Visual analysis of research results of multi-source nighttime light data from the audit perspective

CUI Hongwu1, CHEN Li2* and LI Xinfeng3

1 Shandong University of Science and Technology, College of Geomatics, Qingdao, China, 266590
2 Shandong University of Science and Technology, Auditing Department, Qingdao, China, 266590
3 Shandong Zhengyuan Digital City Construction Co., Ltd, Yantai, China, 264670

First Author’s E-mail: 201982020009@sdust.edu.cn

Abstract. Based on the literature data of multi-source nighttime light research in the core database of Web of science from 2010 to 2020, combined bibliometric analysis and audit analysis, analysed by CiteSpace software, the research results of multi-source nighttime light data are separately audited by integrated analysis, unit analysis, node correlation analysis and word frequency node classification. The main conclusions are: (1) After the initial and stable development of multi-source nighttime light data, it is in the rapid development period; (2) China and the United States are the main research countries; (3) "City" is a research hotspot, and "China" is the main research object; (4) The research fields mainly focus on remote sensing, environmental science, imaging science and photographic technology, geography physical, and others.

1. Introduction
With the development of society, audit becomes more important[1]. The audit conclusions use facts to dig out the internal connections of things and reveal the law of development of things to promote the healthy development of relevant research. This article takes the research results of nighttime lighting as a special case, and tries to audit the related results.

The differences in human activities and the natural phenomena make the brightness of nighttime lighting vary significantly between regions on the earth, which makes multi-source nighttime light data as an important data source for the study of socio-economic and human phenomena[2], and has many research results. Auditing the research results of multi-source nighttime light data will provide evidence and reference for the further research and development of multi-source nighttime light data. The nighttime light image can not only catch the artificial light in towns and villages[3], but also can obtain natural light sources[4] such as fishing boats[5], burning of natural gas[6] and forest fires[7]. Night light data is usually obtained by specific sensors mounted on satellites, which is an important sources. Using multi-source nighttime light data, can effectively realize the research on socio-economic indicators estimation[8], urbanization monitoring and evaluation[9], major event evaluation[10], ecological environment and health effects[11], marine lighting[12] and other aspects.
2. Data sources and research methods

2.1. Data source
This article selects four citation index databases, from 2010 to 2020, under the Web of Science core collection database. To improve the accuracy of retrieval, set the retrieval content as four key words as "nighttime light*", "night-time light*", "night light*", and "light* from nighttime". Using the wildcard characters, '*' and '" ', provided by WOS to control the search scope, and connect by the "or" connective, a total of 1,257 document records were retrieved, as shown in Figure 1.

![Figure 1. Number of related articles published and cited frequency.](image)

It can be seen from Figure 1 that from 2010 to the present, the number of articles on multi-source nighttime light data is increasing, and the growth trend is relatively stable. The number of articles reaches the maximum in 2019, with 284 articles published, but articles published in 2020 is 4 less than that in 2019. The number of citations is increasing, and it basically conforms to the increasing trend of quadratic polynomials. After calculation, its $R^2$ is 0.9861, which shows that this function has a very good fitting effect on the number of citations of multi-source night light data documents. By analyzing the increase in the number of papers and of citation frequency, it can be found that the research on multi-source nighttime light data is the hotspot.

The document record data is output in a plain text file format, and the record content selects the full record and cited references. The number of literature publications is audited on an annual basis. After the preprocessing of the literature records, a total of 1,247 literature records participated in the audit, including 1,132 journals, 94 conferences, 22 reviews, 1 duplicate, and finally compared to the original, 10 pieces of data are discarded.

2.2. Research methods
Combining bibliometric analysis and audit analysis, conduct integrated audits through the number of research subjects, scientific cooperation network analysis for unit and node correlation audits, cluster analysis for word node frequency classification audits, supplemented by CiteSpace software, applications audit analysis explores the development laws of things in the corresponding fields, and effectively supervises the development and changes of things, so as to correctly reflect the nature and laws of things. When the generated graph is audited and interpreted, it mainly focuses on the unique forms of high-frequency nodes, keyword clusters, nodes with high intermediary centrality, and citation year rings. Then auditing and analysing the current research results of nighttime light data from various aspects.

3. Research results audit

3.1. Integrated audit
A full-domain audit analysis of the quantitative changes in the three dimensions of relevant countries,
institutions, and authors, is called the integrated audit. The integrated audit of 1247 articles on multi-source nighttime light was conducted to obtain the change law of relevant results, as shown in Figure 2.

Through audit observations, the development of research on multi-source nighttime light data can be divided into three main stages: first stage (2010-2013) is the initial stage of its research, which development is relatively slow. At this time, the number of individuals conducting research in this field is relatively small; The second stage is from 2014 to 2016. In 2016, the number of institutions is nearly twice that of 2013, and there is a certain increase in the number of countries and authors.

Compared with the first stage, more researchers and institutions realize the value of multi-source night light data. The third stage (2017-2020) is the rapid development stage. The number of countries, institutions and authors increase significantly. It is worth noting that the maximum of those three does not come from the same year. The maximum of authors and institutions is from 2019, while the maximum number of countries is from 2020, and the total number of those three decreases in 2020.

3.2. Unit audit
Unit audit refers to the auditing of individual original authors in the field of research results, which mainly conducts audit analysis through two aspects: author cooperation network and author co-citation network[19]. The author cooperation network is a microscopic scientific cooperation network analysis that can show the cooperative network relationship between scholars who study multi-source nighttime light data. However, the author co-citation network analysis extracts the author's correlation from the entire document bibliography to analyze the distribution of highly cited authors to identify influential scholars in the field.

[Figure 3. Author's cooperation network time zone chart.

In Figure 3 (Author's cooperation network time zone chart), the size of the node corresponds to the amount of articles published by the author, that is, how much their respective research results in the field of multi-source nighttime light data. It is obvious that 2014, the first year of the second stage of the development of multi-source nighttime light data research, there are significantly more highly productive authors than in other years. From the whole, it can be seen that the nodes of authors such as Bailang Yu, Jianping Wu, Kaikai Shi, Zuoqi Chen, Xi Li, Elvidge CD, Chunyang He are significantly larger than others, and the amount of articles published is relatively large. Those authors have their own unique contribution and special influence. At the same time, it can be found in the audit of the connecting lines of each node that a complex interactive network is formed between the authors, and there is a direct or indirect relation between a certain node and any other node.

3.3. Node correlation audit
Node correlation audit is a macroscopic scientific cooperation network analysis audit, which can show the cooperation network relationship between the countries where scholars are located. On this basis, through the timeline method, it is clear that the time for each country (region) to conduct multi-source...
nighttime light data related research. At the same time, the size of the node shows the citation of the multi-source nighttime light data literature of each country (region). The cited status reflects to a certain extent that the degree of influence of various countries on the field of multi-source nighttime light data.

Figure 4. Timeline analysis of country (region) cooperation network.

It can be seen from Figure 4 that the research on multi-source nighttime light data in the United States, Japan, Australia and Germany started relatively early, followed by countries such as China, the United Kingdom, and Italy, and countries such as Hungary, Uruguay, and Ethiopia have just begun research in this field. It can be seen from the degree of nodes that the three countries of the United States (degree of 42), China (degree of 34), and the United Kingdom (degree of 30) have extensive cooperate with other countries who are the top three in the ranking. From the color of the nodes, the citation year rings of the nodes in the United States, Japan and China all have purple rings, indicating that they have high intermediary centrality and are important countries that connect different countries for multi-source nighttime light data research. From the size of the node, although China's research on nighttime light data is relatively late, the number of Chinese articles (716 articles) is much more than the number of American articles (378 articles), and China is ranking first in the world. This indicates that the study of multi-source nighttime light data has a rapid development in China, is widely accepted by Chinese authors, and China plays an increasingly important role in the research and development of global multi-source nighttime light data.

3.4. Word frequency node classification audit

Word frequency refers to the times of occurrences of word in the analyzed document. Word frequency analysis, an analysis method studying the frequency of letters or groups of characters in articles to study the development direction and research hotspots. In this article, the frequency of the keywords is displayed in word clouds, and the frequency of the words appearing in the literature related to multi-source nighttime lighting is determined according to the size of the vocabulary.

It can be seen intuitively in Figure 5 of the word cloud map, from 2011 to 2020, the keyword "China" has occupied a considerable position in the word cloud, which shows that China has always been the main research subject of multi-source nighttime light data. In the first stage of the development of multi-source nighttime lighting, the research hotspots are light pollution, CO₂ emissions, land coverage, population, urbanization, etc. In the second stage, the research hotspots are population, urbanization, electricity consumption, economic activities, urban expansion, etc. In the third stage, the research hotspots are urbanization, impervious surfaces, economic activities, energy consumption, and ecosystem services. Through comparative audit analysis of research directions at different stages, it is found that "city" is the main research meta-domain of multi-source nighttime light, and most research is about "city". Considering the research directions and categories provided by WOS, the research fields of multi-source nighttime light data mainly focus on remote sensing, environmental science, imaging science photography technology and geography physical. At the same time, according to the extracted keywords of relevant research, it is found that the research field of multi-source nighttime light data is diversified.
It also appears in fisheries, offshore natural gas, natural disasters, human diseases, and social war cost evaluations, which are not related to each other. However, this research plays an irreplaceable role in them.

\[
Q_{\text{Modularity}} = \frac{1}{2m} \sum (a_{ij} - p_{ij})\sigma(C_i, C_j)
\]

In formula (1), \( a_{ij} \) is the adjacency matrix in the actual network; \( p_{ij} \) is the expected value of the number of edges connected between node \( i \) and node \( j \) in the null model; \( C_i \) and \( C_j \) represent the communities of node \( i \) and node \( j \) in the network. If \( i \) and \( j \) belong to the same community, then \( \sigma = 1 \); otherwise, \( \sigma = 0 \).

Keyword clustering can easily find similar keywords and perform appropriate node classification. The clustering effect can be evaluated by the \( Q_{\text{Modularity}} \) values. The \( Q_{\text{Modularity}} \) value is 0.4056, when it is greater than 0.3, it means that the obtained network community structure is significant. As shown in Figure 6, the 8 labels can be divided into three categories according to the word class: (1) Research data: artificial light and of-interest data. The main night light data currently used mainly include other small-scale or non-open source data such as DMSP-OLS, NPP-VIIRS, Luojia 1-01. (2) Research method: unmixing method. Data preprocessing mainly uses the invariant target area method, which performs saturation correction, continuity correction and noise removal on night light data. In practical applications, various methods such as threshold method, machine learning algorithm, and multi-source data fusion are used to solve practical problems. (3) Research direction: GDP, CO2 emission, urban heat island, ecosystem services, ntl urban index. The application of multi-source nighttime light data is extremely wide, covering various aspects such as city scope, population, GDP, environment, ecology, electricity, disasters, fishery, oil and gas, health and material reserves.

4. Audit report
Multi-source nighttime light data has been developed rapidly in the past 10 years, the data volume has increased sharply, and the research direction was diversified. Audit research shows that the current integrated research on multi-source nighttime light data is still in a hot state, and related unit continue to increase. Different stages of nighttime light development has different hotspots. The first stage is the initial stage. In this stage, the development of related research is slow, and the number of researchers, institutions and countries are relatively small. The preliminary research of light pollution, CO2 emissions, land coverage, population and urbanization research direction start. The second stage is the steady development stage. More and more researchers have begun to research it, and the number of research institutions and countries has also increased. At this time, the research hotspots are closely related to
people's daily life, such as population, urbanization, electricity consumption and economic activities. The third stage is the rapid development stage. The increased number of research institutions in the previous stage has made more researchers realize the value of multi-source nighttime lighting and actively participate in related research. The research direction is diversified, which mainly focus on urbanization, impervious surface, economic activities, energy consumption and ecosystem services. According to the research direction of nighttime lighting in the past ten years, the main research object is "city" and "China". So far, the unit research of multi-source nighttime light data mainly focuses on remote sensing, environmental science, imaging science and photography technology, geography physical, and also involves fishery, marine natural gas, natural disasters, human diseases and social war cost assessment. Based on its own advantages, "city" will still be the main research meta-domain of multi-source nighttime light, and this research will continue to expand into other fields.

5. Conclusion
On the basis of bibliometrics and audit, this paper analyzes the research results of multi-source nighttime light data by using 1247 relevant articles from web of science. This paper shows the overall research status, development trend and research hotspots. At present, the research on the unit of multi-source nighttime light data is still hot, and related authors, institutions, and countries are constantly increasing. The number of researchers increased significantly in 2014. According to the analysis of node correlation, it is found that most authors have close team cooperation, but there is less communication between teams. The research of multi-source nighttime light data is mainly carried out by China and the United States. The United States has developed earlier than China and has great influence, but the number of Chinese articles was much more than the number of American articles, and China occupies an important position on multi-source nighttime light research. Based on the relevant keywords and word frequencies of multi-source nighttime light data from 2010 to 2020, through word cloud and cluster analysis, it is found that the researched node classification hotspots include carbon emissions, urbanization, human footprint, ecology, population, energy consumption, light pollution and land use, and the research directions is diversified.

Night lights give the night new vitality and value. With audit methods and bibliometric methods, researchers have found that multi-source nighttime light data plays an important role in socio-economic indicator estimation, urban monitoring and evaluation, major event evaluation, ecological environment and health effects and Ocean study. However, some problems have also been discovered from the current audit analysis and research of multi-source nighttime light: (1) The aggregation degree of node points is high, but the correlation degree between nodes is relatively low; (2) The time and progress of unit's research are significantly different, which result in an unbalanced results in integrated audit.

Acknowledgments
I would like to express my gratitude to all those who helped me during the writing of this thesis. I gratefully acknowledge the help of my supervisor, Chen Li, who has offered me valuable suggestions in the academic studies. At the same time, thanks to engineer Li Xinfeng for his guidance on my professional knowledge. Without their patient instruction, the completion of this thesis would not have been possible.

References
[1] Cai, C., Cai, L., Zhu R. (2011) Ten Considerations on Promoting the Innovation and Development of Performance Audit Comprehensively in China. Auditing Research, (04): 32–38.
[2] Yu B., Wang C., Gong W., Chen Z., Shi K., Wu B., Hong Y., Li Q., Wu J. (2021) Nighttime light remote sensing and urban studies: Data, methods, applications, and prospects. Journal of Remote Sensing, 25:342-364.
[3] JIANG W, HE G, LONG T et al., (2018) Potentiality of Using Luojia 1-01 Nighttime Light Imagery to Investigate Artificial Light Pollution. Sensors, 18: 2900.
[4] Li D., Li X., (2015) An Overview on Data Mining of Nighttime Light Remote Sensing. Acta Geod Cartogr Sin, 44: 591–601.
[5] ELVIDGE C D, GHOSH T, BAUGH K et al., (2018) Rating the Effectiveness of Fishery Closures With Visible Infrared Imaging Radiometer Suite Boat Detection Data. Frontiers in Marine Science, 5: 132.
[6] LU W, LIU Y, WANG J et al., (2020) Global Proliferation of Offshore Gas Flaring Areas. Journal of Maps, 16: 396–404.
[7] BERESFORD A E, DONALD P F, BUCHANAN G M. (2020) Repeatable and Standardised Monitoring of Threats to Key Biodiversity Areas in Africa Using Google Earth Engine. Ecological Indicators, 109: 105763.
[8] ELVIDGE C D, BAUGH K E, KIHN E A et al., (1997) Relation between Satellite Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption. International Journal of Remote Sensing, 18: 1373–1379.
[9] YANG C, LI Q, ZHAO T et al., (2019) Detecting Spatiotemporal Features and Rationalities of Urban Expansions within the Guangdong–Hong Kong–Macau Greater Bay Area of China from 1987 to 2017 Using Time-Series Landsat Images and Socioeconomic Data. Remote Sensing, 11: 2215.
[10] CEOLA S, LAIO F, MONTANARI A. (2014) Satellite Nighttime Lights Reveal Increasing Human Exposure to Floods Worldwide. Geophysical Research Letters, 41: 7184–7190.
[11] BAUER S E, WAGNER S E, BURCH J et al., (2013) A Case-Referent Study: Light at Night and Breast Cancer Risk in Georgia. International Journal of Health Geographics, 12: 23.
[12] WALUDA C M, GRIFFITHS H J, RODHOUSE P G. (2008) Remotely sensed spatial dynamics of the Illex argentinus fishery, Southwest Atlantic. Fisheries Research, 91: 196–202.
[13] Liu L., (2005) Mapping knowledge domains of research with document co-citation. Studies in Science of Science, 2005(02): 155–159.
[14] WU L, WANG W, JING P et al., (2020) Travel Mode Choice and Their Impacts on Environment—a Literature Review Based on Bibliometric and Content Analysis, 2000–2018. Journal of Cleaner Production, 249: 119391.
[15] Zhang B, (2014) Structure and Evolution of Co-word Network: Concept and Research Review. Journal of Intelligence, 249: 119391.
[16] ESTABROOKS C A, DERKSEN L, WINTHER C et al., (2004) The Intellectual Structure and Substance of the Knowledge Utilization Field: A Longitudinal Author Co-Citation Analysis, 1945 to 2004. Implementation Science, 3(1): 49.
[17] SUN L, XU X, YANG Y et al., (2020) Knowledge Mapping of Supply Chain Risk Research Based on CiteSpace. Computational Intelligence, 36: 1686–1703.
[18] GUO Y., HUANG Z. , GUO J et al., (2019) Bibliometric Analysis on Smart Cities Research. Sustainability, 11: 3606.
[19] YE Y, GE Y. (2019) A Bibliometric Analysis of Inventory Management Research Based on Knowledge Mapping. The Electronic Library, 37: 127–154.