



Retrieval strategy and possible explanations for the abnormal growth of research publications: re-evaluating a bibliometric analysis of climate change

Fang Liu¹

Received: 30 September 2021 / Accepted: 28 September 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

Climate change and global warming have attracted more and more attention from countries all over the world. A recent study published in *Scientometrics* has evaluated the changing dynamics of climate change-related research publications via a bibliometric analysis and further probed the relationship between climate change-related research output and carbon dioxide emissions. We try to re-evaluate the bibliometric analysis section of the mentioned study and provide three improvement suggestions related to data source, search field and search query respectively. Besides, some new explanations have also been offered for the abnormal increase of research outputs indexed by Web of Science Core Collection in specific years such as 1991. These suggestions and explanations will provide important references for future various bibliometric analyses and research evaluation.

Keywords Research evaluation · Bibliometric analysis · Web of Science Core Collection · Topic search

Introduction

With the increasing global concern over climate change, carbon dioxide emissions have become a key concern for many countries (Chen et al., 2020). A recent interesting study published in *Scientometrics* has evaluated the changing dynamics of climate change-related research publications via a bibliometric analysis and further probed the relationship between climate change-related research output and carbon dioxide emissions (de Gouveia & Inglesi-Lotz, 2021). Although it is an impressive study, we believe there is room for improvement in the bibliometric analysis section of that research.

For example, it is not easy for us to replicate the bibliometric analysis in de Gouveia and Inglesi-Lotz's study since the data source and search field used in their study have not been pointed out clearly. Besides, the abnormal growth of research publications in years such as 1991 has not been explained reasonably, which may mislead many readers. In this paper,

✉ Fang Liu
liu_fang2014@163.com

¹ School of Accounting, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang, China

we try to re-evaluate the bibliometric analysis section of de Gouveia and Inglesi-Lotz's study. Suggestions for improving the retrieval strategy will be made and some new explanations will be provided for the abnormal increase of research outputs. These suggestions and explanations will provide important references for future various bibliometric analyses and research evaluation.

Retrieval strategy

Data source

A clear description of the used data source is essential for conducting a bibliometric analysis. When using Web of Science Core Collection as the data source, the used sub-datasets and their corresponding coverage years should be specified clearly (Hu et al., 2020; Liu, 2019). However, for de Gouveia and Inglesi-Lotz (2021), this key point is confusing to readers.

In the methodology and data analysis section, de Gouveia and Inglesi-Lotz (2021) mentioned that "The starting date of 1956 for the scientometric part of the analysis was chosen because it is the starting date for the Clarivate Analytics database". Web of Science Core Collection covered eight citation indexes and two chemical indexes: Science Citation Index Expanded (1900–present), Social Sciences Citation Index (1900–present), Arts & Humanities Citation Index (1975–present), Emerging Sources Citation Index (2005–present), Conference Proceedings Citation Index (1990–present, including Science and Social Science & Humanities), Book Citation Index (2005–present, including Science and Social Science & Humanities), Current Chemical Reactions (1985–present) and Index Chemicus (1993–present) (Birkle et al., 2020; Liu, 2019).¹ Many institutes only subscribe a customized part of the whole core collection. However, similar to many previous studies as mentioned in Liu (2019), the detailed subscription information was not provided in their study. Thus, we assume de Gouveia and Inglesi-Lotz's institute has subscribed the complete datasets of Web of Science Core Collection from the year 1956.

de Gouveia and Inglesi-Lotz (2021) also mentioned the data source used in their study in the introduction section as following. It seems they have used all the eight citation indexes and two chemical indexes of Web of Science Core Collection as the data source for their study.

The top 50 countries will be chosen via ranking of the information in the first part of the study with data derived from the Clarivate Analytics Core Collection which consists of the—Science Citation Index Expanded, Social Sciences Citation Index, Arts and Humanities Citation Index, Emerging Sources Citation Index, Conference Proceedings Citation Index, Book Citation Index, Current Chemical Reactions and Index Chemicus.

However, in the methodology and data analysis section, de Gouveia and Inglesi-Lotz (2021) further stated the selection of data source as following. According to the latest 2020 Journal Citation Reports, 20,932 journals from four journal citation indexes are covered. Based on the following description, it seems that de Gouveia and Inglesi-Lotz (2021) has only used four journal citation indexes of Web of Science Core Collection as the data source which is in contradiction with the above inference.

¹ For more information, please refer to http://images.webofknowledge.com/WOKRS535R111/help/WOSH_p_database.html.

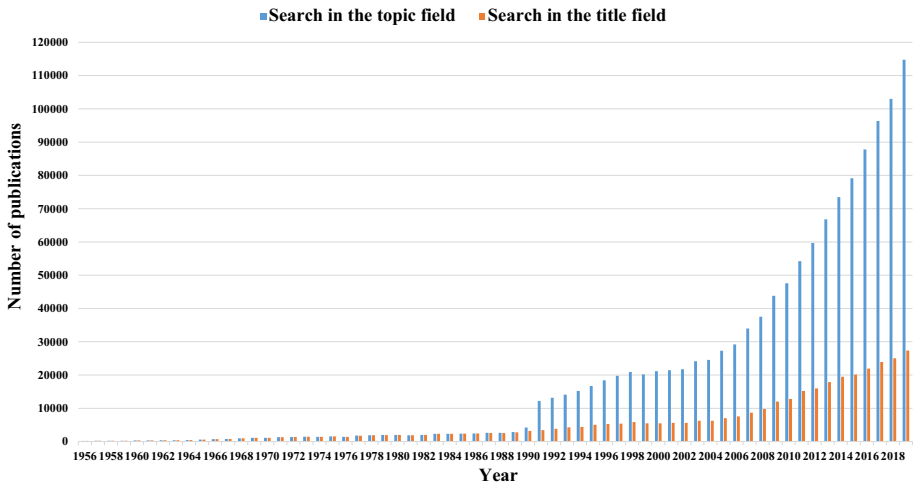


Fig. 1 Dynamics of climate change-related publications in Web of Science Core Collection (topic/title search)

The Core collection is a collection of over 21,000 peer-reviewed, high-quality scholarly journals published worldwide in over 250 social science, humanities and science disciplines (Matthews, 2020). The selection process to make this collection must have the basic principles of selectivity, objectivity and collection dynamics. There is a single set of 28 criteria which focuses on quality and impact when evaluating every journal, hence making the use of the core collection for this analysis is an appropriate one (Clarivate, 2020). This selection was made due to the availability of publication data necessary and because it provides great variability in terms of the types of economies that will be tested in this study.

Luckily, the total numbers of papers in Web of Science Core Collection for the years 1960, 1980, 2000 and 2019 have been provided in Table 3 of de Gouveia and Inglesi-Lotz (2021). By examining the total numbers of publications for these four specific years, we confirm that not only four journal citation indexes but also other citation indexes and chemical indexes were included in their analysis (Data accessed on September 19, 2021; Data values retrieved at different times may differ slightly).

Search field

Some climate change-related keywords were provided in the study by de Gouveia and Inglesi-Lotz (2021). However, it is not clear which field is used to retrieve related records in Web of Science Core Collection for de Gouveia and Inglesi-Lotz’s study. Topic and title are two typical fields used in literature retrieval for bibliometric analysis.² By utilizing the climate change-related keywords and searching in the topic and title fields respectively, Fig. 1 shows the dynamics of climate change-related records. By comparing Fig. 1 with

² Web of Science Core Collection began to support the search in abstract, author keywords and keywords plus fields recently (Liu, 2021).

Fig. 1 in the study by de Gouveia and Inglesi-Lotz (2021), it seems that the topic field was used in de Gouveia and Inglesi-Lotz's study for literature retrieval.

Search query

In the practice of literature retrieval, the selection of appropriate keywords is a very challenging task. There exists a compromise between recall and precision. For the study of de Gouveia and Inglesi-Lotz (2021), about twenty keywords were used to search climate change-related papers. One suggestion is replacing the expression "greenhouse gasses" with "greenhouse gas*". The new expression with a wildcard will contain records with "greenhouse gas"/"greenhouse gases" besides "greenhouse gasses". By using the following updated search query, 5538 new Web of Science Core Collection indexed records published during 1956–2019 can be retrieved. Luckily, from the net increase rate perspective, the small increase (about 0.4%) of the number of new records may not influence the main conclusions in the econometric analysis section of de Gouveia and Inglesi-Lotz (2021). However, there is still room for improvement in the updated search query.

Original search query:

```
TS=("Climate change" OR "global warming" OR "CO2" OR "emissions" OR "carbon dioxide" OR "carbon tax" OR "ETS" OR "emissions trading system" OR "greenhouse gasses" OR "GHG" OR "fossil fuels" OR "global average temperature" OR "sea-level rise" OR "renewable energy" OR "COP" OR "UNFCCC" OR "INDC" OR "IPCC" OR "PPM" OR "Methane" OR "pre-industrial levels of carbon dioxide")
```

Updated search query:

```
TS=("Climate change" OR "global warming" OR "CO2" OR "emissions" OR "carbon dioxide" OR "carbon tax" OR "ETS" OR "emissions trading system" OR "greenhouse gas*" OR "GHG" OR "fossil fuels" OR "global average temperature" OR "sea-level rise" OR "renewable energy" OR "COP" OR "UNFCCC" OR "INDC" OR "IPCC" OR "PPM" OR "Methane" OR "pre-industrial levels of carbon dioxide")
```

Possible explanations for the abnormal growth of research publications

de Gouveia and Inglesi-Lotz (2021) observed a rapid increase of climate change-related records from both an absolute number and a relative share perspective. They also tried to give some possible explanations for the rise of research publications. However, more convincing reasons are needed to explain the abnormal growth of research publications for some specific years in various bibliometric analyses including the study conducted by de Gouveia and Inglesi-Lotz. Unfortunately, these reasons are often ignored by many bibliometric studies.

Comprehensiveness of abstract/author keywords/keywords plus information in Web of Science Core Collection

By using the original search query of de Gouveia and Inglesi-Lotz (2021) in the abstract/author keywords/keywords plus fields, records with climate change-related keywords in corresponding fields in Web of Science Core Collection will be retrieved. Figure 2

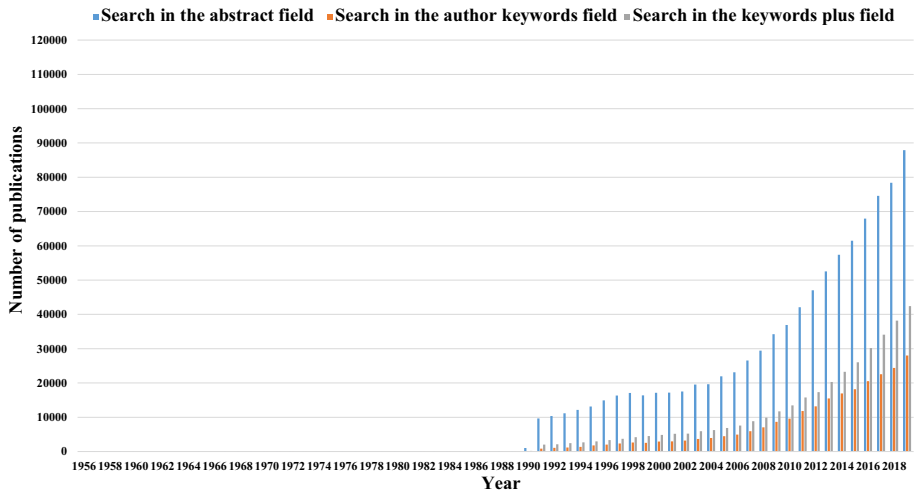


Fig. 2 Dynamics of climate change-related publications in Web of Science Core Collection (abstract/author keywords/keywords plus search)

demonstrates the dynamics of records with climate change-related keywords in abstract/author keywords/keywords plus fields respectively. Please note that the topic search is a combination of search in the title, abstract, author keywords and keywords plus fields.

According to Fig. 1, the numbers of climate change-related records retrieved by the topic search and title search were almost identical before 1990. However, for records retrieved by the topic search, a sudden increase happened for the year 1991. Based on our new data, the number of publications rose from 4206 in 1990 to 12,199 in 1991. However, the number of publications retrieved by the title search only rose slightly from 3208 in 1990 to 3453 in 1991. Besides, for publication years after 1991, the numbers of publications retrieved by the title search increased relatively flat than that retrieved by the topic search. Figure 1 demonstrates the changing gaps between the numbers of records retrieved by the topic search and title search.

Based on Fig. 2, climate change-related publications retrieved by the abstract/author keywords/keywords plus fields were very rare before 1990. This can be explained by a pioneer empirical examination by Liu (2021). Liu (2021) has found that a very high proportion of journal papers published before 1991 lack abstract/author keywords/keywords plus information in Web of Science Core Collection. Since the number of publications retrieved by the title search only increased slightly in 1991, we can conclude that the comprehensiveness of abstract/author keywords/keywords plus information in Web of Science Core Collection is a major cause for the sudden increase of the number of climate change-related publications in 1991 as demonstrated in the study of de Gouveia and Inglesi-Lotz (2021).

Inclusion of a new sub-database

As shown above, the full collection of Web of Science Core Collection has eight citation indexes and two chemical indexes (Birkle et al., 2020; Liu, 2019). Each sub-database may have different coverage years. For example, if we compare the yearly outputs of Web of Science Core Collection indexed publications around the year 1990, we should notice that

Conference Proceedings Citation Index (including Science and Social Science & Humanities) start collecting data of conference papers on a large scale around the year 1990. Abnormal growth of research publications around 1990 may happen due to the inclusion of conference papers. The subscription package of Web of Science Core Collection may be different for different institutional users (including different sub-datasets and different coverage years) (Liu, 2019). Theoretically speaking, the abnormal growth of research publications caused by the inclusion of a new sub-database may happen for any year. For less experienced users, this is a particular concern.

Anomalous growth of source titles in the sub-database

de Gouveia and Inglesi-Lotz (2021) mentioned that more academic publishing platforms that can publish research results also increase the number of climate change-related papers. We assent to this point. We would like to further point out that the abnormal growth of the number of journals/conferences/books in the database also leads to the abnormal growth of the number of publications (Michels & Schmoch, 2012). One example is the inclusion of about 1600 new regional journals in Web of Science during 2007–2009 (Liu, 2017; Testa, 2011; Vanderstraeten & Vandermoere, 2021).

Conclusion

Web of Science Core Collection is widely used as a basic data source in bibliometric analysis and research evaluation (Li et al., 2018; Zhu & Liu, 2020). At the same time, many scholars have begun to raise concerns about bibliometric analysis and the appropriate use of databases (González-Alcaide, 2021; Hicks et al., 2015; Liu, 2019, 2021; Liu et al., 2020). For de Gouveia and Inglesi-Lotz (2021), we suggest the authors details the search strategy to retrieve climate change-related literature. A key point that is easily overlooked is to clarify the used sub-datasets of Web of Science Core Collection and their corresponding coverage years (Liu, 2019).

In general, abnormal increases of the numbers of research outputs are rare or easy to explain. For example, the explosion of COVID-19 related publications in 2020 is due to the COVID-19 pandemic (Cai et al., 2021). If there is an abnormal increase in research outputs and it is not easy to find a reasonable explanation, then we may need to look at the problem of the data source. Comprehensiveness of abstract/author keywords/keywords plus information in Web of Science Core Collection is a major cause for the abnormal increases in research outputs around the year 1990 (Liu, 2021). Inclusion of a new sub-database and anomalous growth of source titles in the sub-database are also two possible causes for abnormal increases of research outputs in some specific years.

Acknowledgements This research was funded by the National Natural Science Foundation of China (71904168) and the Humanities and Social Sciences Foundation of the Ministry of Education of China (19YJC630101) and Zhejiang Provincial Natural Science Foundation of China (LQ18G010005). In addition, we thank the management of the editors and the anonymous reviewers for valuable comments and suggestions which helped us to improve this work. We are responsible for any errors.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

References

- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. https://doi.org/10.1162/qss_a_00018
- Cai, X., Fry, C. V., & Wagner, C. S. (2021). International collaboration during the COVID-19 crisis: Autumn 2020 developments. *Scientometrics*, 126(4), 3683–3692. <https://doi.org/10.1007/s11192-021-03873-7>
- Chen, J., Gao, M., Mangla, S. K., Song, M., & Wen, J. (2020). Effects of technological changes on China's carbon emissions. *Technological Forecasting and Social Change*, 153, 119938. <https://doi.org/10.1016/j.techfore.2020.119938>
- Clarivate. (2020). Editorial selection process. <https://clarivate.com/webofsciencegroup/solutions/editorial/>
- de Gouveia, M., & Inglesi-Lotz, R. (2021). Examining the relationship between climate change-related research output and CO₂ emissions. *Scientometrics*, 126(11), 9069–9111. <https://doi.org/10.1007/s11192-021-04148-x>
- González-Alcaide, G. (2021). Bibliometric studies outside the information science and library science field: Uncontainable or uncontrollable? *Scientometrics*, 126(8), 6837–6870. <https://doi.org/10.1007/s11192-021-04061-3>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>
- Hu, G., Wang, L., Ni, R., & Liu, W. (2020). Which h-index? An exploration within the Web of Science. *Scientometrics*, 123(3), 1225–1233. <https://doi.org/10.1007/s11192-020-03425-5>
- Li, K., Rollins, J., & Yan, E. (2018). Web of Science use in published research and review papers 1997–2017: A selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, 115(1), 1–20. <https://doi.org/10.1007/s11192-017-2622-5>
- Liu, W. (2017). The changing role of non-English papers in scholarly communication: Evidence from Web of Science's three journal citation indexes. *Learned Publishing*, 30(2), 115–123. <https://doi.org/10.1002/leap.1089>
- Liu, W. (2019). The data source of this study is Web of science core collection? Not enough. *Scientometrics*, 121(3), 1815–1824. <https://doi.org/10.1007/s11192-019-03238-1>
- Liu, W. (2021). Caveats for the use of Web of Science Core Collection in old literature retrieval and historical bibliometric analysis. *Technological Forecasting and Social Change*. <https://doi.org/10.1016/j.techfore.2021.121023>
- Liu, W., Tang, L., & Hu, G. (2020). Funding information in Web of Science: An updated overview. *Scientometrics*, 122(3), 1509–1524. <https://doi.org/10.1007/s11192-020-03362-3>
- Matthews, T. (2020). Web of Science platform: Web of Science Core Collection. <https://clarivate.libguides.com/webofscienceplatform/woscc>
- Michels, C., & Schmoch, U. (2012). The growth of science and database coverage. *Scientometrics*, 93(3), 831–846. <https://doi.org/10.1007/s11192-012-0732-7>
- Testa, J. (2011). The globalization of Web of Science: 2005–2010. <http://wokinfo.com/media/pdf/globalwos-essay>
- Vanderstraeten, R., & Vandermoere, F. (2021). Inequalities in the growth of Web of Science. *Scientometrics*, 126(10), 8635–8651. <https://doi.org/10.1007/s11192-021-04143-2>
- Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*, 123(1), 321–335. <https://doi.org/10.1007/s11192-020-03387-8>