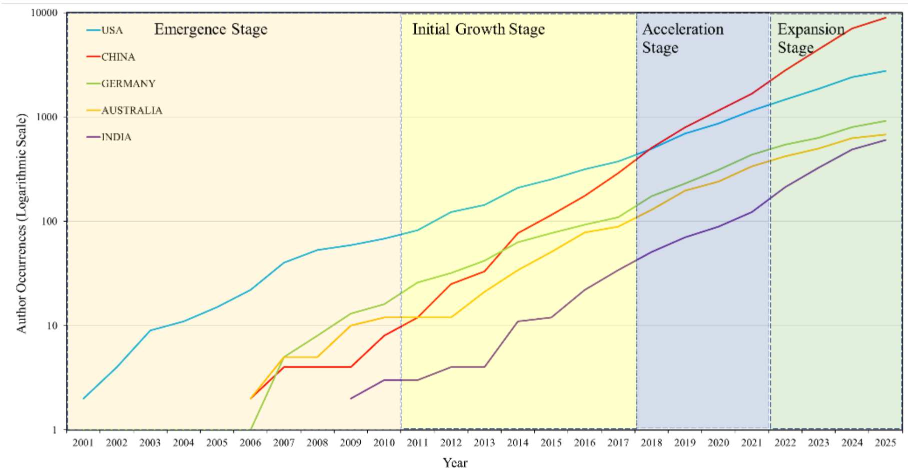

Figure 3. A logarithmic Y-axis is used to balance inter-country output gaps and better reflect stage-wise growth trends.

International cooperation serves as a critical driver for the rapid advancement of research on AI-enabled natural carbon sink. In particular, cross-national collaboration in areas such as remote sensing data sharing, ecosystem monitoring, and carbon estimation model development helps mitigate the limitations imposed by individual countries' technological and data resource constraints. As illustrated in Figure 4, the current international collaboration network exhibits a multilateral structure anchored by China and the United States, with several European and Oceania countries forming high-intensity cooperative clusters. Overall, the network displays a distinct "core-satellite" distribution pattern. In terms of cooperation frequency, the China–U.S. bilateral partnership is the most prominent, with 238 recorded instances—far surpassing any other country pair. This pattern aligns closely with the leading positions of both countries in terms of publication output (China: 1,772 articles; U.S.: 487 articles) and cumulative author appearances (China: 8,901; U.S.: 2,765) (see Table 1). Such intensive collaboration may stem from the complementary strengths of the two nations in natural resource monitoring, remote sensing infrastructure, AI technology development, and global climate governance frameworks (Dai et al., 2023, Han et al., 2024, Meng et al., 2022, Wang et al., 2022). Beyond the China–U.S. axis, China has also developed frequent cooperative ties with countries such as Australia (90 instances), Germany (64), the UK (63), and France (56), forming a regional collaboration network centered on China. Conversely, the United States functions more as a transatlantic hub, maintaining high-frequency partnerships with Germany (85), Canada (55), the UK (59), and France (49). This "dual central-hub" structure suggests that both China and the United States not only dominate in research output but also serve as key coordinators within the global academic network. Germany stands out for its exceptional engagement in international collaboration, maintaining active ties with multiple European countries—including Switzerland, the Netherlands, France, and Italy—each with over 30 instances. This reflects the high level of integration within the European research ecosystem, which is likely bolstered by the EU's systematic support for smart ecological technologies through initiatives such as the "Green Deal," "Digital Transformation," and the "Carbon Neutrality Law" (Fetting, 2020). Although Germany and Australia have comparatively lower publication outputs, they have established high-intensity collaborative networks beyond China–U.S. partnerships. Notable connections include Germany–U.S. (85 instances), Australia–China (90), Germany–France (36), and Germany–Switzerland (38). These patterns align with their high international collaboration rates (Germany: 55.5%; Australia: 46.7%) (see Table 1) and correlate with their leading positions in average citation frequency (Germany: 51.4; Australia: 55.5). Moreover, the Germany–Australia collaboration has reached 32 instances, particularly in fields such as energy cooperation, carbon accounting systems, and regional governance mechanisms (Miehe et al., 2022a). In addition, countries such as France, Canada, the United Kingdom, the Netherlands, and Sweden have established multilateral collaboration chains with major research leaders including China, the United States, Germany, and Australia. These nations rank among the top twenty in collaboration frequency (e.g., China–France: 56; U.S.–UK: 59; Germany–Netherlands: 34). While they may not lead in publication output, they play critical supporting roles in global carbon sink modeling, blue carbon monitoring, and ecological data sharing within multinational projects—highlighting their strategic function as "bridge countries" in the global research network (Miehe et al., 2022b). Notably, while India has witnessed rapid growth in author appearances (599 instances in recent years), its international collaborations remain limited (e.g., U.S.–India: 27; India–Germany: 24), with an international collaboration rate of only 28.5%. This indicates that although India is expanding its domestic research activities, its global cooperation network remains underdeveloped—a common pattern among many developing nations. This highlights the necessity of enhancing model generalizability and fostering broader international collaboration. Overall, the international collaboration network in AI-enabled natural carbon sink research reveals a typical structure of "U.S.–China dominance, multi-country engagement, and regional cluster intensification." The China–U.S. axis forms the backbone of global collaboration, while countries like Germany and Australia—with high collaboration rates and academic influence—anchor the European research core. Meanwhile, countries such as the UK, France, and Canada serve as strategic bridges connecting various research communities. This structure not only accelerates the adaptation of AI-based modeling methods across diverse carbon sink systems but also provides a foundational framework for the establishment of a global carbon monitoring system.

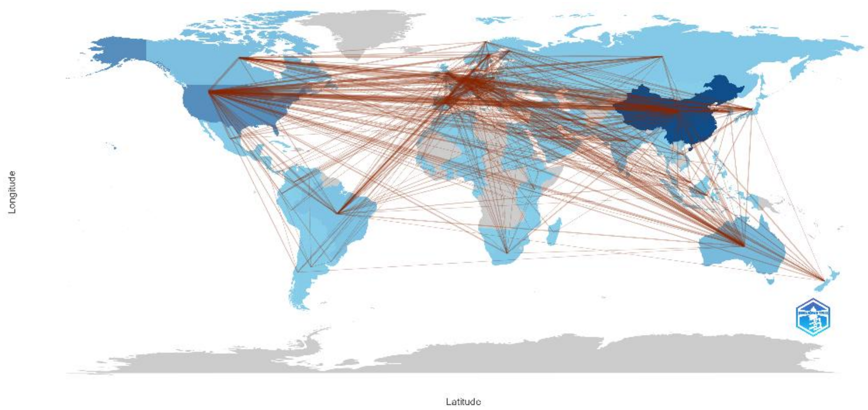

Figure 4. International Cooperation Map.

In summary, the current national participation pattern in AI-enabled natural carbon sink research exhibits clear hierarchical differentiation and structural heterogeneity: On the one hand, China and the United States form a "dual-axis structure" in terms of publication volume and author engagement, representing two distinct research trajectories—high-density output and high-impact citation performance, respectively. On the other hand, countries such as Germany, Australia, and France, despite publishing fewer articles, demonstrate a "high-quality and high-collaboration" research model. Their higher international collaboration rates and average citation counts further reinforce the cohesion of the European research ecosystem. Notably, developing countries such as India have experienced rapid growth in recent years, driven by national policy initiatives and advancements in domestic ecological modeling capabilities. Although their international cooperation networks remain underdeveloped, their research potential is considerable. From the perspective of cooperation structure, the global research network follows a "dual-core with multiple nodes" pattern centered on China and the United States, with each country occupying distinct functional roles—reflecting their organizational capacities and network connectivity within the global landscape of ecological intelligence modeling. These findings also underscore that the current AI-enabled natural carbon sink field faces systemic challenges, including imbalanced regional research capacities, barriers to data sharing, and limited cross-regional model adaptability. Moving forward, it is essential to promote cross-regional model integration, enhance collaborative remote sensing data sharing, and advance policy-driven research initiatives on a global scale to facilitate the seamless application of AI in the intelligent monitoring and management of natural carbon sinks.

3.3. Keyword Clustering and Research Topic Identification

Keywords, as condensed representations of scholarly knowledge, not only capture the semantic orientation of research content but also reflect evolving research hotspots and technological trends over time. In the interdisciplinary domain of AI-enabled natural carbon sink, keyword co-occurrence patterns and clustering reveal the diverse applications of AI methods across different natural carbon sink subsystems (e.g., forests, soils, wetlands), while also illustrating the shifting knowledge consensus and research priorities within the scientific community at various developmental stages.

This study first analyzed the frequency of all keywords and visualized the results using a word cloud (Figure 5) to highlight the distribution of high-frequency terms. Keyword frequency not only reflects the dominant research themes but also provides insight into the interdisciplinary convergence of research topics, methodological frameworks, and ecosystem classifications. Statistical analysis reveals a bimodal clustering pattern within the keyword landscape. On one hand, core methodological terms such as ML (617 occurrences), RF (555), and DL (169) underscore the central role of AI, encompassing key algorithmic families including supervised learning, ensemble methods, and neural networks. On the other hand, domain-specific terms such as SOC (388), forest biomass (155), aboveground biomass (297), and net primary productivity (99) reflect the diversity of carbon

 doi:10.20944/preprints202509.0050.v1

sink systems under investigation, including forests, soils, vegetation, and blue carbon ecosystems. Additionally, keywords related to remote sensing platforms—such as remote sensing (320), lidar (220), Sentinel-2 (148), and Landsat (117)—appear frequently, reflecting the field's strong reliance on spatial data acquisition and remote sensing inversion methods. This indicates a typical research trajectory of "remote sensing data → AI modeling → carbon sink estimation." Furthermore, general-purpose terms such as prediction (404), classification (355), model (306), and uncertainty (147) are frequently observed across studies, indicating that the field encompasses not only carbon stock estimation but also broader issues such as uncertainty analysis, ecosystem simulation, and land-use classification. The interweaving of algorithmic and application-oriented terminology highlights the interdisciplinary nature of AI-enabled natural carbon sink research, which is marked by a strong coupling between methodological approaches and real-world application scenarios.

Building upon the identification of high-frequency keyword distribution, this study further investigates the thematic structure and knowledge segmentation of AI-enabled natural carbon sink research through cluster analysis and structural indicator modeling of core keywords. Based on co-word association strength and semantic clustering results, Figure 6 identifies two primary research clusters: Cluster 1, centered on the "ML and vegetation modeling approach," and Cluster 2, focusing on "soil carbon estimation and digital soil mapping applications." Cluster 1 is defined by core keywords such as ML, RF, and aboveground biomass, with a co-word frequency of 74, a centrality score of 0.34, and an impact value of 5.164. These metrics suggest moderate network connectivity and a relatively high methodological significance within the domain of ecological remote sensing modeling. In contrast, Cluster 2 is represented by keywords such as SOC and digital soil mapping, with a co-occurrence frequency of 176, a centrality score of 0.425, and an impact value of 6.865 — all higher than those of Cluster 1. These indicators suggest that Cluster 2 holds a more active and influential position within the overall research network. To validate the stability and coherence of the clustering structure, Callon's structure index were introduced in this study to model the cohesion and centrality of each cluster. As shown in Figure 7, a two-dimensional distribution map constructed based on Callon Density and Callon Centrality reveals that the "ML" cluster has a centrality score of 1.489 and a density score of 4.09, indicating its role as a strong connective hub within the thematic network, with high structural stability and the second-highest centrality ranking. In contrast, the "SOC" cluster has a centrality score of 1.263 and the highest density score of 4.178. Although its centrality is slightly lower than that of the former, the high density indicates tightly interconnected internal keywords and a more focused research direction, characterizing it as a typical "high-aggregation, high-cohesion" knowledge cluster. In terms of the Cluster Frequency indicator, the "ML" theme contains a total of 12,420 keyword occurrences, slightly higher than that of "SOC" (10,455). This indicates broader presence in the literature and stronger interdisciplinary reach, encompassing fields such as forest modeling, image classification, and time-series prediction. Conversely, the latter cluster, with its more specialized disciplinary focus, may offer enhanced explanatory depth and improved model accuracy. In summary, current research on AI-enabled natural carbon sink exhibits a dual-structured thematic framework: a broad, methodology-oriented cluster centered on "remote sensing—modeling—vegetation monitoring" (Cluster 1), and a focused, application-oriented cluster revolving around "soil—carbon—fluxgeostatistics" (Cluster 2). Each cluster demonstrates distinct advantages in terms of semantic characteristics, indicator performance, and technological preferences. This dual-core structure not only delineates the internal organization of the research network but also offers a solid foundation for future exploration of sub-theme evolution and cross-thematic integration.

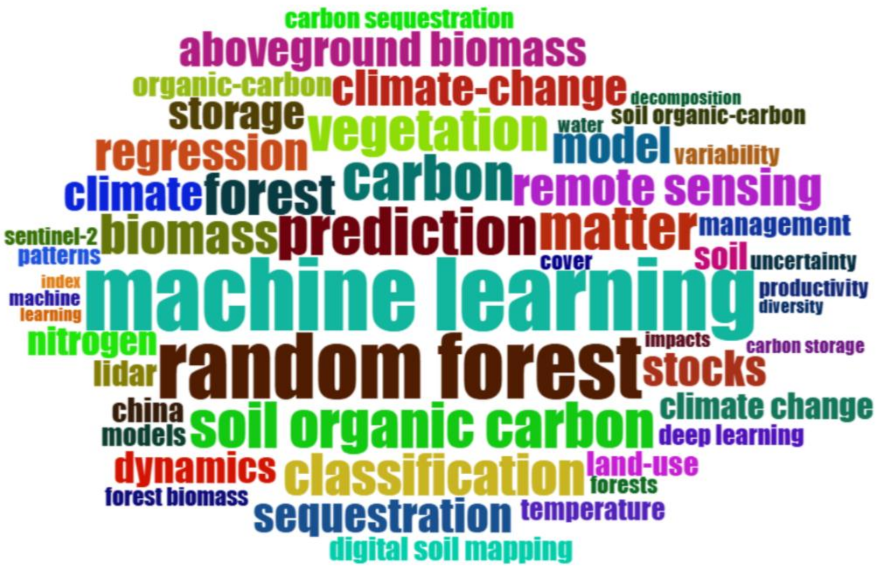

Figure 5. Keyword Word Cloud Visualization.

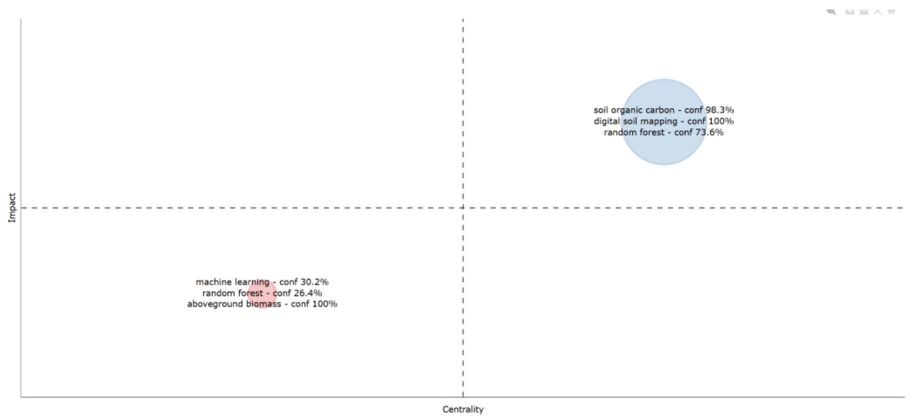

Figure 6. Based on Co-word Association Strength and Semantic Clustering.

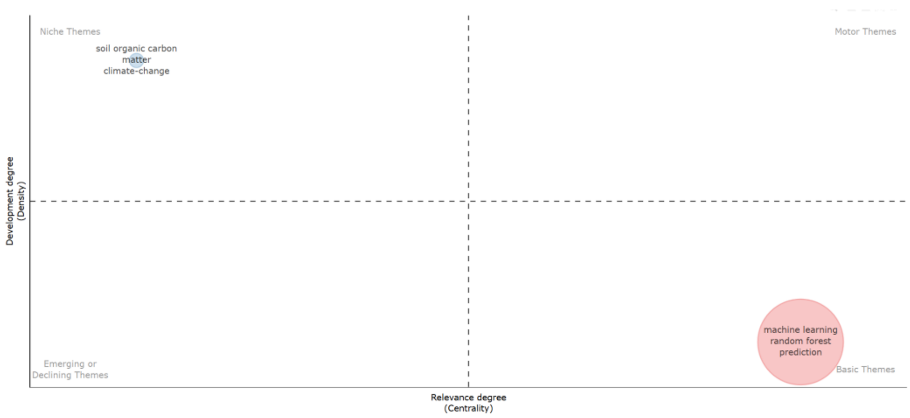

Figure 7. Two-Dimensional Distribution Map Constructed Based on Callon Density and Callon Centrality Coordinates.

To further clarify the semantic structure and core directions of research on AI-enabled natural carbon sink, this study constructed a semantic clustering map based on a keyword co-occurrence network (Figure 8). Relying on the structural properties of the co-word network and its modularity partitioning results, two major thematic clusters were identified. Each cluster consists of a group of keywords that exhibit intrinsic coupling in terms of research objects, technological applications, and methodological approaches, reflecting the structural characteristics of research themes in this interdisciplinary field. Cluster 1 (red) centers on remote sensing image processing and vegetation carbon estimation. It includes keywords such as remote sensing, classification, vegetation, aboveground biomass, forest biomass, LiDAR, Sentinel-2, and cover. This cluster follows a research trajectory utilizing high-resolution satellite imagery (e.g., Landsat, Sentinel series) and spatial recognition algorithm (e.g., image classification, object segmentation) to estimate the carbon stocks and biomass of forest and vegetation systems. The core terms remote sensing and classification exhibit high betweenness centrality and PageRank values, serving as bridging nodes across multiple thematic areas and underscoring their pivotal role in natural carbon sink monitoring and the integration of AI methodologies. This research direction emphasizes continuous spatial coverage, accuracy validation, and the integration of multi-source datasets. Representative methodologies include the synergistic inversion of optical remote sensing and LiDAR data, along with the application of DL-based image classification models (e.g., CNN, U-Net) for producing high-resolution carbon storage maps. Cluster 2 (blue) centers on AI-driven modeling and carbon cycle prediction, with ML and RF as its core elements. It includes keywords such as SOC, prediction, regression, sequestration, carbon, climate change, land use, and temperature. These terms span research domains such as soil carbon estimation, carbon flux modeling, land-use change impact assessment, and studies on climate-related mechanisms. Several keywords (e.g., SOC, carbon sequestration, regression) exhibit high PageRank scores and strong internal cohesion, underscoring their significant influence within the semantic network and establishing them as stable core hotspots in the research domain. This line of research commonly adopts ensemble learning approaches (e.g., RF, XGBoost) or deep neural networks to construct nonlinear carbon stock prediction models, with particular attention to uncertainty estimation and regional generalizability—highlighting the robust representational capacity of AI in modeling complex ecological processes. Overall, the keyword clustering reveals an emerging dual-core structure in AI-enabled natural carbon sink research, characterized by the parallel development of two distinct pathways: a spatial identification path driven by remote sensing imagery and a process modeling path empowered by ML. While the two clusters diverge in terms of research focus (forest vs. soil) and methodological orientation (image recognition vs. process simulation), the semantic network exhibits clear signs of convergence and cross-cluster integration. Notably, keywords such as carbon storage, forest, and model are positioned near the cluster boundaries, reflecting their shared applicability and bridging function across both research pathways. This phenomenon of "boundary terms" highlights future opportunities for tighter interdisciplinary integration, particularly in areas such as data fusion, multi-scale coupling, and ecological model synthesis. To further investigate the temporal dynamics of this semantic structure and track shifts in research priorities, the next section conducts a burst detection and temporal distribution analysis to construct a chronological knowledge map of AI-enabled natural carbon sink research.

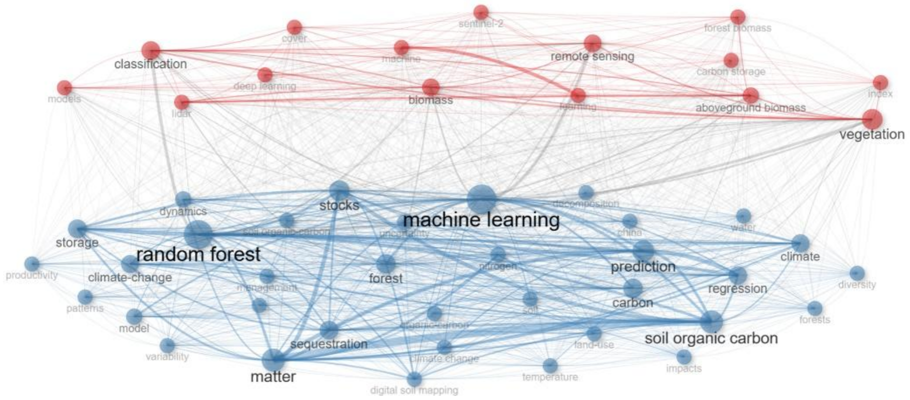

Figure 8. Keyword Clustering Structure.

In summary, the keyword clustering and co-word network analysis reveal the structural patterns of current research on AI-enabled natural carbon sink and establish a coherent knowledge framework for understanding its developmental trajectories, thematic intersections, and technological evolution. To further trace the evolution of research topics and shifting focal areas within this domain, the following section presents a temporal analysis of keyword dynamics. By employing burst detection and annual frequency trend analysis, it outlines the progression of core themes and maps the trajectory of frontier development in this interdisciplinary field.

3.4. Evolution of Research Hotspots and Methodologies

As research on AI-enabled natural carbon sink continues to expand, its thematic focus demonstrates a clear temporal progression. Keywords, as key indicators of knowledge structures, provide insights into both the diffusion of technologies and the shifting of research priorities, as reflected by their frequency and time of emergence. Accordingly, this section introduces a two-dimensional analytical framework that combines a time-series analysis of high-frequency keywords (2001–2025) with a Q1–Q3 emergence time metric to categorize terms into three types—early-sustained, mid-term emerging, and recently-bursting—thereby illuminating the mechanisms underlying thematic transitions. This framework also complements the previously discussed research phases and clustering structures, further enhancing the temporal modeling of knowledge evolution.

Based on the word frequency over time (Figure 9), the field of AI-enabled natural carbon sink between 2001 and 2025 exhibits characteristics of multi-thematic progression and intertwined evolution of techniques and targets. Most keywords display exponential growth after 2020, indicating the widespread penetration of AI methods into natural ecosystem research. However, the early-stage trends reveal significant temporal misalignment and divergence in initial adoption patterns. In terms of appearance time and growth rhythm, terms related to ecological entities—such as forest, biomass, and vegetation—entered the research system significantly earlier than AI-related terms and maintained relatively stable frequencies during 2001–2010. This reflects a research stage primarily focused on remote sensing observations and carbon stock assessments, with limited AI involvement. In contrast, AI-related terms such as RF, prediction, and ML began to appear gradually only after 2012, and their frequency growth remained modest before 2020. This indicates that AI methods initially played a supplementary role in modeling and had not yet established a method-driven research paradigm. Around 2020, the popularization of DL and high-performance computing triggered an explosive rise in AI-related keywords. ML, for instance, surged from 32 mentions in 2018 to 617 in 2025, becoming one of the fastest-growing terms in the field, signifying its transition from an auxiliary tool to a central analytical approach. Notably, keywords such as SOC, carbon sequestration, and classification also exhibited steep growth after 2020, reflecting the increasing demand for AI-driven techniques in subfields like soil carbon modeling and carbon source identification. Meanwhile, terms such as matter, vegetation, and carbon, though not inherently AI-specific, showed parallel increases, suggesting that AI advancements have contributed to the broader quantification of ecological processes. In addition, the trend curves show a sharp convergence in keyword frequency growth after 2022, which may be closely associated with the widespread adoption of models such as ChatGPT and Transformer-based models, the release of large-scale remote sensing foundation models, and the implementation of carbon neutrality strategies. These transformative developments have collectively driven the field into a deeper phase characterized by a “dual engine” of technological tools and application scenarios.

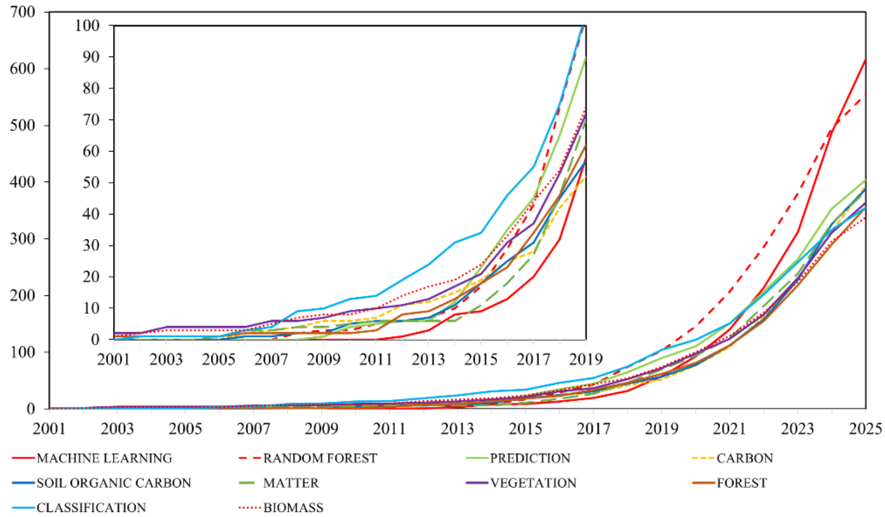

Figure 9. Word Frequency over time.

From an evolutionary trajectory perspective, the lifecycle of keywords and the shifting research focus can be further categorized into five typical patterns of evolution within the field of AI-enabled natural carbon sink. These categories—Emerging-Explosive, Rapidly Evolving, Mid-Term Active, and Early Declining (see Table 2)—are derived from the temporal distribution characteristics of keywords, including their initial appearance (Q1), median activity phase, and peak occurrence (Q3). This classification reflects the dynamic trajectory of technological adoption, encompassing stages from inception and rise to stabilization and eventual obsolescence. Emerging-Explosive keywords—such as ML, DL, and vegetation mapping—have surged since 2022, underscoring the rapid penetration of AI foundation models (e.g., Transformer, ChatGPT) into remote sensing interpretation, language modeling, and multimodal ecological data processing. This trend has been driven by improvements in hardware performance and spatial resolution, alongside the implementation of national carbon peaking and neutrality strategies, which necessitate higher precision in carbon estimation and scenario simulation. These terms exhibit rapid growth over short active periods, exemplifying a "latecomer advantage" trajectory. However, this trajectory also raises concerns about potential "model bubbles" in scenarios where robust ecological ground-truth data are insufficient. Rapidly Evolving keywords—such as RF, prediction, and SOC—typically emerged between 2018 and 2020 and peaked during 2022–2024, following a trajectory of "mid-term rise followed by sustained high frequency." These keywords are commonly linked to applications such as ensemble learning, carbon stock prediction, and soil property modeling. For example, RF has become a standard method for land cover classification and carbon stock modeling due to its robustness and interpretability (Bui et al., 2022). This keyword category represents a resilient "mainstream mid-phase model" trajectory, characterized by sustained applicability over time. Mid-Term Active keywords—such as forest biomass, carbon stocks, and boreal forest—entered mainstream usage after 2008 and have sustained mid-to-high levels of activity throughout the 2020s. These terms generally refer to key ecological scenarios for carbon sink estimation and have benefited from the widespread adoption of high-resolution remote sensing products such as LiDAR and Sentinel-2. While not characterized by explosive growth, their long-term stability reflects a structural "rigid demand" within ecological monitoring research. In contrast, Early Declining keywords—such as ANN, support vector regression (SVR), and radar backscatter—were predominantly active between 2008 and 2016. These keywords represent earlier technological efforts involving traditional AI methods and legacy remote sensing technologies, preceding the full-scale integration of AI into ecological modeling. Their sharp decline in popularity after 2020 suggests they have been largely supplanted by more powerful and adaptive models—such as deep neural networks and graph-based learning frameworks—highlighting the high sensitivity of ecological modeling to algorithmic innovation.

Overall, the current landscape of AI-enabled natural carbon sink research reflects a transition in keyword lifecycles from "tool-centric hotspots" to "system modeling." The research focus has gradually shifted from traditional ML toward multimodal integrated modeling and intelligent semantic understanding, while the research scope has expanded from forests to include soils, wetlands, and microbial mechanisms across multi-scale carbon sink systems. In the future, advancing systematic modeling and accurate estimation in this field will require strengthening the integration between ecological process knowledge and AI modeling logic, while remaining cautious of fragmentation and tool dependency risks associated with rapid technological turnover.

Table 2. Classification of Keyword Evolution Trends in AI+CCUS Research.

Evolution Type	Representative Keywords	Q1	Median	Q3	Trend Interpretation
Emerging-Explosive	ML, china, DL, vegetation mapping, microbial necromass, ch4	2022–2024	2024–2025	2024–2025	Rapid surge after 2022; accelerated integration of AI and natural carbon sinks
Rapidly Evolving	RF, prediction, carbon, SOC, classification	2019–2021	2022–2023	2024	Sharp rise in the mid period; became mainstream methods and indicators
Mid-Term Active	forest biomass, carbon stocks, boreal forest, variable selection	2018–2019	2020–2021	2023–2024	Gained traction around 2018; consistently maintained research popularity
Early Declining	ANN, biomass estimation, pedotransfer functions, radar backscatter, SVR, imaging spectroscopy, glas, water-vapor, tm data, small-footprint lidar, queensland, vegetation structure, jers-1 sar, discrete-return lidar, tree cover	2008–2017	2014–2020	2018–2022	Early-stage methods now experiencing declining attention or marginalization

4. Discussion

4.1. Task Alignment and Applicability Differences of AI Algorithms in Natural Carbon Sink Research

In the deep integration of remote sensing and carbon sink research, AI methods have been widely adopted to address the modeling challenges posed by high-dimensional data, strong nonlinearity, and pronounced cross-scale characteristics in complex ecosystems. These methods have not only achieved breakthroughs in algorithmic efficiency but also exhibit heterogeneous characteristics in terms of task alignment. Different types of AI models exhibit significant differentiation in their suitability for specific carbon sink tasks due to variations in their mathematical mechanisms, dependence on data structures, and levels of interpretability (Ogunyemi, 2023b, Orobinskaya et al., 2024, Xie and Wang, 2024). However, most existing studies focus primarily on model performance evaluation, lacking a systematic understanding of the coupling logic among AI methods, carbon sink tasks, and remote sensing data. Therefore, clarifying the adaptation mechanisms of various AI approaches in different carbon sink research scenarios not only enhances the precision of model selection but also provides a theoretical foundation for building future task-driven modeling systems.

Based on the clustering structure of the co-occurrence network (Figure 6) and the temporal evolution patterns of keywords (Table 2), this study identifies five distinct application paradigms of AI methods in natural carbon sink research: (1) traditional ML (e.g., RF); (2) DL techniques (e.g., DL); (3) classical regression models; (4) early-stage exploratory algorithms (e.g., SVM, ANN); and (5) ensemble modeling approaches. This classification not only reflects the developmental stages of different AI models within the natural carbon sink domain but also highlights their task-specific suitability, as detailed in Table 3. Traditional ML methods, exemplified by RF, have emerged as dominant tools in medium-scale modeling tasks such as soil carbon content estimation and land-use classification, owing to their strengths in feature selection, nonlinear regression, and generalization capability (Ho et al., 2024, Wang et al., 2022). DL methods, on the other hand, are more adept at handling high-dimensional unstructured data, demonstrating superior performance in tasks such as LiDAR point cloud inversion over forested areas, biomass estimation, and semantic segmentation of remote sensing imagery (Liu et al., 2024, Oehmcke et al., 2024). Although models such as SVM and ANN were valuable during the early exploratory phase, they have been increasingly supplanted by deep neural networks owing to their limitations in handling sample sensitivity and scalability. In recent years, hybrid approaches—exemplified by ensemble learning—have begun to demonstrate notable advantages in tackling complex tasks such as multi-source data fusion and error propagation control, highlighting their potential to achieve a balance between high stability and scalability (Hu and Cheng, 2023, Qian et al., 2024, Yewle et al., 2025). The differentiated evolution of these methods collectively shapes the current “adaptability landscape” of AI-assisted carbon sink modeling.

In the field of remote sensing modeling for natural carbon sinks, traditional ML methods—such as RF, SVM, and K-Nearest Neighbors—have been widely adopted for their strong generalization capabilities, low modeling threshold, and high result interpretability (Cheng et al., 2024, Song et al., 2024). As indicated by the keyword clustering results in Figure 6, terms such as RF, prediction, and SOC demonstrate both high frequency and centrality, underscoring the broad applicability of tree-based models—particularly RF—in ecological remote sensing tasks. This trend is further reflected in the keyword evolution graph (Figure 9), where RF has shown rapid growth since 2020 and reached its peak activity in 2024 (Table 2). This pattern highlights the model’s strong capability in handling multi-source heterogeneous remote sensing datasets (e.g., Landsat, MODIS, Sentinel) and its robustness in transferring across different spatial scales. Specifically, traditional ML models have proven effective in tasks such as soil carbon estimation, carbon stock classification, and land cover classification. For example, RF models effectively address the curse of dimensionality in high-dimensional datasets through iterative feature sampling and ensemble voting mechanisms, making them particularly suitable for processing noise-prone remote sensing reflectance data (Suleymanov et al., 2023, Wang, 2024, Zhou et al., 2025). Moreover, these models operate without the need for prior distributional assumptions or transformations, offering robustness to the nonlinearities and multicollinearity commonly encountered in ecological modeling tasks, thereby enhancing computational efficiency. Their inherent advantages in model interpretability—such as variable importance ranking and error attribution—make them particularly valuable for applications in policy development and ecological compensation frameworks. However, as the spatial resolution and temporal complexity of remote sensing data continue to rise, traditional models face growing limitations. Their reliance on expert-driven feature engineering hinders scalability under big data conditions, and their limited capacity to extract contextual semantics constrains their performance in dynamic carbon sink simulations and long-horizon forecasting. Thus, while traditional models remain widely used, their role is gradually transitioning from core modeling engines to auxiliary tools for lightweight evaluation and ensemble integration.

Compared with the lightweight adaptability of traditional ML methods, the recent rise of DL techniques—such as CNN, RNN, and Transformer architectures—in carbon sink remote sensing modeling illustrates an alternative adaptation pathway characterized by strong structural coupling (Günen, 2022, Maguluri et al., 2024, Zhao et al., 2021). An integrated analysis of Figure 9 and Table 2 reveals that keywords such as ML, DL, and learning have shown a marked surge in activity since

 doi:10.20944/preprints202509.0050.v1

2022. These terms can be classified as "Emerging-Explosive" and demonstrate the deepening penetration trend of post-DL technologies in the domain of ecological remote sensing. In particular, DL models show superior fitting capability and spatial resolution in addressing high-dimensional, nonlinear challenges such as image segmentation, feature extraction, and change detection—surpassing the performance of traditional algorithms (Chauhan et al., 2023, Mu et al., 2024). In practical applications, DL techniques are particularly well-suited for tasks involving high-resolution remote sensing imagery—such as land cover extraction, forest stand structure estimation, and fine-scale carbon stock estimation—based on platforms like Sentinel-2, Gaofen, and PlanetScope. CNNs, for example, enable direct construction of automated feature extraction pipelines from raw spectral or band data, significantly streamlining the traditionally labor-intensive feature engineering process (Amziane et al., 2023, Huang et al., 2022). Temporal models like LSTM and Transformer architectures can effectively capture seasonal and interannual dynamics in carbon sink behavior, improving both the temporal robustness and spatial generalizability of model predictions (Niu et al., 2024, Wang et al., 2024). As the integration of DL with multi-source heterogeneous data—including LiDAR, SAR, remote sensing imagery, and in-situ plots—continues to advance, end-to-end frameworks are emerging as a dominant modeling paradigm. These approaches offer promising pathways for multi-scale carbon flux estimation, ecological simulation, and policy modeling. Despite these advantages, DL models still face critical challenges in carbon sink modeling based on remote sensing. First, data scarcity and heterogeneous distribution remain significant challenges. Ecological remote sensing often suffers from sparse plot distribution and spatiotemporal discontinuity, resulting in insufficient or low-quality training data that can substantially compromise model stability (Lyu et al., 2024, Zhang et al., 2025). Second, the issue of model "black-box" is particularly pronounced: DL models often lack transparency in their reasoning processes, which hampers their applicability in policy formulation and scientific knowledge extraction—creating a tension with the high demand for interpretability in ecosystem research (Chen et al., 2023, Tasneem and Islam, 2023). Third, the computational resource threshold remains high: for large-scale regional modeling tasks, deep neural networks demand significantly greater computational power, storage capacity, and energy consumption compared to traditional models, making them difficult to deploy widely in typical research institutions or developing countries (Reiersen et al., 2022, Xi et al., 2023). Therefore, although DL technologies have experienced explosive growth in recent years, their broad application still depends on the coordinated development of supporting mechanisms such as data-sharing frameworks, model compression techniques, and interpretable modeling.

From a temporal evolution perspective, the advancement of AI technologies in carbon sink remote sensing modeling has not followed a linear trajectory; rather, it reflects the characteristics of a "technological substitution cycle" marked by staged experimentation and dynamic replacement. As shown in Table 2, keywords such as ANN, SVR, radar backscatter, and JERS-1 SAR are representative of early-declining techniques that entered the research landscape between 2008 and 2016 but declined rapidly around 2020—illustrating a typical cycle of early experimentation, brief prominence, and rapid obsolescence. The emergence of early AI approaches coincided with two key transitions: the expansion of remote sensing data from optical to radar/microwave sources, and a methodological shift from traditional statistical regression models to intelligent regression models. For instance, SVR gained widespread adoption around 2010 for carbon density estimation and forest structure analysis, largely due to its strong generalization capabilities under conditions of small sample sizes and high-dimensional data (Chen et al., 2010, Diao et al., 2011). However, with the increasing dimensionality of remote sensing data and the growing complexity of computational tasks, the limitations of SVR—particularly its sensitivity to kernel parameters and sample distribution—have become more evident, making it difficult to meet the current demands for high-frequency, multi-source data fusion. Although ANNs are capable of nonlinear modeling, their performance is constrained by shallow network architectures and limited parameter tunability, often leading to issues such as overfitting and unstable convergence (Mas and Flores, 2008). More notably, AI applications during this period were still primarily characterized by a "model-centric, scenariosupplemented" approach, lacking precise representation of the complex structures of ecosystems. For example, early microwave remote sensing variables—such as radar backscatter and small-footprint LiDAR—could supplement gaps in optical data, but their effectiveness was limited by factors such as penetration depth, terrain interference, and spatial consistency, making it difficult to establish stable frameworks for carbon estimation (Guo-qing, 2002, Takahashi et al., 2006, Tsui et al., 2013). The dual constraints of technological bottlenecks and algorithm-task mismatches have catalyzed the rapid emergence of deeper end-to-end DL architectures, signaling a paradigm shift in AI applications—from reliance on empirical parameter tuning to a focus on structure-driven learning. Therefore, the phasing out of early AI methods does not imply their ineffectiveness, but rather reflects a natural process of methodological transition and stage-specific mismatches in applicability. In future model ensemble research, reintegrating lightweight architectures and prior knowledge from earlier algorithms—such as Bayesian ANN or hybrid SVR models—is emerging as a promising pathway to enhance model interpretability and deployment flexibility (Sultana et al., 2022, Wang et al., 2024).

In summary, AI-enabled remote sensing research on natural carbon sinks has begun to take shape around a technological progression spanning traditional modeling, DL, and integrated intelligence. This evolution reflects a distinct shift from empirically driven methods to data-centric paradigms. The lifecycle patterns of various AI techniques (Table 2), along with their task alignment across different carbon sink scenarios (Table 3), converge on a central challenge: how to effectively align algorithmic capabilities with the characteristics of remote sensing data and the specific demands of ecological applications. This "model—data—task" triadic structure is increasingly becoming the central pivot driving the systematization and platformization of AI-based carbon sequestration research (Da Silva et al., 2022a, Wang et al., 2024).

Table 3. Matching Table of Different AI Algorithms and Natural Carbon Sink Tasks.

Algorithm Type	Typical Application Tasks	High-Frequency / Representative Keywords
RF	Soil carbon content estimation, land-use classification, feature importance identification	SOC, digital soil mapping, land-use, classification
DL	Semantic segmentation of remote sensing images, forest carbon stock inversion, LiDAR data processing	DL, aboveground biomass, lidar, vegetation, sentinel-2
Regression Models	Soil property modeling, carbon flux prediction, carbon stock trend fitting	regression, carbon stocks, prediction, carbon sequestration
SVM/ANN	Early exploration of remote sensing–carbon estimation, nonlinear modeling experiments	SVR, ANN, backscatter
Ensemble Modeling Methods	Multi-model ensemble optimization, error propagation control, multi-source data fusion modeling	ensemble learning, uncertainty, variability

4.2. Evolutionary Mechanisms of AI-Enabled Natural Carbon Sink Research

In the field of AI-enabled natural carbon sink research, technological evolution does not follow a linear trajectory; instead, it unfolds through phased breakthroughs and mechanism-level innovations, exhibiting a leapfrogging, cross-modal, and multi-scale integrated development pathway. From the perspective of research evolution, this study divides the field into four typical developmental stages based on the analysis in Section 3.1: the Emergence Stage (2001–2010), the Initial Growth Stage (2011–2017), the Acceleration Stage (2018–2021), and the Expansion Stage (2022–2025) (Figure 2). This temporal framework not only reflects the changing trends in publication output and citation frequency, but also shows a high degree of coupling in terms of methodological

 doi:10.20944/preprints202509.0050.v1

evolution and technological integration (see Table 4). To further clarify the technological trajectories of AI-enabled natural carbon sink research, this section analyzes the annual frequency evolution of keywords (Figure 9) and the lifecycle trajectories of keywords (Table 2).

Table 4. Overview of AI's Cross-Phase Evolution and Methodological Trends in the Field of CCUS.

Development Phase	Representative Methods/Algorithms	Core Research Topics	Dominant Interaction Mechanism	Tech-Problem Paradigm Characteristics
The Emergence Stage (2001–2010)	ANN, SVM, KNN, and other early statistical learning methods	Vegetation classification, preliminary estimation of soil carbon	Problem-Driven	Methods mostly served as auxiliary tools, relying heavily on expert ecological knowledge for modeling
The Initial Growth Stage (2011–2017)	RF、SVR、Ensemble Learning	Spatial interpolation of soil carbon, forest carbon measurement	Data-Driven	Enhanced coupling of remote sensing and field data; AI integrated into high-resolution mapping and modeling
The Acceleration Stage (2018–2021)	DL (e.g., CNN), high-dimensional feature learning	Carbon stock prediction using multi-source remote sensing, scenario simulation	Problem + Data	Models began replacing parts of expert-driven processes; AI embedded in mid-level layers of carbon sink modeling
The Expansion Stage (2022–2025)	Transformer, GPT, temporal prediction models	Multi-scale carbon flow modeling, zero-shot estimation, cross-domain transfer	Algorithm-Driven	AI transformed from a "tool" to a "cognitive agent," contributing to paradigm construction and theoretical abstraction

During the Emergence Stage (2001–2010), AI had not yet become a mainstream tool in natural carbon sink research; however, its early prototypes began to appear in forest ecosystem remote sensing and geostatistical analysis. Research during this phase primarily centered around keywords such as forest, vegetation, and biomass, with a focus on remote sensing inversion for vegetation cover change and carbon stock estimation. The AI methods employed were predominantly early-stage ML algorithms, such as SVM and ANN (e.g., SVM and ANN mostly emerged during Q1, around 2008–2010), which were introduced to address the limitations of traditional remote sensing interpretation and regression models in capturing nonlinear relationships. This period was marked by a strong reliance on moderate- to low-resolution remote sensing datasets (e.g., MODIS, Landsat TM), and due to constraints in computing resources and ecological field data availability, modeling efforts were largely confined to local-scale and static feature estimation tasks (Zhang and MA, 2009, Zhao et al., 2005). Overall, AI approaches during this stage functioned more as "technical experiments" rather than established research paradigms, lacking systematic model integration frameworks and cross-scale modeling logic. The research focus remained largely within the scope of conventional remote sensing applications and had yet to engage deeply with process-based modeling of ecosystem carbon fluxes or scenario simulations. The application of AI in natural carbon sink research was more reliant on knowledge transfer from the remote sensing domain, rather than being inherently driven by scientific questions specific to carbon sink processes.

During the Initial Growth Stage (2011–2017), the application of AI in natural carbon sink research transitioned from the early stage of "methodological experimentation" to a phase of "tool integration." The scope of study expanded beyond forest remote sensing to include soil carbon modeling and carbon stock regression, reflecting a shift toward topics with deeper ecological process logic. The results of keyword evolution indicate that during this stage, terms such as SOC, carbon stocks, and regression experienced a rapid rise in frequency (see Figure 9 and Table 2), marking a notable departure from the earlier remote sensing focus on vegetation classification and biomass estimation. This thematic reconfiguration reflects a shift in research focus from "characterizing forest structure" to "estimating carbon process variables," driven primarily by the institutional expansion of national demands for greenhouse gas accounting and carbon sink quantification (Romijn et al., 2012). On the methodological front, ensemble learning models began to play a dominant role. Ensemble algorithms such as RF and Gradient Boosting have been widely applied in SOC estimation and spatial **interpolation tasks** due to their capabilities in handling **high-dimensional nonlinear data** and assessing **variable importance** (Vašát et al., 2017, Were et al., 2015). At the same time, several studies began incorporating multisource data—including hyperspectral remote sensing, topographic indices, and climate variables—into carbon sink modeling frameworks, significantly improving the accuracy and generalizability of SOC distribution predictions. For instance, Jaber et al. (2011) and Peón et al. (2017) integrated Hyperion hyperspectral data with terrain and climatic variables to construct ensemble prediction models; Song et al. (2017) further combined Extreme Learning Machines with Kriging methods to achieve high-precision SOC estimation at regional scales. From the perspective of knowledge graph structures and keyword clustering, a relatively stable research subdomain began to emerge during this period—centered around "RF" as the core method and "soil carbon" as the core topic (see Figure 6 and Figure 8). This indicates that AI modeling tools were becoming increasingly embedded in natural carbon sink subsystems, particularly forming reproducible modeling pathways in contexts such as agricultural soils, arid ecosystems, and grassland systems. It is important to note, however, that no paradigm-shifting methodologies had yet emerged during this stage. The research remained focused on regression-based models and the interpretation of variable importance, with limited development of end-to-end modeling capabilities or intelligent frameworks capable of dynamic, time-series simulation.

The Acceleration Stage (2018–2021) marked a pivotal transition in the field of AI-enabled natural carbon sink, characterized by a shift from traditional ML to DL methodologies. During this period, both research topics and methodologies exhibited notable increases in complexity and diversification, including the deep fusion of multisource remote sensing data, expansion of modeling task dimensions, and the initial embedding of deep neural networks in simulating complex carbon processes. From a keyword evolution perspective, terms such as ML, prediction, and carbon sequestration frequently co-occurred (see Figure 6 and Figure 9), reflecting a transition in focus from static parameter estimation (e.g., SOC concentration) toward dynamic carbon flux prediction. The most notable breakthrough during this phase was the introduction of DL methods, particularly the widespread application of CNN in high-resolution remote sensing image recognition and land cover segmentation tasks (Pan et al., 2019, Pashaei et al., 2020). In carbon sink research, CNN have been widely applied to forest canopy height estimation and aboveground biomass inversion tasks, demonstrating superior predictive accuracy and spatial resolution compared to traditional models. For example, Shah et al. (2020) employed a CNN to model canopy height using Landsat imagery, achieving better performance than the traditional RF approach, thereby highlighting the CNN's advantage in spatial feature extraction (Shah et al., 2020). Simultaneously, recurrent neural networks such as LSTM began to be applied to time-series carbon flux prediction, effectively capturing long-term dependencies in climate and vegetation dynamics and showing potential for improving dynamic modeling accuracy (Besnard et al., 2019). A structural shift is also evident at the level of research themes. As shown in the Callon Two-Dimensional Distribution Map (Figure 7), keyword clusters related to ML and carbon stock exhibit both high centrality and high density, indicating a high degree of thematic cohesion. This suggests the emergence of a dual-core structure centered on "ML—carbon prediction." During this period, researchers increasingly emphasized model generalization and cross-regional adaptability. Techniques such as transfer learning and domain adaptation were gradually introduced to reconstruct and fine-tune existing models across different ecosystems (Wang et al., 2019, Xu et al., 2020), laying a theoretical and data foundation for the emergence of large-scale AI models in the Expansion Stage. It is worth noting that, while the adoption of deep models improved predictive accuracy, it also introduced new challenges—particularly in terms of model interpretability and the "black box" effect, which limited their acceptance in policy support and natural resource management. In response, researchers began integrating interpretability-enhancing techniques such as Shapley values and attention mechanisms, laying the groundwork for future advancements in model transparency and trustworthiness (Ethayarajh and Jurafsky, 2021, Kersten et al., 2021).

The Expansion Stage (2022–2025) marks a critical transition in AI-enabled natural carbon sink research—from viewing AI as a technical tool to recognizing it as a cognitive paradigm. As shown in Figure 2, both publication volume and average annual citations increased sharply after 2022. Concurrently, keywords such as "ML," "DL," "vegetation mapping," and "microbial necromass" surged in frequency (Figure 9), indicating AI's systemic involvement in model construction, knowledge abstraction, and ecological feedback interpretation. The defining characteristic of this period is the shift in AI's role from "fitting nature" to "explaining nature," reflecting a paradigmatic transition from task substitution to mechanism construction. A new generation of AI architectures, led by Transformer-based models, has gradually been introduced into remote sensing analysis and carbon-related prediction tasks. These models exhibit strong representation and generalization capabilities in applications such as high-resolution image interpretation, long-sequence modeling, and carbon price forecasting, offering new opportunities for simulating complex carbon cycles (Aleissaee et al., 2022, Mou et al., 2023). Since 2022, evaluation criteria in carbon sink research have shifted from traditional metrics—such as classification accuracy or RMSE—toward the ability to simulate and explain systemic processes, including carbon fluxes, climate feedbacks, and ecological disturbances. Researchers have increasingly adopted explainable AI (XAI) techniques, such as attention mechanisms, Shapley value analysis, and causal inference networks, to identify key drivers of changes in carbon sources and sinks, thereby improving model credibility and cross-regional generalizability (Cohrs et al., 2024, Díaz et al., 2022). In parallel, research objects have expanded beyond the "forest—soil" paradigm to include integrative modeling of multidimensional spatial structures—such as microbial carbon fluxes and subsurface carbon storage responses. Notably, the keyword microbial necromass emerged as a burst term for the first time (see Table 2), indicating a deepening of carbon sink cognition toward micro-scale ecological processes. Furthermore, LLM-based approaches for cross-regional carbon estimation have begun to supplement or replace traditional region-specific empirical methods, fostering the development of a systematic research framework that integrates data-driven modeling, model generalization, and process-based mechanisms (Cao et al., 2025, Cheng et al., 2024). With the significant expansion in model scale and application domains, the academic community has increasingly called for standardized datasets and open-platform ecosystems to mitigate risks of model opacity ("black-box" effects) and regional overfitting (Heroux et al., 2024). Against this backdrop, core techniques such as interpretability enhancement, process embedding, and generalization transfer have rapidly emerged, propelling AI-enabled carbon sink research from weak integration to strong interdisciplinarity and systematic development (Kabashkin and Susanin, 2024).

Tracing the development trajectory of the AI-enabled natural carbon sink field from 2001 to the present reveals a highly synchronized co-evolution between technological methodologies and research themes. This dynamic interaction has progressively crystallized into three representative types of coupling mechanisms: Problem-Driven Mechanism. In the early stage, research was oriented around concrete ecological tasks such as "biomass classification" and "soil carbon assessment," guiding the targeted adaptation and task-specific optimization of statistical learning tools such as SVM and RF (Xie et al., 2019). Data-Driven Mechanism. Particularly prominent during the Acceleration Stage, advancements in spatial resolution and multimodal characteristics of remote sensing datasets (e.g., Landsat, Sentinel, and LiDAR) significantly fueled the widespread application of DL models—such as CNNs and ResNet—for high-dimensional feature learning and complex nonlinear modeling (Reichstein et al., 2019). Algorithm-Driven Mechanism. In recent years, AI itself has evolved from a functional fitting tool into a "knowledge-generating entity" within cognitive modeling. This is exemplified by the zero-shot generalization capabilities of Transformer-based models in carbon scenario forecasting, indicating AI's increasing agency in actively reshaping research paradigms (Rasul et al., 2023). These mechanisms are not isolated; rather, they demonstrate a spiral evolutionary trajectory that progresses from problem-driven exploration to data-driven modeling, and ultimately to paradigm reconstruction. The phase division diagram (Figure 2), the semantic network and temporal frequency evolution of keywords (Figure 6 and Figure 9), and the lifecycle-based classification of keywords (Table 2) collectively reveal the co-evolutionary trajectory between AI technologies and thematic developments in carbon sink remote sensing research.

4.3. Application Challenges and Future Prospects

Despite the technological prosperity and data proliferation, the integration of AI into ecosystem carbon sink assessment still faces numerous practical challenges and theoretical bottlenecks. These issues not only hinder the scalability of AI models across ecological types and spatial scales but also highlight the deep disconnection among data foundations, algorithmic design, and decision-making practices. This section systematically examines the key obstacles and future directions of AI-enabled carbon sequestration research across five dimensions: data foundations, model transparency, mechanistic integration, scenario forecasting, and engineering translation.

Although AI methods such as RF and DL have demonstrated considerable predictive accuracy and application potential in subfields such as natural carbon sink estimation, carbon stock forecasting, and soil carbon modeling (see Figure 6), their methodological adaptability and generalization capacity remain limited. As revealed by the keyword clustering network in Figure 6, current AI approaches are highly concentrated around specific technical paradigms (e.g., "RF", "DL") and domain-specific applications (e.g., "SOC", "forest biomass"). This indicates a strong reliance on medium-resolution remote sensing data and existing field plot datasets for model development. While such methods often perform well within localized contexts, they tend to exhibit limited transferability and interpretability when applied across climatic zones, land-use types, or long-term forecasting scenarios (Liu et al., 2024, Wang et al., 2020). DL models may partially address regional heterogeneity through transfer learning, yet there remains a lack of validated pathways for their integration into carbon cycle simulation and mechanistic modeling (Wang et al., 2022). More critically, most current AI models are heavily dependent on high-quality, structured input variables, which struggle to capture the complex, nonlinear feedbacks and stochastic disturbances characteristic of ecosystems (Huang et al., 2021, Rastetter et al., 2023). This leads to a prevailing imbalance of "strong modeling capacity vs. weak process representation", thereby limiting the progression toward process-oriented carbon sink modeling and posing challenges to the scientific rigor of generalized carbon assessment frameworks. Therefore, future research must explicitly account for regional heterogeneity, input variable uncertainty, and cross-scale model adaptability during AI model development. A shift is needed from "task learning" to "mechanism learning", aiming to enhance both the ecological interpretability and broad applicability of AI-driven carbon sequestration research.

The modeling performance of AI in natural carbon sink research is highly contingent upon the quality of input data and the synergistic fusion of multisource datasets. However, current ecological data infrastructures still suffer from significant limitations in spatial-temporal resolution, coverage accuracy, and inter-source consistency (see Figure 3). On one hand, high-resolution remote sensing imagery (e.g., Sentinel-2, Landsat 8) provides fine-grained surface information, but issues such as cloud contamination, temporal discontinuities, and inconsistent radiometric calibration hinder the stable acquisition of global-scale datasets (Belgiu and Drăguţ, 2016, Zhu et al., 2015). On the other hand, field-based plot data—such as SOC concentrations or carbon stock measurements—often suffer from small sampling scales, uneven spatial distributions, and non-standardized recording protocols. These issues are particularly acute in countries of the Global South, significantly undermining the stability and generalizability of AI model training (Dharumarajan et al., 2021, Ugbemuna Ugbaje et al., 2024). More critically, the temporal mismatch across data sources frequently leads to distorted modeling results (Liddicoat et al., 2015). For instance, satellite imagery is typically collected at monthly or quarterly intervals, while soil monitoring data are predominantly annual, rendering temporal alignment infeasible. Similarly, carbon flux observations (e.g., FLUXNET) are constrained by high deployment costs and limited coverage, making it difficult to represent intra-regional heterogeneity (Fang et al., 2024, Kia, 2017). This misalignment of temporal resolution poses a major barrier to accurate modeling of carbon process dynamics using AI. Additionally, the heterogeneity and complexity of ecological variables further exacerbate the challenges of data fusion. These include structural variables (e.g., vegetation indices, topographic factors), process-oriented variables (e.g., temperature, precipitation), and management-driven variables (e.g., land use, fertilization behavior). The semantic ambiguity, high missingness, and non-standard formats among these variable types significantly increase the integration difficulty (Fang et al., 2024, Leitão et al., 2018, Milodowski et al., 2023). Recent studies have explored the integration of remote sensing and ecological data using CNNs and multi-input neural network architectures. While these approaches demonstrate strengths in feature extraction and multi-source data fusion, their performance remains highly sensitive to the quality of input variables and is constrained by the absence of robust adaptive mechanisms capable of dynamically identifying and filtering anomalous or low-quality data (Li et al., 2020, Park et al., 2022, Ye et al., 2019). This limitation not only hinders further improvements in model accuracy but also exacerbates the disparity in AI applicability between data-rich and data-scarce regions. To address this challenge, a critical next step involves the development of an open, unified, and standardized global carbon dataset platform, underpinned by a cross-domain, multi-layered data architecture that encompasses remote sensing, meteorological, biogeochemical, and policy-related variables. Such infrastructure is essential for ensuring the long-term sustainability and scalability of AI applications in carbon sequestration research.

Although AI models have demonstrated unprecedented predictive accuracy and advanced feature extraction capabilities in natural carbon sink modeling, their “black box” nature remains a significant barrier, limiting the interpretation of ecological processes and the application of research to policy. Deep neural network models, such as CNN, LSTM, and Transformer, are widely applied in carbon modeling and remote sensing estimation tasks. However, most of these frameworks focus primarily on pattern recognition and optimizing predictive performance, without systematically embedding the logic of ecological process-based mechanism modeling. As a result, their ability to provide mechanistic explanations remains limited (Li et al., 2024). Additionally, they often lack frameworks for causal inference and visualization of ecological processes, restricting their interpretability and reliability in complex systems such as “carbon sink formation—driving factors—response mechanisms” (Terziyan and Vitko, 2023, Whata and Chimedza, 2022). As seen in the two-dimensional keyword distribution map (Figure 7), while high-centrality terms such as ML and RF are located at the semantic core, there is still a misalignment with process-oriented keywords like carbon dynamics, uncertainty, and SOC. This suggests that current research has yet to achieve an effective coupling between “process modeling” and “result prediction.” This disconnect reflects, on one hand, the insufficient integration of ecological knowledge within AI frameworks, and on the other hand, highlights AI's limitations in addressing uncertainty propagation and interpretation. Recent developments in XAI offer new directions for ecological modeling, including methods such as integrated attribution (e.g., SHAP, LIME), feature importance ranking, and neuron activation visualization (Tempel et al., 2024, Zodage et al., 2024). Some studies have also explored integrating physical priors or ecological rules (such as carbon balance constraints or species distribution logic) into DL architectures to improve ecological plausibility (Diligenti et al., 2017, Roychowdhury et al., 2021). However, this "knowledge-driven AI" approach remains in the exploratory phase and lacks a unified framework or broad consensus. To improve the transparency of AI models in natural carbon sink research, it is crucial to reconstruct these models at the mechanistic level and develop a comprehensive "process—mechanism—prediction" modeling system. This would facilitate the shift from correlation-based prediction to causal mechanism explanation, enhancing the ecological credibility and policy relevance of AI in carbon process simulation.

The application of AI technologies in global carbon sequestration research reveals a striking spatial imbalance, marked by a clear dominance of the Global North and marginal participation from the Global South. As indicated by national publication statistics in Table 1, China and the United States lead the field with 1,772 and 487 publications, respectively—together accounting for nearly 60% of global output. In contrast, regions such as Africa, South America, and Southeast Asia contribute less than 10% collectively, with key ecological zones—like the Congo Basin and the Amazon Rainforest—largely overlooked in existing research. This stark disparity reflects not only the lack of foundational data collection infrastructure in these regions but also broader inequities in scientific resources, remote sensing infrastructure, and algorithmic capacity at the global scale. Given the heavy dependence of AI-driven carbon research on satellite data (e.g., Sentinel, MODIS), ground plot surveys, and long-term ecological monitoring networks (e.g., FLUXNET, LTER), deploying and training robust models in resource-limited contexts remains a major challenge. Moreover, pronounced biophysical heterogeneity, substantial data deficiencies, and frequent cloud cover in these regions have collectively intensified model generalization errors. For instance, due to the absence of critical ecological samples—such as tropical lateritic soils and swamp wetlands—current soil carbon estimation models face significant limitations in delivering equitable and accurate carbon sink assessments at the global level (Huang et al., 2022, Raihan, 2024). In addition, prevailing AI modeling frameworks are typically constructed based on unified training datasets and global parameter optimization strategies, often overlooking regional variations in ecological processes. This shortcoming undermines model generalization in heterogeneous ecosystems and further exacerbates existing inequities (Da Silva et al., 2022b, 2022c). To address this "carbon data divide," the international community must urgently foster collaborative data-sharing mechanisms and invest in local capacity-building. The development of a "Global Carbon Commons"—built on open ecological monitoring platforms, Federated Learning, and edge AI deployment—has emerged as a new consensus in United Nations climate mechanisms and AI ethics frameworks (UNESCO, 2023). Achieving research equity is essential for advancing AI-enabled carbon sink assessment from broad global estimations to context-specific, localized actions—thereby meaningfully contributing to the ecological governance and climate mitigation goals of countries in the Global South.

As AI-enabled carbon sequestration research enters a phase of accelerated development and paradigm reconstruction, there is an urgent need to transition from single-algorithm modeling to a composite paradigm that integrates multi-source data, multi-modal AI, and process-based ecological mechanisms (see **Table 5**). First, at the data level, substantial spatiotemporal heterogeneity exists across remote sensing, meteorological, topographic, soil, and socioeconomic variables. This uneven distribution of data has emerged as a key bottleneck constraining both the generalizability and performance enhancement of ecological models (Liu et al., 2025, Wang et al., 2023). Future efforts must focus on building a multi-source heterogeneous integration platform, combining satellite observations (e.g., Sentinel, Landsat), in-situ monitoring networks (e.g., FLUXNET, SoilGrids), and IoT-based sensing systems. Such integration would enable cross-scale data complementarity and enhance both modeling accuracy and application scalability. Second, regarding modeling paradigms, Transformer architectures, multimodal DL approaches (such as ViT-BERT fusion models), and self-supervised learning strategies are gradually replacing traditional supervised learning models, demonstrating stronger generalization capabilities in ecological scenarios characterized by severe data gaps and sample biases (Caruso et al., 2024, Wang et al., 2022). For example, in recent years, some studies have attempted to incorporate DL and ML models (such as ANN and LightGBM) into forest carbon stock estimation, achieving better predictive performance than traditional methods in tropical regions like the Amazon. However, systematic evaluations of self-supervised time series models remain limited (Dantas et al., 2021, Nguyen and Saha, 2024, Zheng et al., 2025). Meanwhile, the shift from "prediction to explanation" has become a key trend in AI ecological modeling. XAI frameworks are gradually incorporating mechanisms such as SHAP and LIME to help understand the driving processes of carbon sinks and reduce the policy risks associated with black-box models (Bhavana et al., 2024, Sen et al., 2025). Finally, model functions are expanding from static estimation to scenario forecasting and intervention design. Based on a multi-scenario driven "Prediction—Explanation—Intervention" modeling framework, the integration of AI with land use change simulation is progressively providing quantitative support for carbon neutrality policies. For example, multiple studies have shown that moderately increasing the proportion of mixed forests and optimizing forest structure can enhance carbon sink potential and ecological resilience without significantly reducing output (Tian et al., 2024, Zhang et al., 2024). This implies that future AI-enabled carbon sequestration research will no longer be limited to "understanding nature" but will increasingly move towards the practical orientation of "managing nature".

Table 5. Key Challenges and Corresponding Solutions of AI Applications in CCUS.

Research Pain Point	Current Technical Limitations	Future Research Trends	Key Technologies and Methods	Potential Application Value
Data Heterogeneity and Missing Data Issues	Difficult integration of highly heterogeneous spatiotemporal remote sensing and field data; lack of high-quality training datasets	Building an integrated system combining remote sensing, ground-based, and IoT data	Multi-source data fusion, spatiotemporal interpolation, self-supervised learning	Improve model accuracy, generalizability, and regional adaptability
Model Opacity and Lack of Interpretability	Traditional DL "black box" fails to reveal underlying mechanisms	Advancing XAI and causal learning modeling frameworks	SHAP, LIME, causal graphs, feature attribution	Enhance result credibility and support policy formulation
Lack of Process-Based Mechanistic Drivers	Purely data-driven models overlook ecological-climatic process mechanisms	Model integration: Hybrid paradigm combining physical models and AI	Hybrid models, ecological mechanism embedding	Deepen understanding of natural system structure and evolutionary mechanisms
Static Estimation Lacks Predictive Power	Focus on current-state estimation, unable to address future scenario changes	"Prediction—Explanation—Intervention" three-stage modeling framework	Multi-scenario simulation, GNN, reinforcement learning	Support carbon neutrality scenario modeling and decision optimization
Lack of Application-Oriented Transformation	Algorithm engineering disconnected from management,	Promoting a "Technology–Policy–Practice" integration mechanism	Decision support systems, digital twin ecological platforms	Build a "measurable, manageable, and controllable" carbon sink

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2025 doi:10.20944/preprints202509.0050.v1

industry, and governance practices	management system

5. Conclusions

This study systematically reviews the evolutionary trajectory of AI applications in natural carbon sink research. Based on 3,885 relevant publications from 2000 to 2025 and analyzing publication volume and citation trends, the development of AI-enabled natural carbon sink is divided into four typical stages: the Emergence Stage (2001–2010), the Initial Growth Stage (2011–2017), the Acceleration Stage (2018–2021), and the Expansion Stage (2022–2025). Through co-word network analysis, keyword clustering, temporal evolution modeling, and algorithm-topic matrix correlation, this study constructs a relatively clear knowledge map and evolutionary structure of AI-enabled carbon sink. The results indicate that AI models represented by ensemble learning methods such as RF are rapidly penetrating core carbon source-sink modeling scenarios, including aboveground biomass, SOC, and forest carbon stocks, marking a shift in the field from empirical statistics toward a data-driven ecological prediction paradigm.

In terms of methodology, this study integrates bibliometric visualization with conceptual modeling and proposes a multi-dimensional analytical framework—including the classification of research stages (Figure 2), the categorization of keyword evolution types (Table 2), and the Callon two-dimensional map (Figure 7)—thereby expanding the methodological toolbox for AI-enabled natural carbon sink research. The construction of a technology–application matrix (Table 3) also helps clarify the applicability boundaries and research focus of different AI algorithms across various carbon sink scenarios. In addition, the evolution trend chart (Figure 9) and keyword evolution typology reveal the life cycles of specific terms, technological substitution paths, and the progression of stage-specific research hotspots. At the application level, this study highlights the significant value of AI in enhancing carbon sink estimation accuracy, supporting scenario simulation, and informing governance decisions. Especially after 2020, the accelerated application of ML methods has been driven not only by advances in computing power and data resolution but also by policy incentives such as the “dual carbon” goals (e.g., China’s carbon peaking and carbon neutrality targets, and the EU’s Green Deal). With its capabilities in multi-source heterogeneous data integration, scale transformation, and probabilistic inference, AI is providing methodological support for forest management, climate-smart agriculture, and ecosystem service valuation—gradually reshaping traditional carbon assessment systems.

This study also has certain limitations. First, the bibliometric data primarily derived from the WoS Core Collection may underrepresent regional studies and gray literature. Second, the reliance on co-occurrence frequency and keyword strength as primary analytical metrics may fall short in capturing the deeper semantic structures inherent in interdisciplinary discourse. Future research could further incorporate approaches such as knowledge graph modeling, multilingual natural language processing, and contextual semantic analysis to enhance the depiction of research mechanisms and discursive evolution. Moreover, cross-modal modeling that integrates AI algorithm evaluation with ecological process models—such as biogeochemical simulations—remains an important avenue for in-depth exploration.

Overall, this study establishes a comprehensive analytical framework encompassing AI methodologies, ecological mechanisms, and knowledge structures, offering valuable insights for researchers to trace the technological trajectories and future directions of AI-enabled natural carbon sink research. It provides both theoretical and methodological support for developing transparent, interpretable, and standardized carbon modeling systems. In the context of escalating climate change pressures, AI serves not only as a tool for optimization, but also as a key driver of innovation in carbon governance and methodological transformation.

Funding: This research was financially supported by the National Key Research and Development Program of China (Grant No. 2022YFF0801201), the National Natural Science Foundation of China (Grant No. U1911202), and the Key-Area Research and Development Program of Guangdong Province (Grant No. 2020B1111370001).

Data Availability Statement: Geoma183 (2025). *AI-NatureCarbon-Biblio: A bibliometric dataset on AI in natural carbon sequestration*. GitHub. https://github.com/Geoma183/AI-NatureCarbon-Biblio

Acknowledgments: We sincerely appreciate the valuable feedback and insightful suggestions provided by the reviewers, which have significantly improved the quality of this manuscript. We also extend our gratitude to the editorial team for their careful handling of our submission.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Global CCS Institute (2023) Global status of CCS 2023. Melbourne, Global CCS Institute.
2. Aleissaee, A. A., Kumar, A., Anwer, R. M., Khan, S. H., Cholakkal, H., Xia, G. & Khan, F. S. (2022) Transformers in remote sensing: a survey. *Remote. Sens.*, 15, 1860.
3. Alongi, D. M. (2023) Current status and emerging perspectives of coastal blue carbon ecosystems. *Carbon Footprints*, 2, N/A-N/A.
4. Amziane, A., Losson, O., Mathon, B. & Macaire, L. (2023) MSFA-net: a convolutional neural network based on multispectral filter arrays for texture feature extraction. *Pattern Recognition Letters*, 168, 93-99.
5. Aria, M. & Cuccurullo, C. (2017) Bibliometrix: an r-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11, 959-975.
6. Bartell, S. M. (2024) *Biophysical economy: theory, challenges, and sustainability*, CRC Press.
7. Belgiu, M. & Drăguţ, L. (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114, 24-31.
8. Besnard, S., Carvalhais, N., Arain, M. A., Black, A. T., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. P. W., Dutrieux, L. P., Gans, F., Herold, M., Jung, M., Kosugi, Y., Knohl, A., Law, B. E., Paul Limoges, E., Lohila, A., Merbold, L., Roupsard, O., Valentini, R., Wolf, S., Zhang, X. & Reichstein, M. (2019) Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests. *PLoS ONE*, 14.
9. Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. W., Dutrieux, L. P. & Gans, F. (2019) Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests. *PloS one*, 14, e0211510.
10. Bhavana, B., Raju, S. P., Kodipalli, A. & Rao, T. (2024) Carbon dioxide emission by cars: a performance analysis of diverse machine learning regressors and interpretation using explainable AI. *2024 3rd International Conference for Advancement in Technology (ICONAT)*, 1-6.
11. Bracarense, N., Bawack, R. E., Fosso Wamba, S. & Carillo, K. D. A. (2022) Artificial intelligence and sustainability: a bibliometric analysis and future research directions. *Pacific Asia Journal of the Association for Information Systems*, 14, 9.
12. Bricklemyer, R. S., Lawrence, R. L., Miller, P. R. & Battogtokh, N. (2006) Predicting tillage practices and agricultural soil disturbance in north central montana with landsat imagery. *Agriculture, Ecosystems & Environment*, 114, 210-216.
13. Canadell, J. G. & Raupach, M. R. (2008) Managing forests for climate change mitigation. *science*, 320, 1456-1457.
14. Cao, Z., Han, M., Wang, J. & Jia, M. (2025) CarbonChat: large language model-based corporate carbon emission analysis and climate knowledge q&a system. *arXiv preprint arXiv:2501.02031*.
15. Caruso, C. M., Soda, P. & Guarrasi, V. (2024) MARIA: a multimodal transformer model for incomplete healthcare data. *arXiv preprint arXiv:2412.14810*.
16. Change, I. P. O. C. (2014) AR5 climate change 2014: mitigation of climate change. *Geneva, Switzerland*.
17. Chauhan, K., Tomar, H., Kamal, K. & Goel, P. (2023) Feature extraction from image sensing (remote): image segmentation *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE.

 doi:10.20944/preprints202509.0050.v1

18. Chen, C. (2016) *CiteSpace: a practical guide for mapping scientific literature*, Nova Science Publishers Hauppauge, NY, USA.

19. Chen, G., Hay, G. J. & Zhou, Y. (2010) Estimation of forest height, biomass and volume using support vector regression and segmentation from lidar transects and quickbird imagery. *2010 18th International Conference on Geoinformatics*, 1-4.

20. Chen, Y., Fan, K., Chang, Y. & Moriyama, T. (2023) Special issue review: artificial intelligence and machine learning applications in remote sensing, MDPI.

21. Cheng, F., Ou, G., Wang, M. & Liu, C. (2024) Remote sensing estimation of forest carbon stock based on machine learning algorithms. *Forests*, 15, 681.

22. Cheng, Y., Zhou, X., Zhao, H., Gu, J., Wang, X. & Zhao, J. (2024) Large language model for low-carbon energy transition: roles and challenges *2024 4th Power System and Green Energy Conference (PSGEC)*, IEEE.

23. Cohrs, K., Varando, G., Carvalhais, N., Reichstein, M. & Camps-Valls, G. (2024) Causal hybrid modeling with double machine learning—applications in carbon flux modeling. *Machine Learning: Science and Technology*, 5.

24. Da Silva, A. F., Nathaniel, J., Wong, K. C., Watson, C., Wang, H., Singh, J., Chamon, A. A. & Klein, L. (2022a) Netzeroco 2, an ai framework for accelerated nature-based carbon sequestration *2022 IEEE International Conference on Big Data (Big Data)*, IEEE.

25. Da Silva, A. F., Nathaniel, J., Wong, K. C., Watson, C., Wang, H., Singh, J., Chamon, A. A. & Klein, L. (2022b) Netzeroco 2, an ai framework for accelerated nature-based carbon sequestration *2022 IEEE International Conference on Big Data (Big Data)*, IEEE.

26. Da Silva, A. F., Nathaniel, J., Wong, K. C., Watson, C., Wang, H., Singh, J., Chamon, A. A. & Klein, L. (2022c) Netzeroco 2, an ai framework for accelerated nature-based carbon sequestration *2022 IEEE International Conference on Big Data (Big Data)*, IEEE.

27. Dąbrowska-Zielińska, K., Budzynska, M., Tomaszewska, M. A., Malińska, A., Gatkowska, M., Bartold, M. & Malek, I. (2016) Assessment of carbon flux and soil moisture in wetlands applying sentinel-1 data. *Remote. Sens.*, 8, 756.

28. Dai, F., Kahrl, F., Gordon, J. A., Perron, J., Chen, Z., Liu, Z., Yu, Y., Zhu, B., Xie, Y. & Yuan, Y. (2023) US-china coordination on carbon neutrality: an analytical framework. *Climate Policy*, 23, 929-943.

29. Dantas, D., Terra, M. D. C. N., Schorr, L. P. B. & Calegario, N. (2021) Machine learning for carbon stock prediction in a tropical forest in southeastern brazil. *Revista Bosque*, 42, 131-140.

30. Delbeke, J., Runge-Metzger, A., Slingenberg, Y. & Werksman, J. (2019) The paris agreement *Towards a climate-neutral Europe*, Routledge.

31. Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Vasundhara, R., Kumar, K. S. A., Nair, K. M., Hegde, R., Singh, S. K. & Lagacherie, P. (2021) Digital soil mapping of soil organic carbon stocks in western ghats, south india. *Geoderma Regional*, 25.

32. Diao, Y., Zhang, C., Liu, J., Liang, Y., Hou, X. & Gong, X. (2011) Optimization model to estimate mount tai forest biomass based on remote sensing *Conference on Control Technology and Applications*.

33. Díaz, E., Adsuara, J. E., Martínez, Á. M., Piles, M. & Camps-Valls, G. (2022) Inferring causal relations from observational long-term carbon and water fluxes records. *Scientific Reports*, 12.

34. Diligenti, M., Roychowdhury, S. & Gori, M. (2017) Integrating prior knowledge into deep learning. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 920-923.

35. Doney, S. C., Fabry, V. J., Feely, R. A. & Kleypas, J. A. (2009) Ocean acidification: the other CO2 problem. *Annual Review of Marine Science*, 1, 169-192.

36. Ethayarajh, K. & Jurafsky, D. (2021) Attention flows are shapley value explanations *Annual Meeting of the Association for Computational Linguistics*.

37. Fang, J., Bowman, K., Zhao, W., Lian, X. & Gentine, P. (2024) Differentiable land model reveals global environmental controls on ecological parameters.

38. Fang, J., Fang, J., Chen, B., Zhang, H., Dilawar, A., Guo, M. & Liu, S. A. (2024) Assessing spatial representativeness of global flux tower eddy-covariance measurements using data from FLUXNET2015. *Scientific Data*, 11, 569.

39. Fetting, C. (2020) The european green deal. *ESDN Report, December*, 2, 53.

40. Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T. & Olsen, A. (2024) Global carbon budget 2024. *Earth System Science Data Discussions*, 2024, 1-133.

41. Fu, Z., Chen, F., Zhou, S., Li, H. & Jiang, L. (2024) LLMCO2: advancing accurate carbon footprint prediction for LLM inferences. *arXiv preprint arXiv:2410.02950*.

42. Fuss, S., Canadell, J. G., Peters, G. P., Tavoni, M., Andrew, R. M., Ciais, P., Jackson, R. B., Jones, C. D., Kraxner, F., Nakicenovic, N., Le Quéré, C., Raupach, M. R., Sharifi, A., Smith, P. & Yamagata, Y. (2014) Betting on negative emissions. *Nature Climate Change*, 4, 850-853.

43. Green, J. & Keenan, T. F. (2022) The limits of forest carbon sequestration. *Science*, 376, 692 - 693.

44. Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V. & Smith, P. (2017) Natural climate solutions. *Proceedings of the National Academy of Sciences*, 114, 11645-11650.

45. Günen, M. A. (2022) Performance comparison of deep learning and machine learning methods in determining wetland water areas using EuroSAT dataset. *Environmental Science and Pollution Research*, 29, 21092-21106.

46. Guo-Qing, S. B. X. (2002) Radar backscatter model and its application in terrain-effect reduction of radar imagery. *National Remote Sensing Bulletin*.

47. Halkos, G. E. & Aslanidis, P. C. (2024) Green energy pathways towards carbon neutrality. *Environmental and Resource Economics*, 87, 1473-1496.

48. Han, P., Yao, B., Cai, Q., Chen, H., Sun, W., Liang, M., Zhang, X., Zhao, M., Martin, C. & Liu, Z. (2024) Support carbon neutral goal with a high-resolution carbon monitoring system in beijing. *Bulletin of the American Meteorological Society*, 105, E2461-E2481.

49. He, J., Li, Z., Zhang, X., Wang, H., Dong, W., Du, E., Chang, S., Ou, X., Guo, S. & Tian, Z. (2022) Towards carbon neutrality: a study on china's long-term low-carbon transition pathways and strategies. *Environmental Science and Ecotechnology*, 9, 100134.

50. Heroux, M. A., Shende, S., Mcinnes, L. C., Gamblin, T. & Willenbring, J. M. (2024) Toward a cohesive AI and simulation software ecosystem for scientific innovation. *ArXiv*, abs/2411.09507.

51. Ho, V. H., Morita, H., Bachofer, F. & Ho, T. H. (2024) Random forest regression kriging modeling for soil organic carbon density estimation using multi-source environmental data in central vietnamese forests. *Modeling Earth Systems and Environment*, 1-22.

52. Hu, B. & Cheng, Y. (2023) Predicting regional carbon price in china based on multi-factor HKELM by combining secondary decomposition and ensemble learning. *Plos one*, 18, e0285311.

53. Huang, H., Luo, L. & Pu, C. (2022) Self-supervised convolutional neural network via spectral attention module for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 September 2025 doi:10.20944/preprints202509.0050.v1

54. Huang, H., Yang, L., Zhang, L., Pu, Y., Yang, C., Wu, Q., Cai, Y., Shen, F. & Zhou, C. (2022) A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect. *Environmental Research Letters*, 17, 123004.

55. Huang, Y., Song, X., Wang, Y. P., Canadell, J. G., Luo, Y., Ciais, P., Chen, A., Hong, S., Wang, Y., Tao, F., Li, W., Xu, Y., Mirzaeitalarposhti, R., Elbasiouny, H., Savin, I., Shchepashchenko, D., Rossel, R., Goll, D. S., Chang, J., Houlton, B. Z., Wu, H., Yang, F., Feng, X., Chen, Y., Liu, Y., Niu, S. & Zhang, G. L. (2024) Size, distribution, and vulnerability of the global soil inorganic carbon. *Science*, 384, 233-239.

56. Huang, Y., Wang, Y. P. & Ziehn, T. (2021) Nonlinear interactions of land carbon cycle feedbacks in earth system models. *Global Change Biology*, 28, 296 - 306.

57. Intergovernmental Panel On Climate Change (2022) *Global warming of 1.5°c: IPCC special report on impacts of global warming of 1.5°c above pre-industrial levels in context of strengthening response to climate change, sustainable development, and efforts to eradicate poverty*, Cambridge, Cambridge University Press.

58. Jaber, S. M., Lant, C. & Al-Qinna, M. I. (2011) Estimating spatial variations in soil organic carbon using satellite hyperspectral data and map algebra. *International Journal of Remote Sensing*, 32, 5077 - 5103.

59. Justice, C. O., Townshend, J., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P. & Morisette, J. T. (2002) An overview of MODIS land data processing and product status. *Remote sensing of Environment*, 83, 3-15.

60. Kabashkin, I. & Susanin, V. (2024) Unified ecosystem for data sharing and AI-driven predictive maintenance in aviation. *Computers*, 13, 318.

61. Kersten, T., Wong, H. M., Jumelet, J. & Hupkes, D. (2021) Attention vs non-attention for a shapley-based explanation method. *arXiv preprint arXiv:2104.12424*.

62. Khan, M. S., Jeon, S. B. & Jeong, M. (2021) Gap-filling eddy covariance latent heat flux: inter-comparison of four machine learning model predictions and uncertainties in forest ecosystem. *Remote Sensing*, 13, 4976.

63. Kia, S. (2017) Uncertainty associated with scaling spectral indices of carbon fluxes at various spatial and temporal scales, University of Southampton.

64. Kim, S. T., Conklin, S. D., Redan, B. W. & Ho, K. K. H. Y. (2024) Determination of the nutrient and toxic element content of wild-collected and cultivated seaweeds from hawai'i. *ACS Food Science & Technology*, 4, 595-605.

65. Kumar, P., Sajjad, H., Tripathy, B. R., Ahmed, R. U. & Mandal, V. P. (2017) Prediction of spatial soil organic carbon distribution using sentinel-2a and field inventory data in sariska tiger reserve. *Natural Hazards*, 90, 693-704.

66. La Notte, A. (2024) How to account for nature-based solutions as the ecological assets that support economy and society. *Nature-Based Solutions*, 6, 100164.

67. Lal, R. (2004) Soil carbon sequestration to mitigate climate change. *Geoderma*, 123, 1-22.

68. Leitão, P. J., Schwieder, M., Pötzschner, F., Pinto, J. R. R., Teixeira, A. M. C., Pedroni, F., Sanchez, M., Rogass, C., van der Linden, S., Bustamante, M. M. C. & Hostert, P. (2018) From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. *Ecosphere*.

69. Li, S., Wang, H., Song, L., Wang, P., Cui, L. & Lin, T. (2020) An adaptive data fusion strategy for fault diagnosis based on the convolutional neural network. *Measurement*, 165, 108122.

70. Li, Y., Li, M., Li, C. & Liu, Z. (2020) Forest aboveground biomass estimation using landsat 8 and sentinel-1a data with machine learning algorithms. *Scientific reports*, 10, 9952.

71. Li, Y., Wang, X. & Guo, Y. (2024) CNN-trans-SPP: a small transformer with CNN for stock price prediction. *Electronic Research Archive*.

72. Liddicoat, C., Maschmedt, D., Clifford, D., Searle, R. D., Herrmann, T. N., Macdonald, L. M. & Baldock, J. A. (2015) Predictive mapping of soil organic carbon stocks in south australia's agricultural zone. *Soil Research*, 53, 956-973.

73. Liu, G., Migliavacca, M., Reimers, C., Kraft, B., Reichstein, M., Richardson, A. D., Wingate, L., Delpierre, N., Yang, H. & Winkler, A. J. (2024) DeepPhenoMem v1.0: deep learning modelling of canopy greenness dynamics accounting for multi-variate meteorological memory effects on vegetation phenology. *Geoscientific Model Development*.

74. Liu, H., Mou, C., Yuan, J., Chen, Z., Zhong, L. & Cui, X. (2024) Estimating urban forests biomass with LiDAR by using deep learning foundation models. *Remote Sensing*, 16, 1643.

75. Liu, L., Zhou, Y., Ma, J., Zhang, Y. & He, L. (2025) Domain-specific question-answering systems: a case study of a carbon neutrality knowledge base. *Sustainability*, 17, 2192.

76. Liu, M., Lei, H., Wang, X. & Paredes, P. (2025) High-resolution mapping of evapotranspiration over heterogeneous cropland affected by soil salinity. *Agricultural Water Management*.

77. Lyu, G., Wang, X., Huang, X., Xu, J., Li, S., Cui, G. & Huang, H. (2024) Toward a more robust estimation of forest biomass carbon stock and carbon sink in mountainous region: a case study in tibet, china. *Remote. Sens.*, 16, 1481.

78. Ma, Q., Wang, X., Mo, B., Liu, Z., Zhang, Y., Zong, W. & Bai, M. (2025) Knowledge map-based analysis of carbon sequestration research dynamics in forest and grass systems: a bibliometric analysis. *Atmosphere*, 16, 474.

79. Maguluri, L. P., Geetha, B., Banerjee, S., Srivastava, S. S., Nageswaran, A., Mudalkar, P. K. & Raj, G. B. (2024) Sustainable agriculture and climate change: a deep learning approach to remote sensing for food security monitoring. *Remote Sensing in Earth Systems Sciences*, 1-13.

80. Mas, J. F. & Flores, J. J. (2008) The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, 29, 617-663.

81. Masria, A. & Abouelsaad, O. (2025) Artificial intelligence applications in coastal engineering and its challenges–a review. *Continental Shelf Research*, 105425.

82. Mei, L., Rozanov, V., Jiao, Z. & Burrows, J. P. (2022) A new snow bidirectional reflectance distribution function model in spectral regions from UV to SWIR: model development and application to ground-based, aircraft and satellite observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188, 269-285.

83. Meng, G., Wen, Y., Zhang, M., Gu, Y., Xiong, W., Wang, Z. & Niu, S. (2022) The status and development proposal of carbon sources and sinks monitoring satellite system. *Carbon Neutrality*, 1, 32.

84. Miehe, R., Finkbeiner, M., Sauer, A. & Bauernhansl, T. (2022a) A system thinking normative approach towards integrating the environment into value-added accounting—paving the way from carbon to environmental neutrality. *Sustainability*, 14, 13603.

85. Miehe, R., Finkbeiner, M., Sauer, A. & Bauernhansl, T. (2022b) A system thinking normative approach towards integrating the environment into value-added accounting—paving the way from carbon to environmental neutrality. *Sustainability*.

86. Milodowski, D. T., Smallman, T. L. & Williams, M. (2023) Scale variance in the carbon dynamics of fragmented, mixed-use landscapes estimated using model–data fusion. *Biogeosciences*.

87. Moffat, A. M., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G. & Desai, A. R. (2007) Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agricultural and forest meteorology*, 147, 209-232.

88. Mou, C., Xie, Z., Li, Y., Liu, H., Yang, S. & Cui, X. (2023) Urban carbon price forecasting by fusing remote sensing images and historical price data. *Forests*.

89. Mu, H., Li, H. & Chen, Z. (2024) Research on the impact of change features for remote sensing image change detection *MIPPR 2023: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, SPIE.

90. Nguyen, A. & Saha, S. (2024) Machine learning and multi-source remote sensing in forest carbon stock estimation: a review. *arXiv preprint arXiv:2411.17624*.

91. Nie, M., Chen, D., Chen, H. & Wang, D. (2025) AutoMTNAS: automated meta-reinforcement learning on graph tokenization for graph neural architecture search. *Knowledge-Based Systems*, 310, 113023.

92. Niu, X., Liu, Y. N. & Ma, M. (2024) A transformer-based model for time series prediction of remote sensing data *International Conference on Intelligent Computing*.

93. Oehmcke, S., Li, L., Trepekli, K., Revenga, J. C., Nord-Larsen, T., Gieseke, F. & Igel, C. (2024) Deep point cloud regression for above-ground forest biomass estimation from airborne LiDAR. *Remote Sensing of Environment*, 302, 113968.

94. Ogunyemi, F. (2023a) AI-powered carbon accounting: transforming ESG reporting standards for a sustainable global economy. *International Journal of Multidisciplinary Research in Science, Engineering and Technology*, 6.

95. Ogunyemi, F. (2023b) AI-powered carbon accounting: transforming ESG reporting standards for a sustainable global economy. *International Journal of Multidisciplinary Research in Science, Engineering and Technology*, 6.

96. Orobinskaya, V. N., Mishina, T. N., Mazurenko, A. P. & Mishin, V. V. (2024) Problems of interpretability and transparency of decisions made by AI *2024 6th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, IEEE.

97. Padarian, J., Minasny, B. & Mcbratney, A. B. (2019) Using deep learning for digital soil mapping. *SOIL*, 5, 79-89.

98. Padarian, J., Minasny, B. & Mcbratney, A. B. (2019) Online machine learning for collaborative biophysical modelling. *Environmental Modelling & Software*, 122, 104548.

99. Pan, X., Zhao, J. & Xu, J. (2019) An object-based and heterogeneous segment filter convolutional neural network for high-resolution remote sensing image classification. *International Journal of Remote Sensing*, 40, 5892 - 5916.

100. Paris Agreement (2015) Paris agreement *report of the conference of the parties to the United Nations framework convention on climate change (21st session, 2015: Paris). Retrived December*, HeinOnline.

101. Park, J., Choi, B. & Lee, S. (2022) Examining the impact of adaptive convolution on natural language understanding. *Expert Systems with Applications*, 189, 116044.

102. Pashaei, M., Kamangir, H., Starek, M. J. & Tissot, P. E. (2020) Review and evaluation of deep learning architectures for efficient land cover mapping with UAS hyper-spatial imagery: a case study over a wetland. *Remote. Sens.*, 12, 959.

103. Peng, W., Haron, N. A., Alias, A. H. & Law, T. H. (2023) Knowledge map of climate change and transportation: a bibliometric analysis based on CiteSpace. *Atmosphere*, 14, 434.

104. Peón, J., Fernández, S., Recondo, C. & Calleja, J. F. (2017) Evaluation of the spectral characteristics of five hyperspectral and multispectral sensors for soil organic carbon estimation in burned areas. *International Journal of Wildland Fire*, 26, 230-239.

105. Piao, S., He, Y., Wang, X. & Chen, F. (2022) Estimation of china's terrestrial ecosystem carbon sink: methods, progress and prospects. *Science China Earth Sciences*, 65, 641-651.

106. Pimenow, S., Pimenowa, O., Prus, P. & Niklas, A. (2025) The impact of artificial intelligence on the sustainability of regional ecosystems: current challenges and future prospects. *Sustainability*, 17, 4795.

doi:10.20944/preprints202509.0050.v1

107. Protocol, K. (1997) United nations framework convention on climate change. *Kyoto Protocol, Kyoto*, 19, 1-21.

108. Qian, H., Bao, N., Meng, D., Zhou, B., Lei, H. & Li, H. (2024) Mapping and classification of liao river delta coastal wetland based on time series and multi-source GaoFen images using stacking ensemble model. *Ecological Informatics*, 80, 102488.

109. Raihan, A. (2024) Enhancing carbon sequestration through tropical forest management: a review. *Journal of Forestry and Natural Resources*, 3, 33-48.

110. Rastetter, E. B., Griffin, K. L., Kwiatkowski, B. L. & Kling, G. W. (2023) Ecosystem feedbacks constrain the effect of day-to-day weather variability on land–atmosphere carbon exchange. *Global Change Biology*, 29, 6093 - 6105.

111. Rasul, K., Ashok, A., Williams, A. R., Ghonia, H., Bhagwatkar, R., Khorasani, A., Bayazi, M. J. D., Adamopoulos, G., Riachi, R. & Hassen, N. (2023) Lag-llama: towards foundation models for probabilistic time series forecasting. *arXiv preprint arXiv:2310.08278*.

112. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. & Prabhat, F. (2019a) Deep learning and process understanding for data-driven earth system science. *Nature*, 566, 195-204.

113. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. & Prabhat, F. (2019b) Deep learning and process understanding for data-driven earth system science. *Nature*, 566, 195-204.

114. Reiersen, G., Dao, D., Lütjens, B., Klemmer, K., Amara, K., Steinegger, A., Zhang, C. & Zhu, X. (2022) ReforesTree: a dataset for estimating tropical forest carbon stock with deep learning and aerial imagery *Proceedings of the AAAI Conference on Artificial Intelligence*.

115. Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N. & Waldman-Brown, A. (2022) Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55, 1-96.

116. Romijn, E., Herold, M., Kooistra, L., Murdiyarso, D. & Verchot, L. (2012) Assessing capacities of non-annex i countries for national forest monitoring in the context of REDD+. *Environmental science & policy*, 19, 33-48.

117. Rossel, R. V. & Behrens, T. (2010) Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158, 46-54.

118. Roychowdhury, S., Diligenti, M. & Gori, M. (2021) Regularizing deep networks with prior knowledge: a constraint-based approach. *Knowl. Based Syst.*, 222, 106989.

119. Ruiz-Sánchez, R., Arencibia-Jorge, R., Tagüeña, J., Jiménez-Andrade, J. L. & Carrillo-Calvet, H. (2024) Exploring research on ecotechnology through artificial intelligence and bibliometric maps. *Environmental Science and Ecotechnology*, 21, 100386.

120. Safari, A., Sohrabi, H., Powell, S. & Shataee, S. (2017) A comparative assessment of multi-temporal landsat 8 and machine learning algorithms for estimating aboveground carbon stock in coppice oak forests. *International Journal of Remote Sensing*, 38, 6407-6432.

121. Sahu, Y. & Upadhyay, V. K. (2024) Environmental application of digital twins: a review. *Environmental Risk and Resilience in the Changing World: Integrated Geospatial AI and Multidimensional Approach*, 287-295.

122. Sangha, K. K., Ahammad, R., Russell-Smith, J. & Costanza, R. (2024) Payments for ecosystem services opportunities for emerging nature-based solutions: integrating indigenous perspectives from australia. *Ecosystem Services*, 66, 101600.

123. Schimel, D., Pavlick, R., Fisher, J. B., Asner, G. P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K. & Cox, P. (2015) Observing terrestrial ecosystems and the carbon cycle from space. *Global Change Biology*, 21, 1762-1776.

124. Searchinger, T. D., Wirsenius, S., Beringer, T. & Dumas, P. (2018) Assessing the efficiency of changes in land use for mitigating climate change. *Nature*, 564, 249-253.

125. Sen, D., Deora, B. S. & Vaishnav, A. (2025) Explainable deep learning for time series analysis: integrating SHAP and LIME in LSTM-based models. *Journal of Information Systems Engineering and Management*.

126. Shah, S. A. A., Manzoor, M. A. & Bais, A. (2020) Canopy height estimation at landsat resolution using convolutional neural networks. *Mach. Learn. Knowl. Extr.*, 2, 23-36.

127. Soman, M. K. & Indu, J. (2022) Sentinel-1 based inland water dynamics mapping system (SIMS). *Environmental Modelling & Software*, 149, 105305.

128. Song, Y., Yang, L., Li, B., Hu, Y., Wang, A., Zhou, W., Cui, X. & Liu, Y. (2017) Spatial prediction of soil organic matter using a hybrid geostatistical model of an extreme learning machine and ordinary kriging. *Sustainability*, 9, 754.

129. Song, Z., Liang, L., Zhao, X., Zhang, Y., Li, M. & Wang, Y. (2024) Estimation of terrestrial ecosystem NEP coupled with remote sensing and ground data *2024 12th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, IEEE.

130. Streck, C., Keenlyside, P. & Von Unger, M. (2016) The paris agreement: a new beginning. *Journal for European Environmental & Planning Law*, 13, 3-29.

131. Suleymanov, A., Gabbasova, I., Komissarov, M., Suleymanov, R., Garipov, T., Tuktarova, I. & Belan, L. (2023) Random forest modeling of soil properties in saline semi-arid areas. *Agriculture*, 13, 976.

132. Sultana, N., Hossain, S. Z., Abusaad, M., Alanbar, N., Senan, Y. & Razzak, S. A. (2022) Prediction of biodiesel production from microalgal oil using bayesian optimization algorithm-based machine learning approaches. *Fuel*, 309, 122184.

133. Sun, P., Wu, Y., Xiao, J., Hui, J., Hu, J., Zhao, F., Qiu, L. & Liu, S. (2019) Remote sensing and modeling fusion for investigating the ecosystem water-carbon coupling processes. *Science of the total environment*, 697, 134064.

134. Takahashi, T., Yamamoto, K., Miyachi, Y., Senda, Y. & Tsuzuku, M. (2006) The penetration rate of laser pulses transmitted from a small-footprint airborne LiDAR: a case study in closed canopy, middle-aged pure sugi (cryptomeria japonica d. Don) and hinoki cypress (chamaecyparis obtusa sieb. Et zucc.) Stands in japan. *Journal of Forest Research*, 11, 117-123.

135. Tasneem, S. & Islam, K. A. (2023) Development of trustable deep learning model in remote sensing through explainable-AI method selection *2023 IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, IEEE.

136. Tempel, F., Groos, D., Ihlen, E. A. F., Adde, L. & Strümke, I. (2024) Choose your explanation: a comparison of SHAP and GradCAM in human activity recognition. *arXiv preprint arXiv:2412.16003*.

137. Terziyan, V. Y. & Vitko, O. (2023) Causality-aware convolutional neural networks for advanced image classification and generation. *Procedia Computer Science*.

138. Tian, H., Zhu, J., Lei, X., Chen, X., Zeng, L., Jian, Z., Li, F. & Xiao, W. (2024) Timber and carbon sequestration potential of chinese forests under different forest management scenarios. *Advances in Climate Change Research*, 15, 1121-1129.

139. Tsui, O. W., Coops, N. C., Wulder, M. A. & Marshall, P. L. (2013) Integrating airborne LiDAR and space-borne radar via multivariate kriging to estimate above-ground biomass. *Remote Sensing of Environment*, 139, 340-352.

140. Ugbemuna Ugbaje, S., Karunaratne, S., Bishop, T. F. A., Gregory, L., Searle, R., Coelli, K. & Farrell, M. (2024) Space-time mapping of soil organic carbon stock and its local drivers: potential for use in carbon accounting. *Geoderma*.

141. Unesco (2023) *UNESCO's recommendation on the ethics of artificial intelligence: key facts*, United Nations Educational, Scientific and Cultural Organization.

142. Vahedi, A. A. (2017) Monitoring soil carbon pool in the hyrcanian coastal plain forest of iran: artificial neural network application in comparison with developing traditional models. *Catena*, 152, 182-189.

143. Vašát, R., Kodešová, R. & Borůvka, L. (2017) Ensemble predictive model for more accurate soil organic carbon spectroscopic estimation. *Computers & Geosciences*, 104, 75-83.

144. Velastegui-Montoya, A., Montalván-Burbano, N., Peña-Villacreses, G., de Lima, A. & Herrera-Franco, G. (2022) Land use and land cover in tropical forest: global research. *Forests*, 13, 1709.

145. Wang, B., Chen, Y., Yan, Z. & Liu, W. (2024) Integrating remote sensing data and CNN-LSTM-attention techniques for improved forest stock volume estimation: a comprehensive analysis of baishanzu forest park, china. *Remote. Sens.*, 16, 324.

146. Wang, F., Liu, J., Qin, G., Zhang, J., Zhou, J., Wu, J., Zhang, L., Thapa, P., Sanders, C. J. & Santos, I. R. (2023) Coastal blue carbon in china as a nature-based solution toward carbon neutrality. *The Innovation*, 4.

147. Wang, J., Chen, Y., Feng, W., Yu, H., Huang, M. & Yang, Q. (2019) Transfer learning with dynamic distribution adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11, 1 - 25.

148. Wang, J., Feng, L., Palmer, P. I., Liu, Y., Fang, S., Bösch, H., O Dell, C. W., Tang, X., Yang, D., Liu, L. & Xia, C. (2020) Large chinese land carbon sink estimated from atmospheric carbon dioxide data. *Nature*, 586, 720-723.

149. Wang, J., Sun, X., Cheng, Q. & Cui, Q. (2020) An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *The Science of the total environment*, 143099

150. Wang, J., Yu, G., Han, L., Yao, Y., Sun, M. & Yan, Z. (2024) Ecosystem carbon exchange across china's coastal wetlands: spatial patterns, mechanisms, and magnitudes. *Agricultural and Forest Meteorology*, 345, 109859.

151. Wang, L., Bai, Y., Wang, J., Qin, F., Liu, C., Zhou, Z. & Jiao, X. (2022) A transferable learning classification model and carbon sequestration estimation of crops in farmland ecosystem. *Remote Sensing*, 14, 5216.

152. Wang, M., Wang, Y., Teng, F. & Ji, Y. (2023) The spatiotemporal evolution and impact mechanism of energy consumption carbon emissions in china from 2010 to 2020 by integrating multisource remote sensing data. *Journal of Environmental Management*, 346, 119054.

153. Wang, N., Guo, Z., Shang, D. & Li, K. (2024) Carbon trading price forecasting in digitalization social change era using an explainable machine learning approach: the case of china as emerging country evidence. *Technological Forecasting and Social Change*.

154. Wang, P., Haron, N. A., Alias, A. H. & Law, T. H. (2023) Knowledge map of climate change and transportation: a bibliometric analysis based on CiteSpace. *Atmosphere*, 14, 434.

155. Wang, Q., Zhang, M., Jiang, X. & Li, R. (2022) Does the COVID-19 pandemic derail US-china collaboration on carbon neutrality research? A survey. *Energy Strategy Reviews*, 43, 100937.

156. Wang, Q., Zhang, W., Xia, J., Ou, D., Tian, Z. & Gao, X. (2024) Multi-scenario simulation of land-use/land-cover changes and carbon storage prediction coupled with the SD-PLUS-InVEST model: a case study of the tuojiang river basin, china. *Land*, 13, 1518.

157. Wang, T. (2024) Improved random forest classification model combined with c5. 0 algorithm for vegetation feature analysis in non-agricultural environments. *Scientific Reports*, 14, 10367.

158. Wang, T., Zhang, Y., Yue, C., Wang, Y., Wang, X., Lyu, G., Wei, J., Yang, H. & Piao, S. (2025) Progress and challenges in remotely sensed terrestrial carbon fluxes. *Geo-Spatial Information Science*, 28, 1-21.

159. Wang, X., Wang, L., Li, S., Wang, Z., Zheng, M. & Song, K. (2022) Remote estimates of soil organic carbon using multi-temporal synthetic images and the probability hybrid model. *Geoderma*.

160. Wang, X., Zhu, J., Yan, Z., Zhang, Z., Zhang, Y., Chen, Y. & Li, H. (2022) LaST: label-free self-distillation contrastive learning with transformer architecture for remote sensing image scene classification. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.

161. Were, K., Bui, D. T., Dick, Ø. B. & Singh, B. R. (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an afromontane landscape. *Ecological Indicators*, 52, 394-403.

162. Whata, A. & Chimedza, C. (2022) Evaluating uses of deep learning methods for causal inference. *IEEE Access*, PP, 1-1.

163. Wiesmeier, M., Barthold, F., Blank, B. & Kögel-Knabner, I. (2011) Digital mapping of soil organic matter stocks using random forest modeling in a semi-arid steppe ecosystem. *Plant and soil*, 340, 7-24.

164. Wong, D. (2018) VOSviewer. *Technical Services Quarterly*, 35, 219-220.

165. Wong, W. Y., Al-Ani, A. K. I., Hasikin, K., Khairuddin, A. S. M., Razak, S. A., Hizaddin, H. F., Mokhtar, M. I. & Azizan, M. M. (2021) Water, soil and air pollutants' interaction on mangrove ecosystem and corresponding artificial intelligence techniques used in decision support systems-a review. *Ieee Access*, 9, 105532-105563.

166. Xi, L., Shu, Q., Sun, Y., Huang, J. & Song, H. (2023) Carbon storage estimation of mountain forests based on deep learning and multisource remote sensing data. *Journal of Applied Remote Sensing*, 17, 014510 - 014510.

167. Xie, T. & Wang, X. (2024) Investigating the nonlinear carbon reduction effect of AI: empirical insights from china's provincial level. *Frontiers in Environmental Science*, 12, 1353294.

168. Xu, W., He, J. & Shu, Y. (2020) Transfer learning and deep domain adaptation. *Advances and Applications in Deep Learning*.

169. Ye, F., Li, X. & Zhang, X. (2019) FusionCNN: a remote sensing image fusion algorithm based on deep convolutional neural networks. *Multimedia Tools and Applications*, 78, 14683-14703.

170. Yewle, A. D., Mirzayeva, L. & Karakuş, O. (2025) Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. *arXiv preprint arXiv:2502.06062*.

171. Zhang, H., Sun, Y., Shi, W., Guo, D. & Zheng, N. (2021) An object-based spatiotemporal fusion model for remote sensing images. *European Journal of Remote Sensing*, 54, 86-101.

172. Zhang, J., Li, H., Miao, Y., Zhou, Z., Lyu, H. & Gong, Z. (2025) Remote sensing retrieval method based on few-shot learning: a case study of surface dissolved organic carbon in jiangsu coastal waters, china. *IEEE Access*, 13, 3014-3025.

173. Zhang, R. & Ma, J. (2009) State of the art on remotely sensed data classification based on support vector machines. *Advances in Earth Science*, 24, 555-562.

174. Zhang, Y., Zhang, Y., Yang, J., Wu, W. & Tao, R. (2024) Integrating system dynamics, land change models, and machine learning to simulate and predict ecosystem carbon sequestration under RCP-SSP scenarios: fusing land and climate changes. *Land*, 13, 1967.

175. Zhao, S., Ke, C., Dong, X., Li, J. & Feng, X. (2005) Landsat TM multispectral classification using support vector machine method in low-hill areas *Image Processing and Pattern Recognition in Remote Sensing II*, SPIE.

176. Zhao, W., Wu, Z. & Yin, Z. (2021) Estimation of soil organic carbon content based on deep learning and quantile regression *2021 ieee international geoscience and remote sensing symposium igarss*, IEEE.

177. Zheng, M., Wen, Q., Xu, F. & Wu, D. (2025) Regional forest carbon stock estimation based on multi-source data and machine learning algorithms. *Forests*, 16, 420.

178. Zhong, C., Li, T., Bi, R., Sanganyado, E., Huang, J., Jiang, S., Zhang, Z. & Du, H. (2023) A systematic overview, trends and global perspectives on blue carbon: a bibliometric study (2003–2021). *Ecological Indicators*, 148, 110063.

179. Zhou, Y., Wei, G., Wang, Y., Wang, B., Quan, Y., Wu, Z., Liu, J., Bian, S., Li, M. & Fan, W. (2025) Estimating regional forest carbon density using remote sensing and geographically weighted random forest models: a case study of mid-to high-latitude forests in china. *Forests*, 16, 96.

180. Zhu, Z., Wang, S. & Woodcock, C. E. (2015) Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote sensing of Environment*, 159, 269-277.

181. Zodage, P., Harianawala, H., Shaikh, H. & Kharodia, A. (2024) Explainable AI (XAI): history, basic ideas and methods. *International Journal of Advanced Research in Science, Communication and Technology*, 3.