Visualization applications of the bibliometric software SATI and Histcite

Jiaxin Zhuang[1], Qin Nie[2*], Wang Man[3], Chenghai You[4]

College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China

[1]zhuangjx126@126.com
[2*]nieqinhongyi@163.com
[3]manwang@xmut.edu.cn
[4]chenghai.you@qq.com

Abstract. Compared with the traditional literature review, the use of bibliometric software can carry out scientific and quantitative analysis of a large number of documents, showing the development context and research hotspots in this field. In order to explore the visual analysis effect of the combination of different bibliometric software, the forest biomass in remote sensing was used as the key word to carry out the theme research. Using the title analysis software SATI and the citation analysis software HistCite, combined with the keyword literature in the China National Knowledge Infrastructure database and the Web of Science database for visual quantitative analysis. The results show that the combination of functions of different bibliometric software can effectively complement the analysis results and make them more comprehensive and accurate.

Keywords: Computer software; SATI; HistCite; Remote sensing; Forest biomass

1 Introduction

For researchers who first come into contact with an unfamiliar discipline, it is very difficult to quickly find the key information, figure out the past and present development history of the research object, and analyze the hot spots in this field in the face of numerous literatures. However, the emergence of bibliometric software makes all of this simple. We only need to export some part of information from the database and import it into the bibliometric software to analyze the key information we need. By far the most common literature metrology software Citespace, VOSviewer, CiteNet, etc.

However, this approach of using bibliometric software for analysis also has disadvantages. First, some bibliometric software is inherently complex to operate and in-

V. Escudero et al. (Eds.): BADS 2022, AHCS 11, pp. 223–232, 2023.
https://doi.org/10.2991/978-94-6463-102-9_26

stead requires time cost to learn the software. Second, the literature sources of individual software are often single databases, and differences in literature information from different databases can lead to different analysis results.

In this paper, we used a combination of SATI and HistCite for bibliometric analysis, and selected two databases, China National Knowledge Infrastructure (CNKI) and Web of Science (WOS), to analyze "forest biomass" as a keyword in remote sensing in the interdisciplinary field of subject and computer, and after the experiment The results of the analysis in this field were quickly obtained, and the research hotspots of forest biomass in recent years were clarified. It is verified that the combination of different bibliometric software and databases has the advantage that the analysis results are effectively complementary and more accurate.

2 Materials and Methods

2.1 Methods

In this study, a bibliometric analysis on the topic of "forest biomass" was conducted with the help of SATI, a caption analysis tool, and HistCite, a citation mapping software.

SATI is a catalog analysis tool programmed in C# and designed and developed on the SATI is a C# programming tool designed for visual analysis of imported bibliographies through formatting and field extraction of the bibliographies, and has various functions such as basic statistics, natural language processing, word frequency statistics, knowledge matrix construction, co-occurrence analysis, and cluster analysis [1]. In this paper, we used SATI to analyze the hotspot research and publication trends of literature with the keyword "forest biomass", and imported 4625 valid literature from CNKI database into SATI to visualize and analyze the number and year of literature and keyword hotspots, and then summarize the progress of forest biomass The visual analysis was conducted to summarize the progress in the field of forest biomass.

HistCite is a free and open source citation mapping analysis software developed by Eugene Garfield, the inventor of Science Citation Index (SCI). It uses the frequency of citations among the literature as an indicator to locate the important literature in the field of study by means of chronological graphs and visualizes the relationships, importance and development history among individual papers [2]. In this paper, 2900 retrieved literature in the field of forest biomass were imported into HistCite software, analyzed based on the WOS core database, and the research history of forest biomass was presented in a chronological diagram. Finally, the 30 classical literature with the highest influence in this field were located with the help of HisctCite citation chronology function.

2.2 Data

CNKI is the most abundant and complete platform for Chinese literature research at present, which is used as the original data source for importing bibliographic analysis of SATI software for literature retrieval. The keyword "forest biomass" was searched

by December 31, 2021, and the literature types were journals and papers. 1351 articles and 516 articles from Chinese literature journals and 2791 articles from English journals were obtained. A total of 4625 effective literatures were obtained after screening and de-weighting.

WOS is the largest comprehensive academic resource database covering the most natural Sciences in the world. In this paper, SCI-Expanded and Social Sciences Citation Index in the core database of WOS are used as literature sources. The keyword "forest-biomass" was used as the subject search, and the time range was set as "all years". The search time was as of December 31, 2021. A total of 2900 valid literatures were obtained and imported into HistCite software for citation chronological analysis.

3 Results of the analysis based on SATI and CNKI databases

3.1 Number and year of literature

The CNKI database was searched for literature on the topic of "forest biomass", and a total of 4625 valid papers published between 1968 and 2021 were obtained, which were imported into SATI for literature analysis.

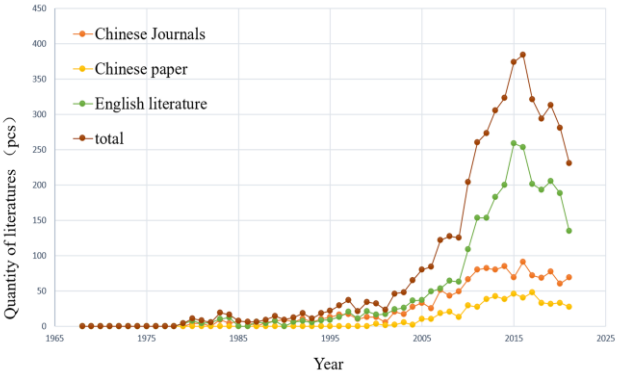

Fig. 1. Trends of annual publications in forest biomass related fields

The trend of the number of published journals and papers in the field of forest biomass at home and abroad was analyzed, and the results are shown in Figure 1. It can be found that the overall trend of the number of published literature per year is increasing. The research phases of forest biomass can be divided into: slow growth period (1968-1991), fluctuating growth period (1992-2006), and rapid growth period (2007-2021) years according to the different literature numbers. The average number of publications per year was 5.25 in the slow growth period, 37.87 in the fluctuating growth period, and 262.47 in the rapid growth period, and reached a historical peak of 384 publications per year in 2016, while the number began to fall back to 231 in 2021.

3.2 Keyword hotspot analysis

The frequency of keywords reflects the hot spots of research directions. The 4625 documents were imported into SATI for keyword analysis, with a total number of 18,463 keywords and an average of 3.99 keywords per article, and the similar keyword segments were combined to organize the top 10 keywords, as shown in Table 1.

Table 1. Forest biomass field high frequency keyword

Number	*Keyword*	*Count*	*Percentage*
1	生物量 (Biomass)	874	4.7338
2	森林生物量 (Forest Biomass)	646	3.4989
3	碳储量 (Carbon Reserve)	303	1.6411
4	地上生物量 (Above Ground Biomass)	236	1.2782
5	Bioenergy	117	0.6337
6	碳密度 (Carbon Density)	117	0.6337
7	Lidar	92	0.4983
8	森林生态系统 (Forest Ecosystem)	89	0.4820
9	遥感 (Remote Sensing)	74	0.4008
10	Climate Change	63	0.3412

The software was used to give the keyword co-linear network knowledge map, distinguishing between Chinese and English keywords, as shown in Figure 2.

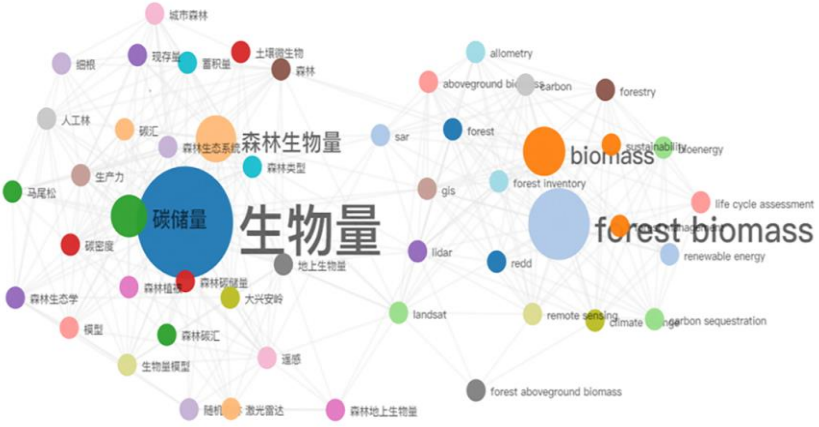

Fig. 2. High frequency keyword co-occurrence atlas

The larger circle of the keyword co-occurrence network knowledge map indicates the more frequent occurrence of the keyword in the literature, and the connecting line indicates the cooperation of different keyword domains. Combining Table 1 and Figure 2, we can see that the most popular indicators (or hotspots) in the research field around the most frequent biomass and forest biomass are carbon stock, biomass energy and carbon density, while Synthetic Aperture Radar (SAR), LiDar and biomass model construction are commonly used by researchers in remote sensing monitoring. In addition, climate change and forest ecology are also hot spots of interest.

4 The results were analyzed based on HistCite and WOS databases

HistCite software and SATI tool can do statistical analysis on the distribution of publication time, core authors, and journal distribution in the field of forest biomass, but it lacks visual analysis mapping. In addition, SATI is difficult to analyze the contribution of important literature in a certain field. Therefore, in this paper, we use the unique citation analysis function of HistCite software to statistically analyze the national distribution of forest biomass publications and classical literature.

HistCite has several common metrics:

1) Global Citation Score (GCS)

That is, the Citation Score displayed from the WOS database, shows the overall Citation times of the literature in academia, which is independent of the number of local literature.

2) Local Citation Score (LCS)

It shows the Citation times of the literature in the Local database. If the value is high, it indicates that the literature is the key literature in the research field of interest, and the total part is TLCS.

3) Local Cited References (LCR)

It shows the number of citations of the literature in the local database.

4) Cited references (CR)

It shows the number of references cited in the WOS database. The larger the value, the more likely it is to be a review article.

4.1 Analysis of published country

The 2900 valid documents obtained from WOS were imported into HistCite software for analysis, and the top 10 countries were ranked according to TLCS as shown in Table 2.

Table 2. A country with high frequency of publication in forest biomass

Number	*Country*	*Recs*	*TLCS*	*TGCS*
1	USA	903	3653	33161
2	Canada	288	1253	8596

Number	*Country*	*Recs*	*TLCS*	*TGCS*
3	China	439	1127	10390
4	UK	198	1093	11639
5	Germany	220	824	7549
6	France	135	769	6594
7	Italy	150	678	5505
8	Sweden	178	613	4716
9	Brazil	230	590	6192
10	Finland	170	526	4360

Among the top 10, the U.S. literature is No. 1 in terms of number (903/records) and TLCS (3653/time) ranking, and has a clear advantage over No. 2 Canada, indicating that the research in the field of forest biomass is in an absolute dominant position. The number of Chinese articles is 439, but the TLCS is only 1127 times, and the number of Canadian articles is only 288, and the TLCS value of 1253 times is similar to that of China, indicating that there are more independent articles in China, and the articles are less cited in related fields. Other countries such as UK, Germany and France are at the top of the number of publications and TLCS, indicating in-depth and high impact research in the field.

4.2 Analysis of chronological evolution

The unique HistCite literature citation chronogram function was used to locate the 30 most classic literature in the field of forest biomass between 2008 and 2021, and the literature details are listed in Table 3, and the citation relationships among the classic literature are shown in Figure 3.

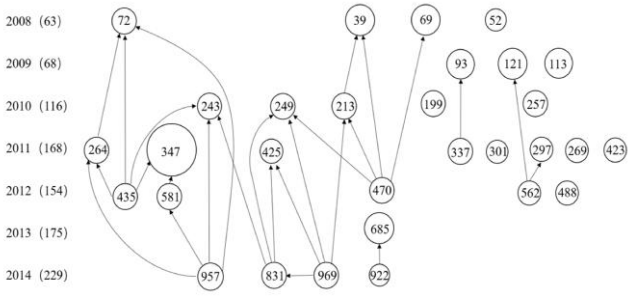

Fig. 3. The most valuable 30 references in the field of forest biomass

The literature citation chronology graph function locates classic literature by citation size. Using the Graph Maker function in HistCite software, the citation chronology graph of forest biomass literature, conditional on the LCS Count and with the number of nodes set to 30, also reveals the research history of the field. Figure 3 shows the publication dates, development process, and cross-citation relationships of the 30

most influential papers in the field of forest biomass research. The year of publication is shown on the left side of the figure, and the circles parallel to the year (the number in the circle is the literature number) represent the literature published in that year. The frequency of citation of the literature is analyzed according to the size of the circle, which is proportional to the LCS value, and the citation relationship between the literature is analyzed according to the arrows between the circles, and the specific nodal literature is shown in Table 3.

Table 3. Details of the 30 most valuable articles in the field of forest biomass

Number	*Serial Number*	*Node Information (Author/Year of publication/journal)*	*LCS*	*GCS*
1	39	Blackard JA, 2008, REMOTE SENS ENVIRON, V112, P1658	65	253
2	52	Wang XP, 2008, FOREST ECOL MANAG, V255, P4007	37	107
3	69	Boudreau J, 2008, REMOTE SENS ENVIRON, V112, P3876	65	249
4	72	Nogueira EM, 2008, FOREST ECOL MANAG, V256, P1853	52	160
5	93	Zhao KG, 2009, REMOTE SENS ENVIRON, V113, P182	54	229
6	113	Piao SL, 2009, NATURE, V458, P1009	44	770
7	121	Keith H, 2009, P NATL ACAD SCI USA, V106, P11635	55	450
8	199	Guo ZD, 2010, FOREST ECOL MANAG, V259, P1225	49	111
9	213	Powell SL, 2010, REMOTE SENS ENVIRON, V114, P1053	57	286
10	243	Djomo AN, 2010, FOREST ECOL MANAG, V260, P1873	45	122
11	249	Koch B, 2010, ISPRS J PHOTOGRAMM, V65, P581	55	224
12	257	Lucas R, 2010, IEEE J-STARS, V3, P576	43	171
13	264	Feldpausch TR, 2011, BIOGEOSCIENCES, V8, P1081	50	291
14	269	Thiffault E, 2011, ENVIRON REV, V19, P278	58	261
15	297	Mckechnie J, 2011, ENVIRON SCI TECHNOL, V45, P789	51	224

Number	*Serial Number*	*Node Information (Author/Year of publication/journal)*	*LCS*	*GCS*
16	301	Frazer GW, 2011, REMOTE SENS ENVIRON, V115, P636	47	165
17	337	Englhart S, 2011, REMOTE SENS ENVIRON, V115, P1260	62	164
18	347	Saatchi SS, 2011, P NATL ACAD SCI USA, V108, P9899	197	1129
19	423	Sandberg G, 2011, REMOTE SENS ENVIRON, V115, P2874	41	113
20	425	Clark ML, 2011, REMOTE SENS ENVIRON, V115, P2931	36	146
21	435	Feldpausch TR, 2012, BIOGEOSCIENCES, V9, P3381	56	257
22	470	Avitabile V, 2012, REMOTE SENS ENVIRON, V117, P366	35	137
23	488	Eckert S, 2012, REMOTE SENS-BASEL, V4, P810	36	131
24	562	Schulze ED, 2012, GCB BIOENERGY, V4, P611	36	176
25	581	Clark DB, 2012, J VEG SCI, V23, P1191	42	110
26	685	Shabani N, 2013, RENEW SUST ENERG REV, V23, P299	58	140
27	831	Laurin GV, 2014, ISPRS J PHOTOGRAMM, V89, P49	42	148
28	922	Carnbero C, 2014, RENEW SUST ENERG REV, V36, P62	47	187
29	957	Sileshi GW, 2014, FOREST ECOL MANAG, V329, P237	64	175
30	969	Fassnacht FE, 2014, REMOTE SENS ENVIRON, V154, P102	37	169

The bibliography reveals the progress of forest biomass research and the origin of classical literature. Carbon storage measurement of forest biomass [3] and environmental impact analysis [4] have always been a research hotspot. After the development of field measurement method and volume model estimation method in early forest biomass measurement, In 2008, J.A. Blackard et al [5]. used Moderate Resolution Imaging Spectroradiometer (MODIS) images to predict biomass and found a good correlation. The regression measurement of forest biomass by combining ground sample data with remote sensing optical data, liDAR, SAR and hyperspectral databegan to prevail on a large scale. The combination and comparison of multi-source remote sensing data, such as liDAR and hyperspectral data, is also a common method for estimating bio-

mass [6]. The forest biomass measurement model has been continuously improved. In 2012, Eldpausch T et al [7]. integrated tree height into the pan-tropical forest biomass estimation, and Sandra Eckert added the texture information of WorldView-2 satellite image of the northeast Madagascar rainforest into the model to improve the effect. With the expansion of the team of researchers and the change of technology, the standards of forest biomass, such as plot size, model and prediction method, are constantly updated and improved 8.

5 Conclusions

In this paper, with the help of bibliometric software SATI and HistCite, we analyze the progress and hotspots of forest biomass research from the perspectives of number and year of literature, keyword hotspots, country of publication and chronological evolution analysis by visualizing and quantifying the literature of forest biomass in CNKI database and WOS database, respectively. Thus, the advantages of combining the two software were verified.

The difficulty of having a large amount of literature and unknown key results for a study in an unfamiliar field often leads to lay people spending a lot of time on organizing literature data. In this paper, we combine the unique advantages of different bibliometric software and use SATI and HistCite analysis to sort out the overall overview, progress and hotspots of forest biomass research. This analysis can provide a convenient and clear way of thinking for future researchers who are struggling to organize the lengthy literature in a time-consuming manner.

References

1. Liu, Q.Y., Ye, Y. (2012) The Technology and Method of Bibliographic Information Mining and the Realization of SATI Software -- Taking Chinese and Foreign Library and Information Science as an example. Journal of Information Resources Management. 2(01):50-58.
2. Kang, G.G., Zhen, Z.F., Xu, Y.J., Chen, X. (2009) Research on Student satisfaction: Current situation, Evolution Path and Frontier -- Based on Web of Science database. Modern intelligence. 34(08):29-36+41.
3. PIAO, S. L., Fang, J.Y., Ciais, P., et al. (2009) The carbon balance of terrestrial ecosystems in China. Nature: International weekly journal of science.458(9).
4. Sassan, S., Nancy, H., Sandra, B., et al. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences of the United States of America.108(24).
5. Blackard, J., Finco, M.V., Helmer, E.H., et al. (2007) Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sensing of Environment, 112(4).
6. Laurin, G., Chen, Q., Jeremy, A., et al. (2014) Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. ISPRS Journal of Photogrammetry and Remote Sensing, 89.

7. Eldpausch, T., loyd, L., Lewis, S., et al. (2012) Tree height integrated into pantropical forest biomass estimates. Biogeosciences, 9(110).
8. Frazer, G., Magnussen, S., Wulder, M., et al. (2011) Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sensing of Environment.115(2).