






Review

Forestry Big Data: A Review and Bibliometric Analysis

Wen Gao ^{1,2} , Quan Qiu ^{2,3} , Changyan Yuan ^{2,4}, Xin Shen ¹ , Fuliang Cao ¹ , Guibin Wang ¹ and Guangyu Wang ^{2,*} 

¹ Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China

² Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada

³ College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China

⁴ School of Economics and Management, Beijing Forestry University, Beijing 100083, China

* Correspondence: guangyu.wang@ubc.ca; Tel.: +1-604-822-2681

Abstract: Due to improved data collection and processing techniques, forestry surveys are now more efficient and accurate, generating large amounts of forestry data. Forestry Big Data (FBD) has become a critical component of the forestry inventory investigation system. In this study, publications on FBD were identified via the Web of Science database, and a comprehensive bibliometric analysis, network analysis, and analysis of major research streams were conducted to present an overview of the FBD field. The results show that FBD research only began nearly a decade ago but has undergone an upswing since 2016. The studies were mainly conducted by China and the US, and collaboration among authors is relatively fragmented. FBD research involved interdisciplinary integration. Among all the keywords, data acquisition (data mining and remote sensing) and data processing (machine learning and deep learning) received more attention, while FBD applications (forecasting, biodiversity, and climate change) have only recently received attention. Our research reveals that the FBD research is still in the infancy stage but has grown rapidly in recent years. Data acquisition and data processing are the main research fields, whereas FBD applications have gradually emerged and may become the next focus.

Keywords: forestry; big data; bibliometric analysis; VOSviewer; Bibliometrix; Citespace



Citation: Gao, W.; Qiu, Q.; Yuan, C.; Shen, X.; Cao, F.; Wang, G.; Wang, G. Forestry Big Data: A Review and Bibliometric Analysis. *Forests* **2022**, *13*, 1549. <https://doi.org/10.3390/f13101549>

Academic Editors: Xiaoli Zhang, Dengsheng Lu, Xiujuan Chai, Guijun Yang, Langning Huo and Timothy A. Martin

Received: 22 June 2022

Accepted: 20 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Global data has increased in an unprecedented manner during the last few decades. In the previous five years, the global volume of created and copied data increased nearly ninefold, and it has doubled at least every two years [1]. As a result of the explosion of available information, the concept of big data was born [2]. The term “big data” first appeared in the 1980s, and the official concept was launched at the EMC World 2011 in Las Vegas [3]. Typically, big data refers to data that contains greater variety, increased volume, and faster velocity [4]. These characters are also known as the three Vs. [5]. Variety: the data comes from different sources and has become more varied in format. It has broken out of the previously defined categories of structured, semistructured, and unstructured data [6]. Volume: this includes the amount of data gathered, saved, and computed. The volume of data is enormous, larger than both terabytes and petabytes. The magnitude and upsurge of data exceed traditional storage and analysis techniques [4]. Velocity: the speed at generating and processing data to meet the demands and challenges of growth and development. For time-bound processes, big data should be utilized as it flows into the organization so that its value is optimized [4]. Big data is more continuously produced, as opposed to small data. Among the types of velocity related to big data are the generation frequency and handling, recording, and publishing frequency [7]. Veracity, value, and also other “Vs” are added by some organizations to describe big data. The Vs of big data are

therefore regularly referred to as the “four Vs” or the “five Vs”. Big data introduces new challenges and opportunities because of its volume, velocity, and veracity [8].

Forests dominate the Earth’s land surface, occupying 4.06 billion hectares, or approximately 31% of the total land area, and contributing about 75% and 86% of the total primary production and biomass of terrestrial ecosystems, respectively [9,10]. As one of the major components of the Earth’s ecosystems, forests play an increasingly essential role in improving livelihoods, maintaining ecological balance, regulating water and atmospheric cycles, reducing soil erosion, and alleviating climate change [9,11,12]. Traditional forestry surveys are mainly based on sampling techniques and visual methods, which are labor-intensive and relatively inefficient, making it challenging to conduct large-area forest resource surveys [13]. These techniques were significantly improved after the 1920s when rapid photo interpretation introduced aerial photogrammetry, photo interpretation, and mathematical statistics into forest surveys [14]. Remote sensing technology is able to offer detailed contiguous spatial, multidimensional, and vast spectral data. With the application of remote sensing technology in forestry, scientists and forestry investigation staff can make accurate predictions of forest structure parameters based on the structural and spectral characteristics of remote sensing data [15–17].

Currently, after a century of continuous and rapid technological progress, more and more advanced data collection and processing technologies are being employed to make forestry surveys more accurate and intelligent, thus generating more and more forestry data. As a kind of active remote sensing technique, Light Detection and Ranging (LiDAR) is, for example, able to effectively penetrate the forest and provide knowledge of the forest’s horizontal distribution and vertical canopy structure [18]. LiDAR methods have been successfully applied to forestry inventory and have been attempted for harvested wood assessments [19]. Radio Frequency Identification (RFID) is a barcode technology that uses optical scanners to read tags [20]. Trees on a plot can be tagged with inexpensive and long-lasting RFID tags that can store data about most features of the trees [21]. Since the velocity and efficiency of forestry data collection have been dramatically improved, a large amount of data has been generated. Big data can simplify the complexity of calculation, which is not only beneficial to data mining and utilization, but which can also make forestry data processing more efficient. The application of big data technology to forestry can consist of various present opportunities. Forestry Big Data (FBD) is now becoming an intrinsic aspect of the forestry inventory investigation system. As an essential research area, a study of the current situation, key areas, and new trends in FBD would generate important results.

Bibliometrics is an interdisciplinary science that qualitatively and quantitatively analyzes all knowledge of publications through mathematical and statistical methods [22–24]. As a relatively mature method, it generally measures the number of scholarly publications and citations, coauthorship, and cocitation among authors, institutes, and journals, as well as keywords. This technique is commonly utilized to quickly evaluate the merits of a research field, evaluate its current status, and determine its direction [25,26]. To date, bibliometric analysis has been widely used as part of complementary research in a wide range of fields [27], including fine roots, forest carbon sequestration, and medical big data [6,28,29]. Several articles on FBD have been completed in the past several years. This study analyzes the research through a bibliometric analysis and visualization. The objective of this study is to investigate the characteristics of the FBD field and to systematically and objectively illustrate the past research status, research hotspots, and emerging research trends in order to promote a clearer idea of the development trends and outlook of FBD.

The remainder of this article begins with an overview of FBD in Section 2, explaining how the field has been defined and developed in the past. Section 3 introduces the structured methodology used to gather and further refine the literature that will be evaluated in this study. Some general observations are also made in Section 3 before presenting a detailed analysis using VOSviewer, Citespace, and Bibliometrix network analysis tools in Sections 4 and 5. An evaluation of the leading research streams is presented in Section 6.

Section 7 is the discussion, while Section 8 summarizes the results and some limitations of this study.

2. The Definition and Applications of FBD

The definition of FBD has not been clearly defined. It is based on the big data technology application in forestry, which refers to many aspects of forestry. As for forestry resources management, the present studies are focused on the reform of the management system [3]. The majority of forestry data collection systems are built by government and research institutions, with relatively little participation from private companies [30,31]. To achieve digital forestry, forestry data need to be combined with big data platforms for collection, cleaning, storage, calculation, analysis, and information mining. Developed countries have already noticed the importance of sharing forestry data, while developing countries have not [32].

Forestry is an economic field that deals with natural products, especially wood [33]. Among all subjects of forestry output, wood processing has many applications in forestry big data. Sawmilling operations are one of the most important units in the timber supply chain, and long-term data monitoring is a requirement to improve the operational efficiency and make the processes more accurate [34]. Data collection, exchange, and connectivity throughout the timber supply chain have become increasingly important. There appears to be strong potential for optimizing the integration of harvester production data throughout the timber supply chains [35]. Products and technologies are now available for the digitization of forest operations. Areas of increased efficiency include wearable technology to map individual saplings, optimized harvesting programs based on knowledge of tree and product information, high-resolution mapped forests, improved machine navigation using automation and robotics, digital silvicultural processing, and networking tracking products [36]. A new direction in sustainable adaptive forest management is climate-smart forestry (CSF), which requires much more data than traditional forestry and relies on technologies such as the Internet of Things (IoT) for the continuous, real-time monitoring of optimal conditions to improve the potential of forests to adapt to and mitigate climate change [37]. In addition, big data plays a huge role in forest catastrophe monitoring and control [38].

3. Research Methodology and Initial Data Statistics

3.1. Data Collection

We performed a literature search on the Web of Science Core Collection (WoSCC) database from its inception to 30 April 2020. The search string is as follows: TOPIC = (“big data” AND (“forestry” OR “forest*” OR “plantation*” NOT “random forest*”), including all languages. The literature type was set up as “article” and “review”. The data retrieved included a total of 1080 papers, including 992 articles and 88 reviews. Publications from WoSCC were exported in plain text format. All information from the databases, including the number of documents, the number of citations, authors, affiliations, countries, titles, keywords, publication year, journals, and references, were downloaded and integrated for bibliometric analysis and visualization.

3.2. Initial Data Statistics

The first publication in WoSCC in the field of FBD was in 2012. Both in terms of articles and reviews, the annual publications generally show an upward trend (Figure 1). It is worth noting that there was a significant increase in 2016, 2018, and especially in 2020. We separated the whole period into four stages according to annual production and growth rates: period I (2012–2015), period II (2016–2017), period III (2018–2019), and period IV (2020–2022). In period I, there were no more than ten publications per year, but the citation rate of publications in 2014 was relatively high. In period II, both publications and citations improved, showing an overall stable trend. The citation rate had increased significantly. In period III, the number of publications was still growing and the citation rate increased

slightly. During period IV, the growth in the number of publications is particularly notable, reaching a peak of publications in 2021. The citation rate started to drop in 2020, however.

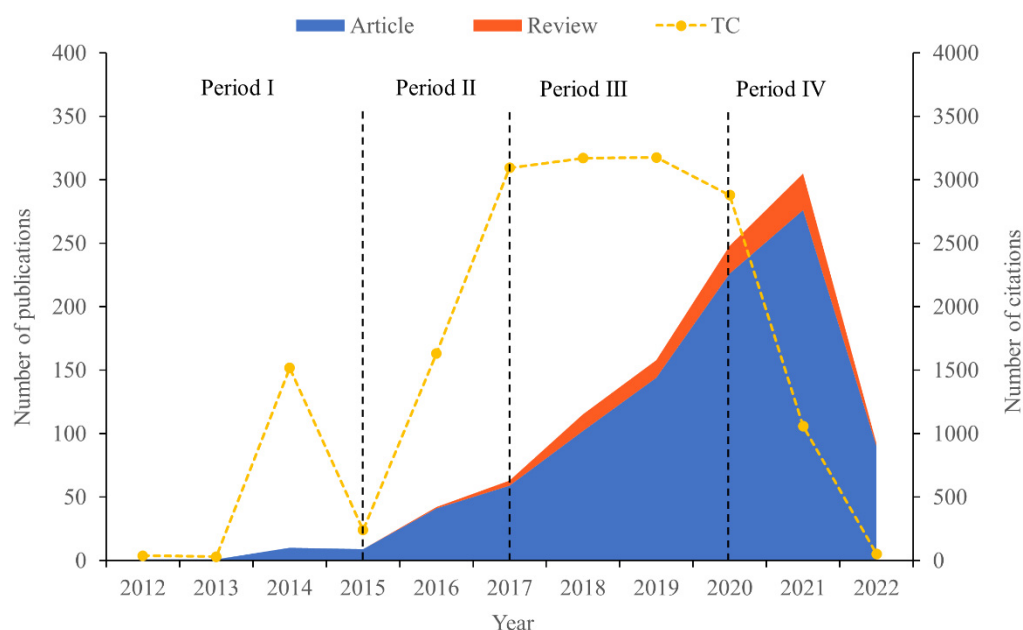


Figure 1. Annual trends in FBD-related publications and citations. TC = total number of citations.

3.3. Statistic Analysis

The bibliometric analysis in our research principally concentrates on three sections to render a thorough analysis and to uncover the current status and the future of FBD research. The three sections are (1) bibliometric analysis, (2) network analysis of publications, and (3) main research streams of FBD, presented in Sections 4–6, respectively. The bibliometric analysis using Citespace and Bibiometrix provides additional data statistics, including author, affiliation, country, sources, keyword, and reference statistics. Network analysis of publications uses VOSviewer, Citespace, and Bibliometrix to perform the coauthorship networks of authors, institutes, countries, and journals, and the dual-map overlay of the existing literature of FBD. The third section presents the timeline view of clusters of keywords and references, strategic diagrams, and thematic evolution of FBD-related publications using Bibliometrix and Citespace. In our strategic diagram and thematic evolution analysis, in order to make the results more concise and intuitive, we removed the terms that already appeared in the search string, such as “forest”, “big data”, etc.

VOSviewer (Version 1.6.18, Leiden, The Netherlands) (<https://www.vosviewer.com>, accessed on 30 May 2022) [39], which is an accessible, intuitive, and user-friendly bibliometric visualizer [40], was used to visualize complex cocitation and coauthorship networks of FBD research, including the collaboration and time trends amongst institutes and authors. A node indicates a source, an author, or an institute. The node’s size indicates the quantity of publications/citations, the line’s boldness indicates the connection’s intensity, and the node’s color indicates different groups or times.

Citespace (Version 5.8.R3, Philadelphia, USA) (<https://citespace.podia.com>, accessed on 30 May 2022) [41] is a free application based on Java that allows the visualization and analysis of trends in the scientific literature. It was used to conduct an analysis of the knowledge domains and emerging trends in FBD, including dual-map overlay, cluster analysis, a timeline view of references, and keywords and citation bursts.

Based on the R language, the Bibliometrix R Package (Version 4.0.0, Naples, Italy) (<https://www.bibliometrix.org>, accessed on 30 May 2022) [42] is an open-source instrument that is used to carry out science mapping [42,43]. We used the Bibliometrix Package to calculate the general statistics of publications, citations, authors, institutes, countries, and

journals. An analysis of the thematic evolution of FBD research was also conducted to categorize its development into different periods.

4. Bibliometric Analysis

4.1. Active Authors, Institutes, Countries, and Journals

According to Figure 2a, even the most high-yield authors have published no more than ten articles. The most prolific authors are MOSAVIA A (10 publications with 206 citations), KHOSHGOFTAAR TM (9 publications with 81 citations), and LI Y (7 publications with 51 citations) (Table 1). LI Y and WANG Y published their research relatively early, publishing their first results in the FBD field in 2016. BRISCO BJ only started publishing his FBD research in 2020 but has received 276 citations. The H-index of active authors ranges from 3 to 7. MOSAVI A and KHOSHGOFTAAR TM have the highest H-index. BRISCO B also has the highest mean citations per article (5 publications with 276 citations). The top 10 institutes with the highest output in FBD research are shown in Figure 2b. Tsinghua University has published the most papers in the FBD field ($n = 25$), accounting for 2.31% of all publications, followed by the University of Chinese Academy of Sciences ($n = 23$) and the University of Florida ($n = 23$). From 2012 to April 2022, 91 countries published articles on FBD research. The top 10 most published countries published approximately 72.78% ($n = 786$) of the world's articles (Figure 2c). China showed the highest output, with 301 FBD-related articles, followed by the US ($n = 182$) and India ($n = 59$).

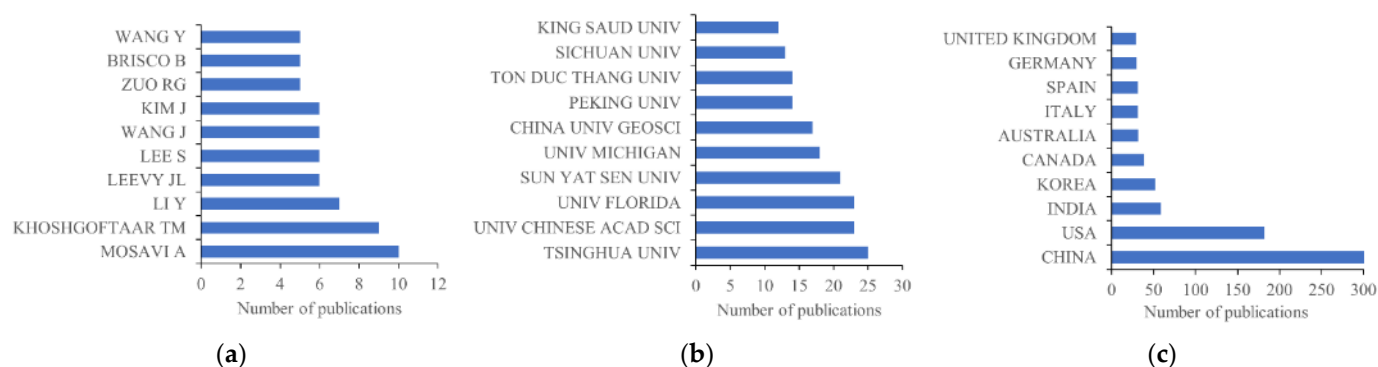


Figure 2. The top 10 most productive authors (a), institutes (b), and countries (c) for FBD-related publications.

Table 1. The production of the top ten authors over time.

Author	NP	TC	AC	H-Index	PY-Start
MOSAVI A	10	206	20.60	7	2020
KHOSHGOFTAAR TM	9	81	9.00	7	2019
LI Y	7	51	7.29	5	2016
LEEYV JL	6	62	10.33	5	2019
LEE S	6	113	18.83	4	2018
WANG J	6	62	10.33	4	2019
KIM J	6	54	9.00	3	2018
ZUO RG	5	225	45.00	5	2017
BRISCO B	5	276	55.20	4	2020
WANG Y	5	44	8.80	4	2016

The full names of the abbreviations in this table are: NP = The total number of publications; TC = The total number of citations; AC = The average number of citations per article; PY-start = The year of the first publication.

We identify the most active journals according to Bradford's law. Our dataset had thirty-five core journals (Figure 3), and the top five were: *Remote Sensing* (ISSN: 2072-4292), *IEEE Access* (ISSN: 2169-3536), *Journal of Big Data* (ISSN: 2196-1115), *Sustainability* (ISSN: 2071-1050), and *Sensors* (ISSN: 1424-8220).

burst. Global forest loss and gain were mapped using Earth observation satellite data from 2000 to 2012 in this paper [46]. No references with a strong citation burst have continued to maintain this level of burst.

References	Year	Strength	Begin	End	2012 - 2022
Hansen MC, 2013, SCIENCE, V342, P850	2013	3.79	2016	2018	
Natekin A, 2013, FRONT NEUROBOTICS, V7, P0, DOI 10.3389/fnbot.2013.00021, DOI	2013	2.98	2016	2017	
del Rio S, 2014, INFORM SCIENCES, V285, P112, DOI 10.1016/j.ins.2014.03.043, DOI	2014	3.71	2017	2019	
Fan JQ, 2014, NATL SCI REV, V1, P293, DOI 10.1093/nsr/nwt032, DOI	2014	2.59	2017	2019	
Hampton SE, 2013, FRONT ECOL ENVIRON, V11, P156, DOI 10.1890/120103, DOI	2013	2.58	2017	2018	
LeCun Y, 2015, NATURE, V521, P436, DOI 10.1038/nature14539, DOI	2015	4.5	2018	2019	
Zhou LN, 2017, NEUROCOMPUTING, V237, P350, DOI 10.1016/j.neucom.2017.01.026, DOI	2017	2.98	2018	2019	
Chen M, 2014, MOBILE NETW APPL, V19, P171, DOI 10.1007/s11036-013-0489-0, DOI	2014	2.98	2018	2019	
Schmidhuber J, 2015, NEURAL NETWORKS, V61, P85, DOI 10.1016/j.neunet.2014.09.003, DOI	2015	2.76	2018	2019	
Krizhevsky A, 2017, COMMUN ACM, V60, P84, DOI 10.1145/3065386, DOI	2017	2.49	2018	2019	

Figure 5. The top 10 references with the strongest citation bursts.

5. Network Analysis of Publications

Figure 6 provides a timeline visualization of the coauthorship networks of authors. The node color indicates the mean time of author collaborations and the node size indicates the number of associations. The distance between nodes and the link thickness indicates the level of cooperation among authors. Setting a threshold of ≥ 2 for the number of papers published by an author in the field, a total of 266 out of 4724 authors were eligible. Most of them do not connect with each other, however: only 13 authors have collaborative relationships. The cooperation between authors is relatively scattered. The author with the most collaborative relationships is Liu Xiaoping, followed by Chen Qian, Ye Tingting, and Yang Xuchao. There has been an increase, over time, in the number of authors who had cooperative relationships. Author cooperation was mainly distributed in 2017 and 2020.

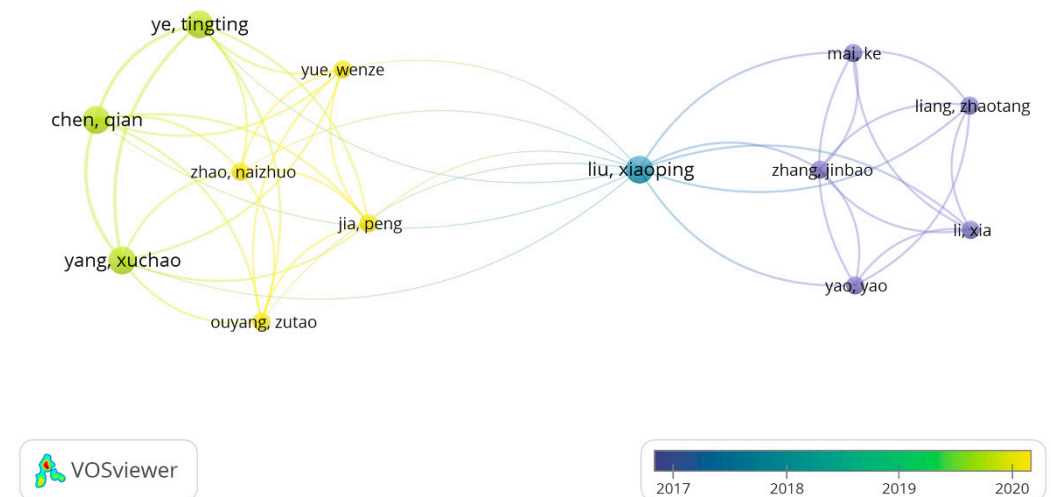


Figure 6. Timeline visualization of cooperation amongst authors.

Figure 7 shows the timeline visualizations of co-authorship networks of institutes constructed in the same way as for the authors. This process showed that a number of institutes are interconnected with each other. We set the threshold for the minimum number of publications by an institute at six. As a result, 40 institutes were selected from the 1834 institutes. A quantitative disparity is evident in the collaborative network. The University of the Chinese Academy of Sciences is the leading institute with the most collaborative relationships and is at the center of the collaborations. It is worth noting, additionally, that the Chinese Academy of Sciences maintains a close cooperative relationship with most institutes. Cooperation takes place, on average, in most institutes, from 2019 to 2020.

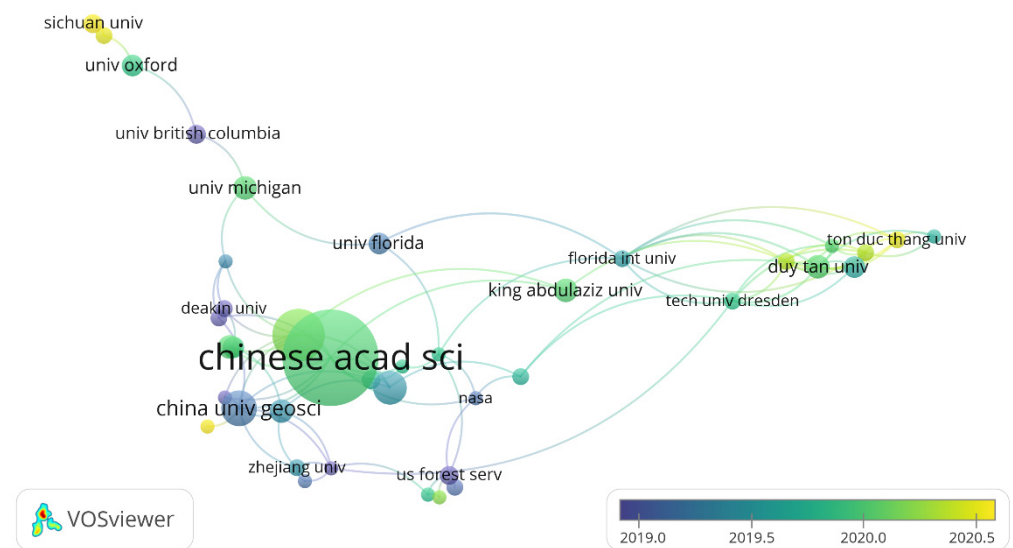


Figure 7. Timeline visualization of cooperation among institutes.

Figure 8 presents the global geographic mapping of publications and collaborations. The color shade of the country reflects the number of publications in that country, where the darker the color, the higher the number. The link between countries shows a cooperative relationship with thicker lines, indicating a closer relationship between countries. As seen from the collaboration map, Asia, Europe, America, Africa, and Oceania all participated in FBD research, but the participation of African countries was relatively low. The research is mainly distributed in the United States, China, Australia, and European countries, forming a complex cooperation network. Although the United States has not published as much research as China, it has a more complex and diverse cooperative relationship than China. The United States not only has the biggest number of academic ties with China, but also has good cooperative relations with almost all other countries. Overall, the results of the collaboration assessment indicate that the US and China are in a leading position in FBD research and have maintained good cooperative relations with countries from all over the world.

Country Collaboration Map

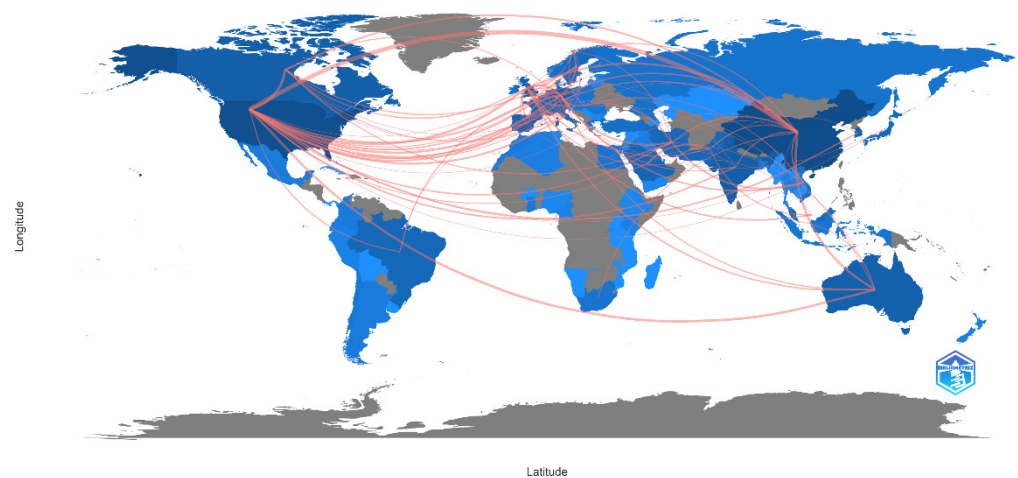


Figure 8. Country/region collaboration map.

Figure 9 presents the cocitation network of journals. We set the lowest number of citations at 30, and 473 journals met this threshold from the total of 18,407 journals. Every node refers to a journal, and the node size indicates the number of citations that the journal

has received. Each color represents a cluster that is formed based on cocitations. Cocitation relationships between journals are represented by links, and the thicker the link, the more citations by journals. The most frequently cited journals associated with the FBD field are found in six distinct groups (yellow, green, teal, blue, purple, and red). These are: *Science of The Total Environment* (ISSN: 0048-9697), *Machine Learning* (ISSN: 1573-0565), *Science* (ISSN: 0036-8075), *Natural Resources Research* (ISSN: 1520-7439), *Remote Sensing of Environment* (ISSN: 0034-4257), and *IEEE Access* (ISSN: 2169-3536), respectively.

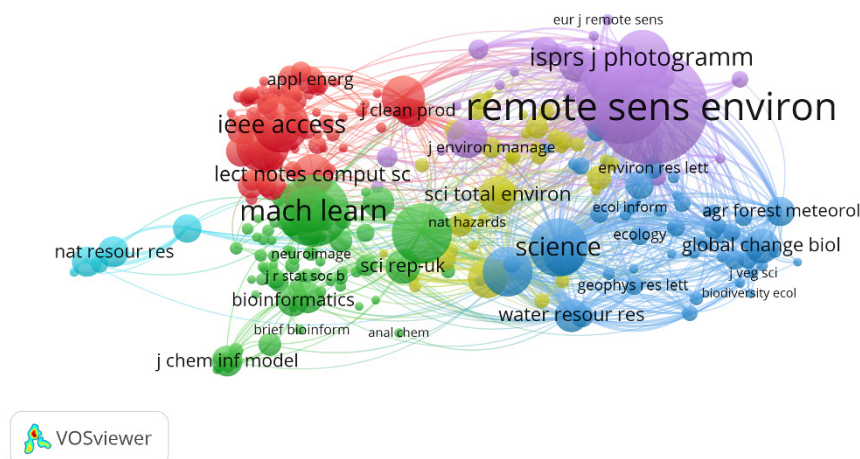


Figure 9. Journal cocitation network.

As shown by the dazzling blue and red lines in Figure 10, the dual map overlay of FBD research shows five main citation paths. The base map on the left is generated by mapping the citing journals, and the cited journals creating the base map are on the right. The different colors represent the different disciplines to which the journals belong. Research publications are therefore listed on the left and the references are listed on the right. The line path represents the citation relationship. By using citation relationships, the two maps are connected together to offer a perspective of the citation distribution of disciplines within FBD journals. The primary focus of the published articles containing FBD research was on journals in the field of “ecology, earth, marine” and “mathematics, systems, mathematical”. In contrast, most of the cited articles were published in journals in the field of “molecular, biology, genetics”, “plant, ecology, zoology”, “earth, geology, geophysics”, and “systems, computing, computer”. These publications are mainly outward-looking, absorbing knowledge from outside fields. This reflects the fact that FBD research involves interdisciplinary integration including mathematics, computational science, biology, molecular biology, ecology, and geography.

From the published journals, it is evident that FBD research is multidisciplinary and intersectional. The creation and development of big data are ultimately a result of the information explosion in this computer era [2] and the acquisition of big data in forestry is tied to RS, GIS, and GPS [47]. Logically, remote sensing and computational science are also closely related to FBD. According to Bradford’s law, the top five active journals are *Remote Sensing*, *IEEE Access*, the *Journal of Big Data*, *Sustainability*, and *Sensors*. Two of these journals are well-known journals in the remote sensing field, which also shows the vital role of remote sensing technology in the field of FBD. Apart from *Remote Sensing* and *Computational Science*, subject categories of knowledge that were found to be most influential include, but are not limited to, Ecology and Earth Science, Mathematics and System Science, Biology, and Genetics and Geology.

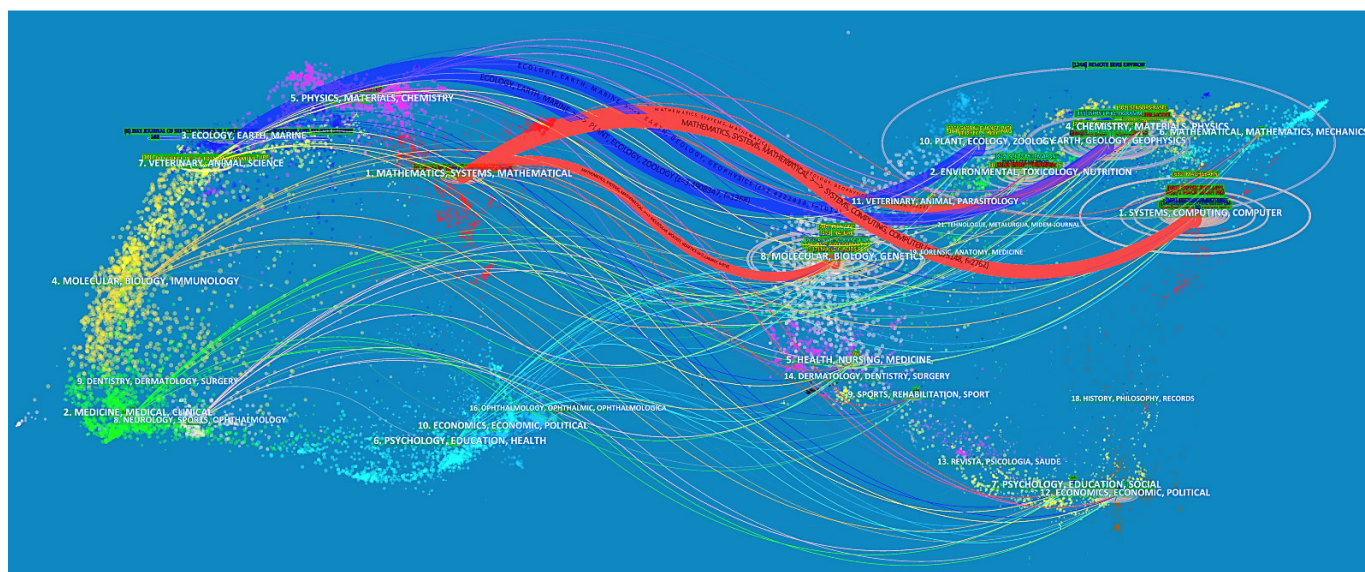


Figure 10. The dual-map overlay of FBD publications.

6. Main Research Streams of FBD

6.1. Timeline Distribution of the Cluster Analysis of the Keywords

We demonstrate the current situation in FBD research with a timeline pattern of existing clusters (Figure 11). This timeline view helps to discover better the trends in the hot topics in FBD research at different periods. The modularity Q is 0.4565 and the weighted mean silhouette S is 0.7095. Nine cocitation clusters are presented with their keywords. Cluster #1 “feature selection”, #2 “gaussian process”, #3 “data science”, and #7 “remote sensing” have the longest time span, which has lasted from the emergence of the FBD research to the present day. Cluster #0 “Google Earth engine”, #1 “feature selection”, #2 “gaussian process”, #3 “data science”, #4 “machine learning”, and #5 “feature extraction” are the most critical areas of FBD research, and they continue to be active research topics. Clusters #6 “risk” and #8 “identification” were emerging hotspots in FBD research, although no new literature has yet been published in 2022 on these themes. This does not mean that these topics are no longer being researched, however, because the research period of this paper is only until April 2022, which does not fully represent the whole of 2022. Using the timeline view, in addition to the curves linking these important nodes, we can view the central knowledge structure. The location of nodes marks the first year that the keywords appeared, and the node size represents the keyword frequency. It is worth noting that the essential nodes (centrality > 0.1) are externally covered by the pink outer ring. FBD research focused mostly on the keywords “remote sensing”, “model”, “algorithm”, “forest”, “management”, “deep learning”, “artificial intelligent”, “machine learning”, “data mining”, “classification”, “prediction”, and “identification”. Over time, the keywords change and become more diverse, with researchers placing great emphasis on machine learning and data mining in FBD research.

6.2. Timeline Distribution of the Cluster Analysis of the References

To have a better understanding of the core content of all the references, we use Citespace to describe the clustering of references, as shown in Figure 12a. All references are divided into 13 categories: the modularity Q is 0.8325 and the weighted mean silhouette S is 0.9478. The nine clusters with the highest K value are #0 “data mining”, #1 “Google Earth engine”, #2 “mineral prospectivity mapping”, #3 “monitoring”, #4 “electronic health records”, #5 “Apache Spark”, #6 “deep learning”, #7 “k-nearest neighbors”, and #9 “class imbalance”. We also performed a timeline view for clusters (Figure 12b). Each cluster was marked in a certain period, representing the time distribution and the precedence relationships between the clusters. The locations of the nodes show the first year that the

references appeared. We found that “data mining” and “k-nearest neighbors” are early fields in FBD research. Of all the clusters, the start time of “electronic health records” is the latest, with the research contents continuing to expand. The current hotspots in FBD research are “Google Earth engine”, “mineral prospectivity mapping”, “monitoring”, “electronic health records”, and “class imbalance”. These research hotspots are relatively late as a whole, however, because the development of FBD started late. The rise and fall of the research topics in the above nine clusters might indicate that the research in these clusters is reaching an end, maturing, or has changed to a new research path due to breakthrough discoveries.

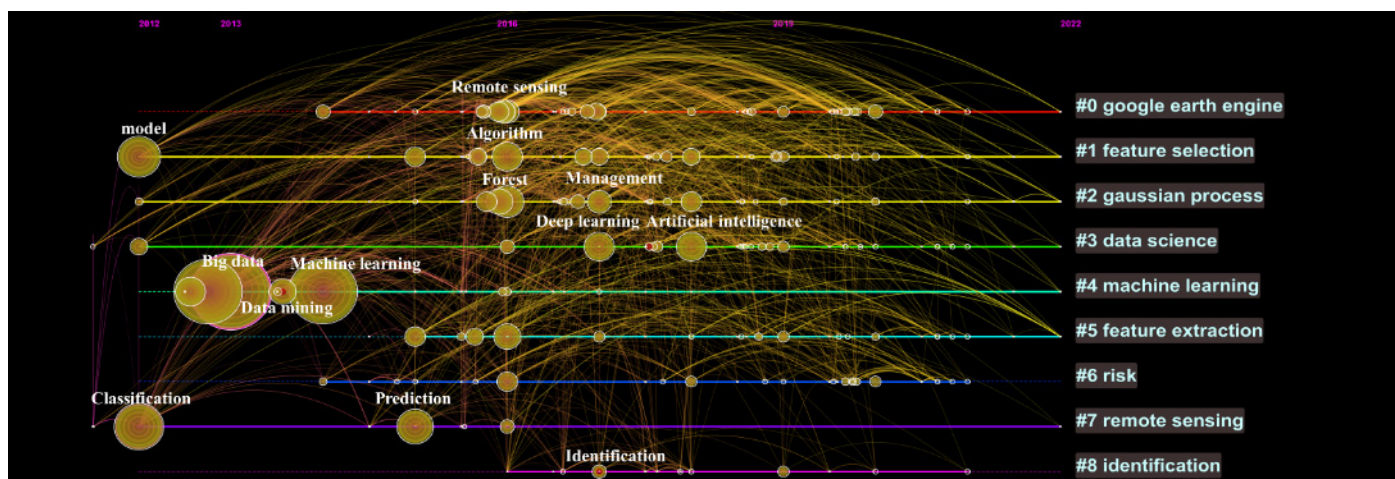


Figure 11. Timeline distribution of the cluster analysis of the keywords.

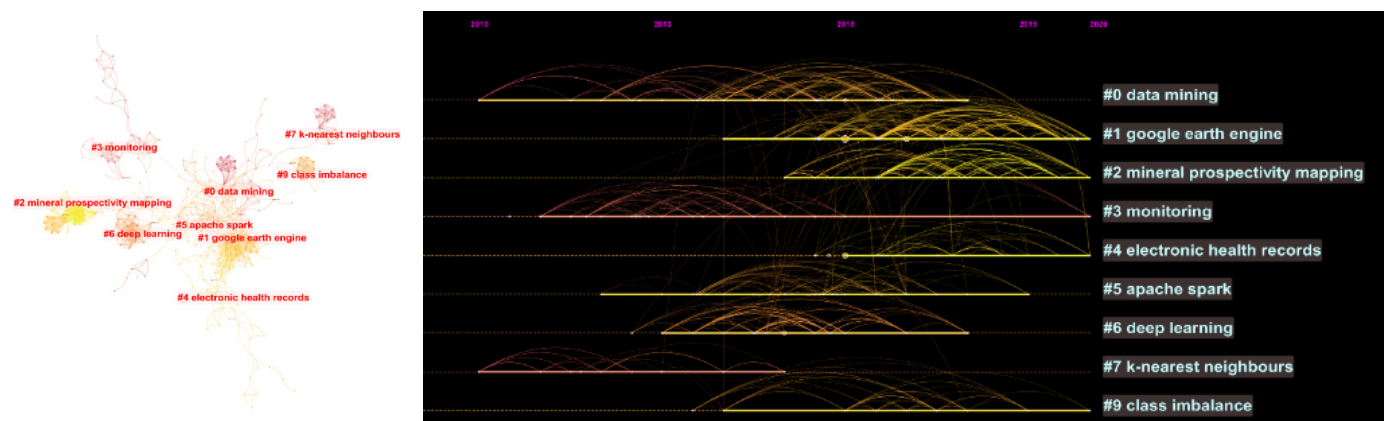


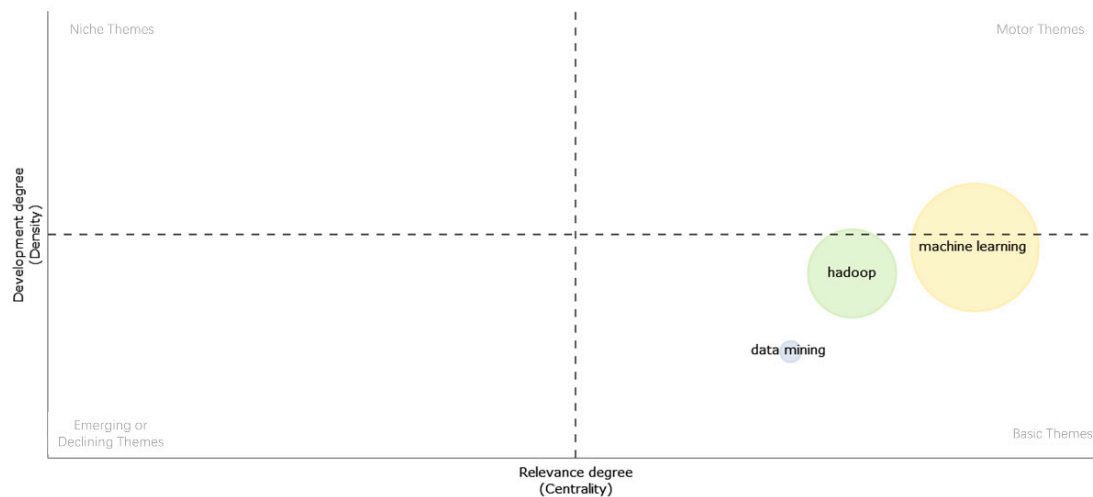
Figure 12. Visualization of cocited references. (a) Cluster analysis; (b) Timeline contribution of the top nine clusters.

6.3. Emerging Research Areas of FBD

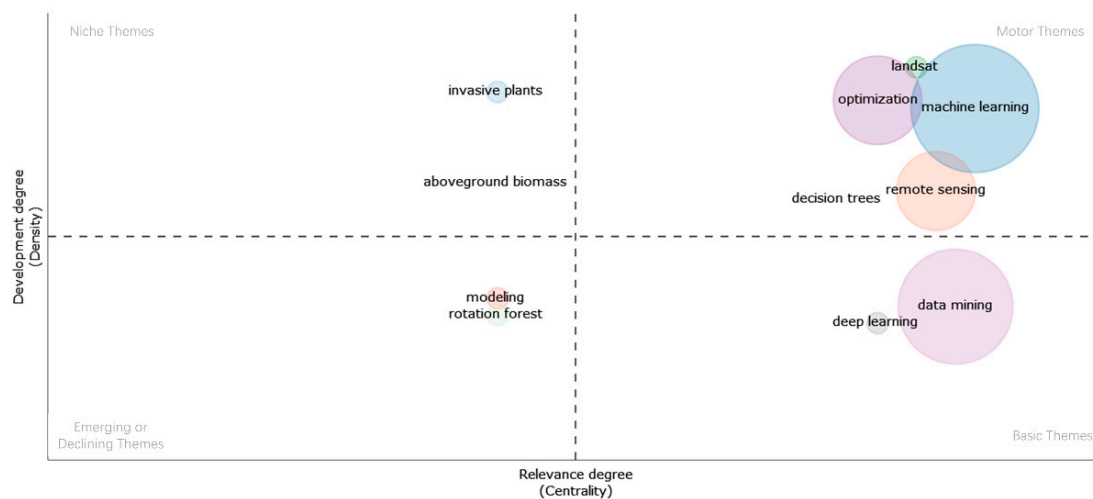
6.3.1. Strategic Diagram of FBD-Related Publications

We divide the time since FBD research emerged into four consecutive subperiods based on the publication volume, as shown in Figure 13: 2012–2015, 2016–2018, 2019–2020, and 2021–2022, in anticipation of revealing the evolution of FBD research over time. A strategic diagram based on centrality and density divides all topics into four quadrants, giving us a clear picture of the key research topics in FBD research. Density reveals the level of cohesion within a cluster. Centrality refers to the level of interaction among clusters [43,48]. The isolated themes (top left) indicate that the subject is highly specialized and is studied by relatively few researchers. The motor themes (top right) are critical to FBD research: these are the themes that are well developed and imperative for the field of FBD research. The emerging themes (bottom left) are relatively weakly developed compared

to themes in the other quadrants: they are themes that need to be developed as the field, as a whole, is moving towards them. Finally, the fundamental themes (bottom right) are important research directions for the future but have not received sufficient attention [48]. Specifically, each circle represents a topic that is named by the highest frequency keyword. The circle size depends on the keyword frequency with the circle area increasing with increasing frequency.

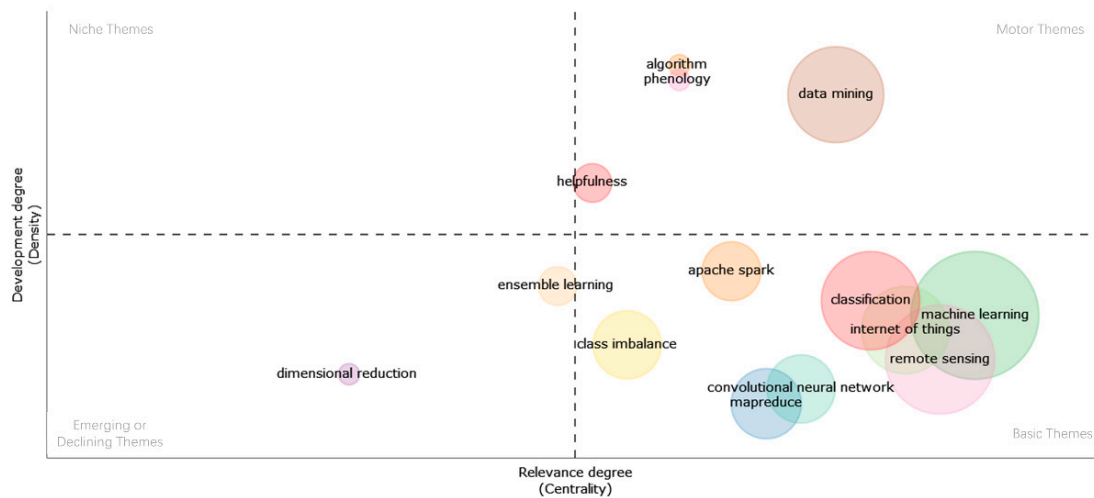


(a) 2012–2015

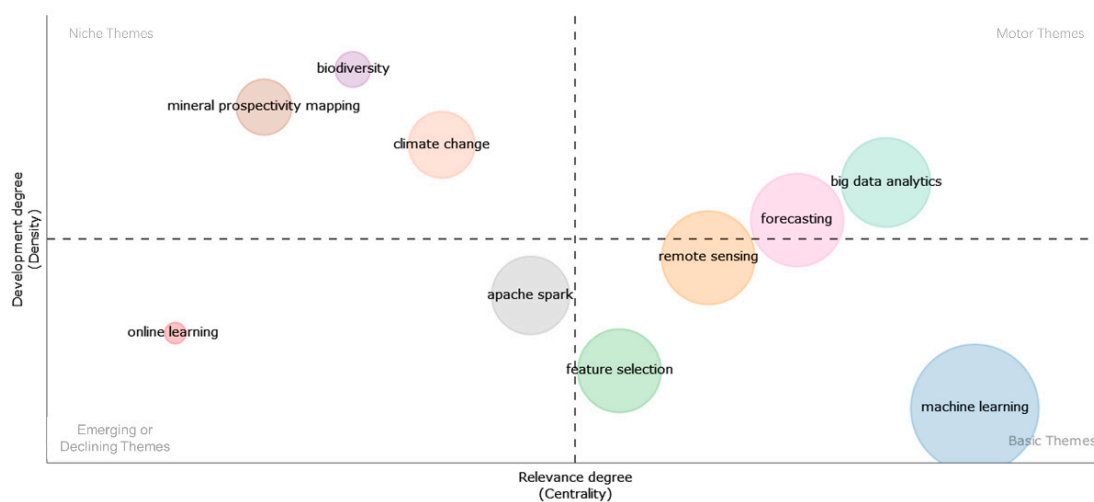


(b) 2016–2017

Figure 13. *Cont.*



(c) 2018–2019



(d) 2020–2022

Figure 13. Strategic diagram of FBD-related publications (2012–2022).

There were no isolated or emerging themes in the first subperiod (2012–2015). Research related to “Hadoop”, “machine learning”, and “data mining” were the future research trends. Among these topics, “machine learning” was partially well developed. In the second subperiod (2016–2017), the topic “remote sensing” was well developed, as were “landsat”, “optimization”, and “decision tree”, with “data mining” remaining a future research trend. Research on “invasive plants” and “aboveground biomass” appeared to be isolated, while the research on “modeling” and “rotation forest” was an emerging theme at this time. In the third subperiod (2018–2019), the topic of “internet of things” arose for the first time, as did “class imbalance”, “Apache Spark”, “mapreduce”, “convolutional neural network”, and “classification”, which were all the basic themes. In addition, “dimensional reduction” became an emerging theme. In the fourth subperiod (2020–2022), “forecasting” and “big data analytics” turned into well-developed themes. At the same time “Apache Spark” became an emerging theme, with research related to “remote sensing” and “machine learning” remaining the basic theme and critical directions for future research. The research

on “biodiversity”, “mineral prospectivity mapping”, and “climate change” became highly specialized themes.

Throughout the four subperiods in FBD research, and since its first appearance in the first phase, the topic “machine learning” appeared throughout the four subperiods. Its circle area was always the largest, which implies machine learning has become a significant research topic. With different themes relating to artificial intelligence appearing one after the other, there is no doubt that computational science is playing an essential part in the development of FBD research. It is noteworthy that the algorithms involved in the different subperiods included rotation forest, decision tree, and convolutional neural network, amongst others. Recently (2015–2022), the theme “remote sensing” has gained momentum and contributed significantly to FBD research’s development. In the first subperiod, research in relation to “Hadoop” was the fundamental theme of the research in this field, but in the last two subperiods it was replaced by “Spark”. At the same time, themes focused on the applications of big data were emerging including “aboveground biomass”, “phenology”, “forecasting”, “biodiversity”, and “climate change”.

6.3.2. Thematic Evolution of FBD-Related Publications

In Figure 14, we outline the thematic evolution of FBD-related publications. “Landsat”, “remote sensing”, and “optimization” from the second subperiod converged, in the third subperiod, to become “remote sensing”, and only a tiny branch of “climate change” was separated out in the fourth subperiod. Some initial themes were split and recombined into new themes for the next subperiod. For example, “machine learning”, which emerged in the first subperiod, spawned into “remote sensing”, “classification”, and “internet of things” in the third subperiod. Different topics, such as “ensemble learning” and “data mining”, were, however, re consolidated into “machine learning” in the fourth subperiod. It can be seen that the development of several technical themes eventually presented as the application of big data: “data mining”, “classification”, and “class imbalance” integrated into “forecast” in the last subperiod. In addition, “big data analytics” of the fourth subperiod evolved from “data mining”, “convolutional neural network”, and “internet of things” in the third period. According to the evolution of the themes, we can see that “remote sensing” and “machine learning” were the principal elements that formed the basic framework of the thematic research. The recent two years have seen the emergence of different applications of FBD which are now new hot topics: these represent important research directions for the future.

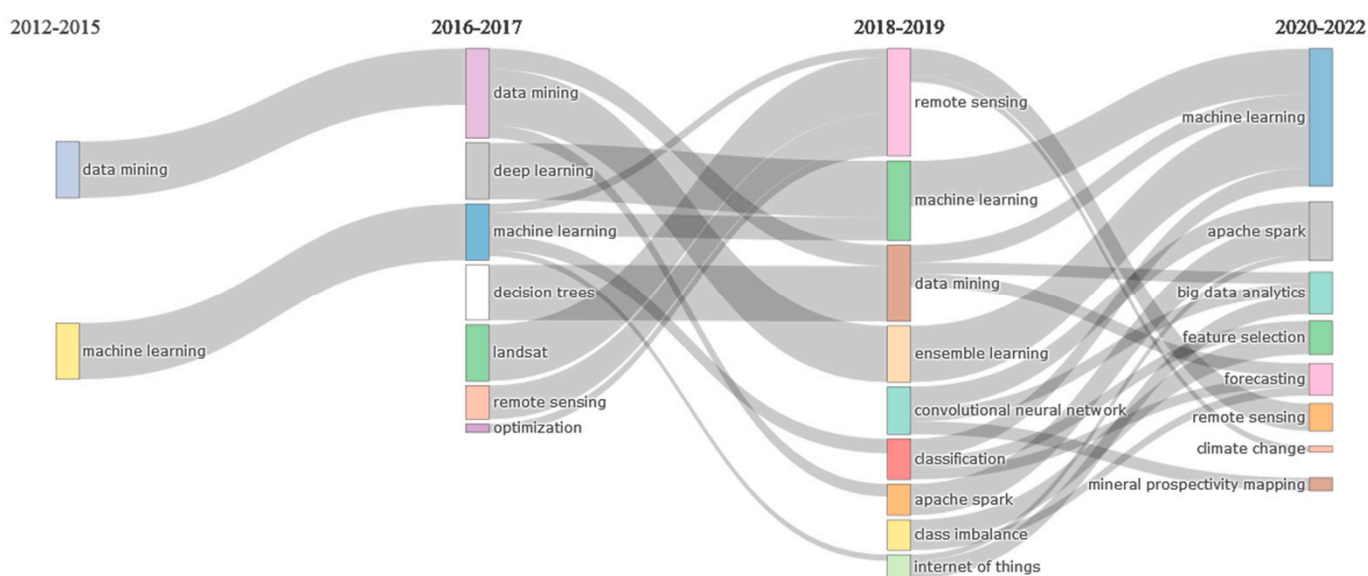


Figure 14. Thematic evolution of FBD-related publications (2012–2022).

7. Discussion

In this research, we analyzed the knowledge base and emerging trends of FBD research, highlighting the focus themes and future research directions. We observed a steady increase in the number of publications in the field of FBD. Since the formal concept of big data was not introduced in the EMC World until 2011 [3], research in the field of FBD started late with only a few sporadic articles each year until 2016. It is worth noting, however, that, in 2014, the number of citations was relatively high. This was due to the publication, in 2014, of a highly cited article on scalable nearest neighbor algorithms for high-dimensional data [49]. It was evident from this high citation rate that scientists have an interest in developing new algorithms to handle big data. Due to the significant effect of big data technology in forestry information management, the study of big data in forestry has become more and more critical [14,50]. After 2016, the number of publications rose rapidly. There was a significant increase in 2016, in 2018, and especially in 2020. Based on previous work [51], we divided the research history into four subperiods: period I (2012–2015), which had low publications and low citations; period II (2016–2017), with an increase in both publications and citations; period III (2018–2019), with a growing publication rate and the highest citation rate; period IV (2020–2022), with an increasing publication rate and decreasing citation rate. The low citation rate of papers in 2020 is due to the basic rule of article citation: more recent articles are less likely to have been cited as much as older articles.

Our research showed that even the most productive authors have not published more than ten articles. Research was published from 2016 onwards, and the most prolific author, MOSAVIA A. BRISCO BJ, with a research focus on the Google Earth engine [52–54], started his FBD research in 2020 but has already obtained 276 citations. Most of the researchers have no academic communication with each other and collaborations among the authors are relatively scattered. The number of authors where collaborations occur is increasing over time. Tsinghua University has published the most papers in FBD, but the University of the Chinese Academy of Sciences is the institute with the most collaborations. It is at the center of all collaborations and maintains close partnerships with most institutes. Ninety-one countries have published FBD research, with China having the highest output, similarly to other research fields [51,55,56], followed by the United States and India. Asia, Europe, the Americas, Africa, and Oceania are all involved in FBD research, forming a complex collaborative network, with African countries being the least engaged. The United States and China are leading the way in FBD research, with the United States having a more complex and diverse set of collaborations, although it is inferior to China in terms of volume.

The timeline of references and keywords suggested that studies associated with FBD involve computational science including data science, data mining, deep learning, k-nearest neighbors and class imbalance. As one of the earliest research topics to emerge, data mining is a comprehensive concept which literally means creating new knowledge from the vast amount of information that is available [57]. Most of the data mining work is realized through algorithmic tools provided by machine learning [58]. Deep learning, which is a popular topic in machine learning, has achieved excellent results in classifying and recognizing images and other media. It has been applied in forest structural parameter estimation [59]. The k-nearest neighbors algorithm is a nonparametric supervised learning method, while imbalanced classification is a classification problem posing a challenge for predictive modeling as machine learning algorithms [60,61]. The research showed that almost all computational science topics started at a relatively early stage, which implies that computer science is the foundation for the development of big data in forestry and, as such, played a massive role in the initial stages of its development. Studies related to FBD also involve data acquisition, including the Google Earth engine, feature selection, and monitoring. The Google Earth engine can provide geoinformation data and monitoring, and so can obtain and record the growth of trees in real-time.

According to the analyses of the references and keywords, some items have had the strongest citation bursts according to the bursts in references and keywords. One of these

bursts is “data mining”. It was the first keyword to appear and has lasted relatively longer than the other keywords. Data mining methods improve the accuracy and reliability of forest management decisions [62]. They can also be used to clean and extract data. We can use deep learning to classify and identify the cleaned and processed data and improve our results accordingly. One of the references with the highest citation burst was a *Nature* article on deep learning [45]. Deep learning has been applied to individual tree detection, tree species classification, and forest abnormality monitoring, including forest fires [63]. We also found that keyword citation bursts occurred for “pattern recognition” and “identification” in the last five years. In addition, an article in *Science* on the mapping of global forest loss using Earth observation satellite data was found to have a high citation burst rate [46], indicating that remote sensing data is a current hotspot for FBD research.

Looking at the thematic evolution chart for FBD keywords, all of the keywords can be broadly classified into several categories: data acquisition, data processing, and application. In terms of data acquisition, remote sensing and data mining are undoubtedly the most important parts, especially the acquisition and processing of remote sensing data that have been fully developed: the Google Earth engine has clearly played an important part in this. The Internet of Things emerged in the third subperiod. The Internet of Things is a system of computing devices, digital machines and other interrelationships which uses a common unique identifying Internet Protocol (IP) to form a global network and transmit data over the network. It has been applied to several disciplines [64]. The processing of big data involves knowledge of distinct computer disciplines, including machine learning, which has been instrumental for the development of FBD. Different topics about artificial intelligence have appeared one after another in different subperiods, including rotation forest, decision tree, and convolutional neural network. Computational science undoubtedly plays a vital role in the development of FBD. It is also worth noting that “Hadoop” was replaced by “Spark” in the last two subperiods. They are both open-source frameworks developed by the Apache Software Foundation and are both widely used in big data architectures. Apache Spark has become a unified engine for massive data analytics across various workloads [65]. Combined with the current development of big data technology, previous researchers found that the distributed system of Hadoop can be used to address the storage problem of large-scale forestry big data. In contrast, the memory-based Spark computing framework can be used to achieve fast processing of data in real-time [32]. In the later stages of development, studies on the applications of forestry big data, such as above-ground biomass, phenology, forecasting, monitoring, biodiversity, and climate change, have emerged [14,63,66,67]. There are, however, still many challenges facing FBD, including how to combine multiple sources of data, how to manage data, how to mine data, and how to analyze and utilize the data collected, which still require forestry researchers to continue with their research processes.

8. Conclusions and Limitation

Based on our scientific analysis, we determined that the development of research on FBD is still imperfect and in the infancy stage. The entire field started late and had a quite low publication rate before 2016, but the field has grown explosively from 2016 to the present. The most productive authors have not, however, published more than ten articles, and collaboration amongst authors is still relatively fragmented. Cooperation is, however, gradually improving. All five continents are involved in FBD research, forming a complex network of collaborations, with the United States and China leading the way in FBD research. The United States has the most complex and diverse collaborations, although not as many as China in terms of the volume of published research emerging from these collaborations. FBD is involved in interdisciplinary integration, including Remote Sensing and Computational Science, Ecology and Earth Science, Mathematics and System Science, Biology, and Genetics and Geology. The research involving data acquisition and data processing has, however, received more attention. The current research focus is data mining, remote sensing, machine learning, and deep learning. These research hotspots

continue to flourish, and forestry applications in big data have gradually emerged including forecasting, biodiversity, and climate change.

Despite the progress presented in this study, this study still has some limitations. We tried our best to improve the retrieval strategy but must have missed some articles due to the restrictions of machine retrieval. In addition, all the results were created using machine algorithms. Compared to manual induction, machine induction is still insufficient. Additionally, whilst research using large amounts of data has been conducted for a long period of time, it was not named “big data” prior to the first use of this term (in 2011), and so this research is not included in this paper. This research is, however, still very relevant to the origin and development of FBD and should be studied in depth. In addition, no results related to wood processing were obtained in this paper, but wood is a critical piece of forestry products and has relevant applications in terms of big data, which needs further research.

Author Contributions: Conceptualization, G.W. (Guangyu Wang) and C.Y.; methodology, C.Y.; validation, Q.Q.; formal analysis, W.G.; writing—original draft, W.G. and X.S.; writing—review and editing, Q.Q., F.C. and G.W. (Guibin Wang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by APFNet (2017SP2-UBC) and China Scholarship Council (no. 202108680002; no. 202008440171; no. 202006510054).

Data Availability Statement: All datasets presented in this study can be found within the article.

Acknowledgments: W.G., Q.Q. and C.Y. are grateful for support from the CSC (China Scholarship Council) Scholarship (no. 202108680002; no. 202008440171; no. 202006510054). We also would like to thank the creators of the three extraordinary tools: Chaomei Chen (Citespace), Nees Jan van Eck, and Ludo Waltman (VOSviewer), and Massimo Aria and Corrado Cuccurullo (Bibliometrix R Package).

Conflicts of Interest: The authors declare no conflict to interest.

References

- Chen, M.; Mao, S.; Liu, Y.; Chen, M.; Mao, S.; Liu, Y. Big Data: A Survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [\[CrossRef\]](#)
- Fan, J.; Han, F.; Liu, H. Challenges of Big Data Analysis. *Natl. Sci. Rev.* **2014**, *1*, 293–314. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, L.; Hao, T.; Chi, T. Evaluation on China’s Forestry Resources Efficiency Based on Big Data. *J. Clean. Prod.* **2017**, *142*, 513–523. [\[CrossRef\]](#)
- Sagiroglu, S.; Sinanc, D. Big Data: A Review. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013; pp. 42–47.
- Lu, R.; Zhu, H.; Liu, X.; Liu, J.; Shao, J. Toward Efficient and Privacy-Preserving Computing in Big Data Era. *IEEE Netw.* **2014**, *28*, 46–50. [\[CrossRef\]](#)
- Liao, H.; Tang, M.; Luo, L.; Li, C.; Chiclana, F.; Zeng, X.J. A Bibliometric Analysis and Visualization of Medical Big Data Research. *Sustainability* **2018**, *10*, 166. [\[CrossRef\]](#)
- Kitchin, R.; McArdle, G. What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets. *Big Data Soc.* **2016**, *3*. [\[CrossRef\]](#)
- Baru, C.; Bhandarkar, M.; Nambiar, R.; Poess, M.; Rabl, T. Benchmarking Big Data Systems and the BigData Top100 List. *Big Data* **2013**, *1*, 60–64. [\[CrossRef\]](#)
- FAO. *Global Forest Resources Assessment 2020*; FAO: Rome, Italy, 2020.
- Shen, X.; Cao, L.; Chen, D.; Sun, Y.; Wang, G.; Ruan, H. Prediction of Forest Structural Parameters Using Airborne Full-Waveform LiDAR and Hyperspectral Data in Subtropical Forests. *Remote Sens.* **2018**, *10*, 1729. [\[CrossRef\]](#)
- Shen, X.; Cao, L.; Coops, N.C.; Fan, H.; Wu, X.; Liu, H.; Wang, G.; Cao, F. Quantifying Vertical Profiles of Biochemical Traits for Forest Plantation Species Using Advanced Remote Sensing Approaches. *Remote Sens. Environ.* **2020**, *250*, 112041. [\[CrossRef\]](#)
- Busch, J.; Ferretti-Gallon, K. What Drives Deforestation and What Stops It? A Meta-Analysis. *Rev. Environ. Econ. Policy* **2017**, *11*, 3–23. [\[CrossRef\]](#)
- LaBau, V.J.; Bones, J.T.; Kingsley, N.P.; Lund, H.G.; Smith, W.B. *A History of the Forest Survey in the United States: 1830–2004*; US Department of Agriculture, Forest Service: Washington, DC, USA, 2007.
- Zou, W.; Jing, W.; Chen, G.; Lu, Y.; Song, H. A Survey of Big Data Analytics for Smart Forestry. *IEEE Access* **2019**, *7*, 46621–46636. [\[CrossRef\]](#)
- Foody, G.M. Remote Sensing of Tropical Forest Environments: Towards the Monitoring of Environmental Resources for Sustainable Development. *Int. J. Remote Sens.* **2003**, *24*, 4035–4046. [\[CrossRef\]](#)

16. Wulder, M.A.; White, J.C.; Nelson, R.F.; Næsset, E.; Ørka, H.O.; Coops, N.C.; Hilker, T.; Bater, C.W.; Gobakken, T. Lidar Sampling for Large-Area Forest Characterization: A Review. *Remote Sens. Environ.* **2012**, *121*, 196–209. [\[CrossRef\]](#)
17. Buddenbaum, H.; Seeling, S.; Hill, J. Fusion of Full-Waveform Lidar and Imaging Spectroscopy Remote Sensing Data for the Characterization of Forest Stands. *Int. J. Remote Sens.* **2013**, *34*, 4511–4524. [\[CrossRef\]](#)
18. Schulte to Bühne, H.; Pettorelli, N. Better Together: Integrating and Fusing Multispectral and Radar Satellite Imagery to Inform Biodiversity Monitoring, Ecological Research and Conservation Science. *Methods Ecol. Evol.* **2018**, *9*, 849–865. [\[CrossRef\]](#)
19. Borz, S.A.; Proto, A.R. Application and Accuracy of Smart Technologies for Measurements of Roundwood: Evaluation of Time Consumption and Efficiency. *Comput. Electron. Agric.* **2022**, *197*, 106990. [\[CrossRef\]](#)
20. Weinstein, R. RFID: A Technical Overview and Its Application to the Enterprise. *IT Prof.* **2005**, *7*, 27–33. [\[CrossRef\]](#)
21. Farve, R. *Using Radio Frequency Identification (RFID) for Monitoring Trees in the Forest: State-of-the-Technology Investigation*; United States Department of Agriculture (USDA): Washington, DC, USA, 2014.
22. Merigó, J.M.; Cancino, C.A.; Coronado, F.; Urbano, D. Academic Research in Innovation: A Country Analysis. *Scientometrics* **2016**, *108*, 559–593. [\[CrossRef\]](#)
23. Keathley-Herring, H.; Van Aken, E.; Gonzalez-Aleu, F.; Deschamps, F.; Letens, G.; Orlandini, P.C. Assessing the Maturity of a Research Area: Bibliometric Review and Proposed Framework. *Scientometrics* **2016**, *109*, 927–951. [\[CrossRef\]](#)
24. Moed, H.F. Bibliometric Indicators Reflect Publication and Management Strategies. *Scientometrics* **2000**, *47*, 323–346. [\[CrossRef\]](#)
25. Merigó, J.M.; Mas-Tur, A.; Roig-Tierno, N.; Ribeiro-Soriano, D. A Bibliometric Overview of the Journal of Business Research between 1973 and 2014. *J. Bus. Res.* **2015**, *68*, 2645–2653. [\[CrossRef\]](#)
26. Železnik, D.; Blažun Vošner, H.; Kokol, P. A Bibliometric Analysis of the Journal of Advanced Nursing, 1976–2015. *J. Adv. Nurs.* **2017**, *73*, 2407–2419. [\[CrossRef\]](#)
27. Wang, B.; Zhang, Q.; Cui, F. Scientific Research on Ecosystem Services and Human Well-Being: A Bibliometric Analysis. *Ecol. Indic.* **2021**, *125*, 107449. [\[CrossRef\]](#)
28. Huang, L.; Xia, Z.; Cao, Y. A Bibliometric Analysis of Global Fine Roots Research in Forest Ecosystems during 1992–2020. *Forests* **2022**, *13*, 93. [\[CrossRef\]](#)
29. Huang, L.; Zhou, M.; Lv, J.; Chen, K. Trends in Global Research in Forest Carbon Sequestration: A Bibliometric Analysis. *J. Clean. Prod.* **2020**, *252*, 119908. [\[CrossRef\]](#)
30. Bovenzi, M. Metrics of Whole-Body Vibration and Exposure-Response Relationship for Low Back Pain in Professional Drivers: A Prospective Cohort Study. *Int. Arch. Occup. Environ. Health* **2009**, *82*, 893–917. [\[CrossRef\]](#)
31. Ruslandi; Roopsind, A.; Sist, P.; Peña-Claros, M.; Thomas, R.; Putz, F.E. Beyond Equitable Data Sharing to Improve Tropical Forest Management. *Int. For. Rev.* **2014**, *16*, 497–503. [\[CrossRef\]](#)
32. Zhao, M.; Li, D.; Long, Y. Forestry Big Data Platform by Knowledge Graph. *J. For. Res.* **2021**, *32*, 1305–1314. [\[CrossRef\]](#)
33. Lummitsch, S.; Findeisen, E.; Haas, M.; Carl Sascha Lummitsch, C.; Carl, C. The Perspective of Optical Measurement Methods in Forestry. *Photonics Educ. Meas. Sci.* **2019**, *11144*, 346–351. [\[CrossRef\]](#)
34. Borz, S.A.; Păun, M. Integrating Offline Object Tracking, Signal Processing, and Artificial Intelligence to Classify Relevant Events in Sawmilling Operations. *Forests* **2020**, *11*, 1333. [\[CrossRef\]](#)
35. Hartsch, F.; Kemmerer, J.; Labelle, E.R.; Jaeger, D.; Wagner, T. Integration of Harvester Production Data in German Wood Supply Chains: Legal, Social and Economic Requirements. *Forests* **2021**, *12*, 460. [\[CrossRef\]](#)
36. Keefe, R.F.; Zimbelman, E.G.; Picchi, G. Use of Individual Tree and Product Level Data to Improve Operational Forestry. *Curr. For. Rep.* **2022**, *8*, 148–165. [\[CrossRef\]](#)
37. Torresan, C.; Garzón, M.B.; O'grady, M.; Robson, T.M.; Picchi, G.; Panzacchi, P.; Tomelleri, E.; Smith, M.; Marshall, J.; Wingate, L.; et al. A New Generation of Sensors and Monitoring Tools to Support Climate-Smart Forestry Practices. *Can. J. For. Res.* **2021**, *51*, 1751–1765. [\[CrossRef\]](#)
38. Li, J.; Zhang, C.; Liu, J.; Li, Z.; Yang, X. An Application of Mean Escape Time and Metapopulation on Forestry Catastrophe Insurance. *Phys. A Stat. Mech. Its Appl.* **2018**, *495*, 312–323. [\[CrossRef\]](#)
39. van Eck, N.J.; Waltman, L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics* **2010**, *84*, 523–538. [\[CrossRef\]](#)
40. Mourao, P.R.; Martinho, V.D. Forest Entrepreneurship: A Bibliometric Analysis and a Discussion about the Co-Authorship Networks of an Emerging Scientific Field. *J. Clean. Prod.* **2020**, *256*, 120413. [\[CrossRef\]](#)
41. Chen, C. A Glimpse of the First Eight Months of the COVID-19 Literature on Microsoft Academic Graph: Themes, Citation Contexts, and Uncertainties. *Front. Res. Metr. Anal.* **2020**, *5*, 607286. [\[CrossRef\]](#)
42. Aria, M.; Cuccurullo, C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *J. Informetr.* **2017**, *11*, 959–975. [\[CrossRef\]](#)
43. Wang, X.; Xu, Z.; Qin, Y. Structure, Trend and Prospect of Operational Research: A Scientific Analysis for Publications from 1952 to 2020 Included in Web of Science Database. *Fuzzy Optim. Decis. Mak.* **2022**. [\[CrossRef\]](#)
44. Zhou, L.; Zhang, L.; Zhao, Y.; Zheng, R.; Song, K. A Scientometric Review of Blockchain Research. *Inf. Syst. E-Bus. Manag.* **2021**, *19*, 757–787. [\[CrossRef\]](#)
45. Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)

46. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [[CrossRef](#)] [[PubMed](#)]
47. Lam-Gordillo, O.; Baring, R.; Dittmann, S. Ecosystem Functioning and Functional Approaches on Marine Macroinvertebrate Fauna: A Research Synthesis towards a Global Consensus. *Ecol. Indic.* **2020**, *115*, 106379. [[CrossRef](#)]
48. Forlano, C.; De Bernardi, P.; Yahiaoui, D. Entrepreneurial Universities: A Bibliometric Analysis within the Business and Management Domains. *Technol. Forecast. Soc. Change* **2021**, *165*, 120522. [[CrossRef](#)]
49. Muja, M.; Lowe, D.G. Scalable Nearest Neighbor Algorithms for High Dimensional Data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2227–2240. [[CrossRef](#)] [[PubMed](#)]
50. Hasan, S.S.; Zhang, Y.; Chu, X.; Teng, Y. The Role of Big Data in China's Sustainable Forest Management. *For. Econ. Rev.* **2019**, *1*, 96–105. [[CrossRef](#)]
51. Liu, T.; Yang, L.; Mao, H.; Ma, F.; Wang, Y.; Zhan, Y. Knowledge Domain and Emerging Trends in Podocyte Injury Research From 1994 to 2021: A Bibliometric and Visualized Analysis. *Front. Pharmacol.* **2021**, *12*, 3508. [[CrossRef](#)]
52. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Brisco, B.; Homayouni, S.; Gill, E.; DeLancey, E.R.; Bourgeau-Chavez, L. Big Data for a Big Country: The First Generation of Canadian Wetland Inventory Map at a Spatial Resolution of 10-m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform. *Can. J. Remote Sens.* **2020**, *46*, 15–33. [[CrossRef](#)]
53. Amani, M.; Ghorbanian, A.; Ahmadi, S.A.; Kakooei, M.; Moghimi, A.; Mirmazloumi, S.M.; Moghaddam, S.H.A.; Mahdavi, S.; Ghahremanloo, M.; Parsian, S.; et al. Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5326–5350. [[CrossRef](#)]
54. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 152–170. [[CrossRef](#)]
55. Kousis, A.; Tjortjis, C. Data Mining Algorithms for Smart Cities: A Bibliometric Analysis. *Algorithms* **2021**, *14*, 242. [[CrossRef](#)]
56. Herrera-Franco, G.; Montalván-Burbano, N.; Carrión-Mero, P.; Jaya-Montalvo, M.; Gurumendi-Noriega, M. Worldwide Research on Geoparks through Bibliometric Analysis. *Sustainability* **2021**, *13*, 1175. [[CrossRef](#)]
57. Colonna, L. A Taxonomy and Classification of Data Mining. *SMU Sci. Technol. Law Rev.* **2013**, *16*, 309.
58. Mannila, H. Data Mining: Machine Learning, Statistics, and Databases. In Proceedings of the 8th International Conference on Scientific and Statistical Data Base Management, SSDBM 1996, Stockholm, Sweden, 18–20 June 1996; Institute of Electrical and Electronics Engineers Inc.: Interlaken, Switzerland; pp. 2–8.
59. Liu, H.; Shen, X.; Cao, L.; Yun, T.; Zhang, Z.; Fu, X.; Chen, X.; Liu, F. Deep Learning in Forest Structural Parameter Estimation Using Airborne LiDAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 1603–1618. [[CrossRef](#)]
60. Krawczyk, B.; Woźniak, M.; Schaefer, G. Cost-Sensitive Decision Tree Ensembles for Effective Imbalanced Classification. *Appl. Soft Comput. J.* **2014**, *14*, 554–562. [[CrossRef](#)]
61. Sun, S.; Huang, R. An Adaptive K-Nearest Neighbor Algorithm. In Proceedings of the 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2010, Yantai, China, 10–12 August 2010; pp. 91–94.
62. Zhang, W.Q. Application Research of Data Mining Technology on Growth Management of Forestry. *Adv. Mater. Res.* **2014**, *846–847*, 995–998. [[CrossRef](#)]
63. Diez, Y.; Kentsch, S.; Fukuda, M.; Caceres, M.L.L.; Moritake, K.; Cabezas, M. Deep Learning in Forestry Using Uav-Acquired Rgb Data: A Practical Review. *Remote Sens.* **2021**, *13*, 2837. [[CrossRef](#)]
64. Keefe, R.F.; Wempe, A.M.; Becker, R.M.; Zimbelman, E.G.; Nagler, E.S.; Gilbert, S.L.; Caudill, C.C. Positioning Methods and the Use of Location and Activity Data in Forests. *Forests* **2019**, *10*, 458. [[CrossRef](#)]
65. Salloom, S.; Dautov, R.; Chen, X.; Peng, P.X.; Huang, J.Z. Big Data Analytics on Apache Spark. *Int. J. Data Sci. Anal.* **2016**, *1*, 145–164. [[CrossRef](#)]
66. Moradi, F.; Darvishsefat, A.A.; Pourrahmati, M.R.; Deljouei, A.; Borz, S.A. Estimating Aboveground Biomass in Dense Hyrcanian Forests by the Use of Sentinel-2 Data. *Forests* **2022**, *13*, 104. [[CrossRef](#)]
67. Klimetzek, D.; Stăncioiu, P.T.; Paraschiv, M.; Niță, M.D. Ecological Monitoring with Spy Satellite Images— the Case of Red Wood Ants in Romania. *Remote Sens.* **2021**, *13*, 520. [[CrossRef](#)]