Scientific review of climate science: a bibliometric analysis of trends

Maria Goncharova, and *Elizaveta* Sokolova*

Ural State University of Economics, Yekaterinburg, Russia

Abstract. This research analyses the metadata of 15550 climatology publications indexed in the OpenAlex data catalogue of the types 'article' (95%), 'book' (3%) and 'book-chapter' (2%). The research covers the period from 1851 to the present. Bibliometrix and VOSviewer software tools were used for data processing and analysis, and Scimago Graphica was used to visualise the results. As a result of the research, new data on the periods of authors' publication activity by subject, supported by the systematisation of the main scientific achievements and changes of those times, were obtained. The structure of access to publications in climatology and its impact on the attention of the scientific community, thematic diversity and the most relevant current research areas were also analysed.

1 Introduction

The increasing relevance has been noted every year in the global dialog on climate change due to the intensification of the greenhouse effect. Climate, as a physiographic attribute of the environment, influences economic activity, including agrarian specialisation, industrial location and transport. Awareness of meteorological and climatological fundamentals is necessary to understand the atmospheric processes that affect the ecology of an enterprise or region. Climatology, as a section of meteorology, studies the regularities of climatic formation, geographical distribution and historical dynamics.

The first ideas about climate were formed in ancient Greece. In the 17th-18th centuries, descriptions of climates based on meteorological observations appeared. E. Halley, J. Hadley and M. V. Lomonosov proposed ideas about the influence of atmospheric circulation on climate. At the beginning of the 19th century, A. Humboldt started systematic description of climates and creation of climatic maps. The task of climate forecasting for the next decades and centuries, complicated by the need to take into account the growing anthropogenic influences, remains one of the key tasks.

Previously, similar studies based on bibliometric analysis methods have been conducted. There are papers that use bibliometric and scientometric methods to analyse publications on climate change [7, 8, 9], extreme weather events [10], climatology of individual cities, regions and countries [11, 12, 13], and other highly specialised topics. However, studies that analyze climate science as a knowledge field in its diversity without territorial and sectoral limitations were not found.

* Corresponding author: sokolova_ev@usue.ru

2 Materials and methods

The study is based on metadata analysis of 15550 publications on climatology (containing various forms of the word 'climatology' in the title) of the types 'article' (95%), 'book' (3%) and 'book-chapter' (2%) indexed in the OpenAlex data catalogue. The time range of the analysis covers the entire available period from 1851 to the present. The vast majority of publications (96%) are written in English. Italian accounts for 1.1%, and French, Portuguese, Spanish and Chinese are even less represented, each accounting for 0.2% of the total number of publications.

The current access structure of climate science publications, based on the sample analysed, has the following distribution by type of access: 65% of publications are closed access, 13% are hybrid publications, 4% have green status, 10% have bronze status, and 7% of publications are gold access.

The choice of data source for this study was driven by several key factors confirmed by recent research. According to a study by Culbert J. and his colleagues, which analyzed a metadata set of 16.8 million papers published in the OpenAlex, Web of Science, and Scopus databases, OpenAlex demonstrated an average number of source citations and internal coverage quite comparable to that reported by recognized databases [1]. In a separate study by Alperin J.P. and colleagues, OpenAlex was also compared with Scopus, and the results showed that OpenAlex and Scopus presented comparable results in terms of Open Access status, with OpenAlex identifying more papers in the 'Hybrid' category [2]. Simard M.A. and colleagues also found that most of the Open Access journals indexed in DOAJ are also listed in OpenAlex, unlike Web of Science and Scopus [3]. Thus, OpenAlex has greater coverage, which is critical for bibliometric research. Another important advantage of OpenAlex, why it was chosen for this study, is its openness and the ability to upload metadata of an unlimited number of publications, which is not available in other international databases. The platform's commitment to full openness and free access means a move towards more transparent and accessible scientific data, which benefits the research community as a whole [4].

Bibliometrix and VOSviewer software tools were used for data processing and analysis in this study, while Scimago Graphica was chosen to visualise the results. Bibliometrix and VOSviewer are classic tools for bibliometric research. Scimago Graphica, on the other hand, was chosen for its flexible specification, ability to generate interactive visualisations, and support for a variety of data types [5].

3 Results

Although the dataset analysed includes publications dating from 1851 onwards [Abich], in the early years the publication of research on climatology was erratic, with only a few papers published each year. During this period, W. Keppen developed a classification of climates, and J. Hahn compiled a monograph, A Manual of Climatology. The first significant increase in the number of publications was in 1896 and continued until 1913, at which time an average of 43 papers were published annually (whereas only 29 papers were published in the entire period before). This increase is mainly due to the increased publication of papers in the American Meteorological Society's Monthly weather review journal starting in 1986. During this period, 49% of the papers from this source included in the analysed dataset were published. Most of the articles published during this period were climatological data, summaries and analyses for specific regions, most often South and North America.

Overall, the 20th century was marked by the growth of a global network of meteorological observations. Fundamental reference publications on climate, including climate atlases and monographs, appeared. Soviet scientists (O.A. Drozdov, E.S. Rubinstein, etc.) played a

leading role in this direction, focusing on complex climatology. In the middle of the 20th century, the idea of the heat balance of the Earth and the atmosphere as the basis of climate was formulated. These ideas were developed in the works of M.I. Budyko (USSR), G. Landsberg, D. Miller in the USA and others. In parallel, studies were carried out to assess the climate-forming role of moisture turnover in the USSR, USA, Japan, and Germany. Attention was paid to climatological processing of observations in mountainous areas (F. Steinhauser in Austria, M. Konchek in Czechoslovakia, etc In 1930, T. Bergeron in Norway proposed the concept dynamic climatology, which stimulated the study of the climate-forming role provided by the general circulation in the atmosphere. In the USSR, B.P. Alisov, V.A. Bugaev, V.A. Giorgio, B.L. Dzerdzeyevsky, H.P. Pogosyan, T.V. Pokrovskaya, S.P. Khromov and others worked on this problem, in the Germany there was G. Flon, in France where there was P. Pedelabord. In the USSR, climatic descriptions were accompanied by analyses of circulation conditions, which formed the basis for the classification of the globe's climates by B.P. Alisov (1952). Successes were also achieved in climatology of the tropics (in India, USA, China, FRG).

Due to the growth of cities and changes in the natural environment, interest in the study of microclimate and local climate related to anthropogenic changes and possible climate ameliorations has increased (works by R. Geiger in Germany, S.A. Sapozhnikova, I.A. Goltsberg in the USSR). Practice stimulates the development of agroclimatology and other applied climatological disciplines. In the field of palaeoclimatology since the 30s of the 20th c. generalisations appear by C. Brooks (Great Britain), G. Flon and others. In Germany R. Scherhag, in the USSR V.Y. Wiese and E.S. Rubinstein studied climatic changes of that time. An important part of the problem concerning natural climate change is to find out the influence that solar activity has on the climate; here the research by F. Baur in Germany, H. Willett in the USA, T.V. Pokrovskaya, L.A. Vitels and others in the USSR stands out.

The next significant spikes in publication activity were in 1965-1966, 1983-1984, 1995-1996, 2000, 2006, 2010, and 2013. Despite these spikes, the field as a whole has developed fairly smoothly, albeit with some rapidity, which is evident in the many small increases in the number of publications.

Currently, open access publications account for only a third (34.8%) of the entire sample. However, the dynamics of the distribution of open access and ‘closed’ access publications has changed over time. Up to and including 1949, open access publications prevailed, accounting on average for 66 per cent of the total number of publications per year. From 1950 onwards, the situation changed significantly, with open access publications averaging only a quarter of the total. However, as of 2019, open access publications account for half of the total volume of publications (± 10%).

The results calculated correlation coefficients (using Pearson and Spearman methods) revealed a moderate positive correlation between the type of publication dissemination (open access or non-open access) and the number of citations for the full data sample. For publications that were cited 2 or more times, the correlation was weaker but still statistically significant.

The relationship is especially noticeable when analysing the types of publications depending on the number of citations of articles. Most of the analysed sample (52%) consists of publications with zero citations, among which open access publications account for 23%. Among the publications with non-zero number of citations we can distinguish approximately equal ranges: 1-3 citations, 4-11 citations, 12-35 citations, and 36 and more citations. According to the results of the analysis, the publications with a large number of citations are dominated by open access publications. In the group of publications with 1-3 citations 31% are open access publications, and in the group with 4-11 citations - 46%. Starting from the group of publications with 12-35 citations, the number of open access publications starts to dominate, making up 55%, and in the group of publications with 36 and more citations it

reaches 62%. Among the 15 publications with the highest number of citations (more than 1000), two thirds are open access publications.

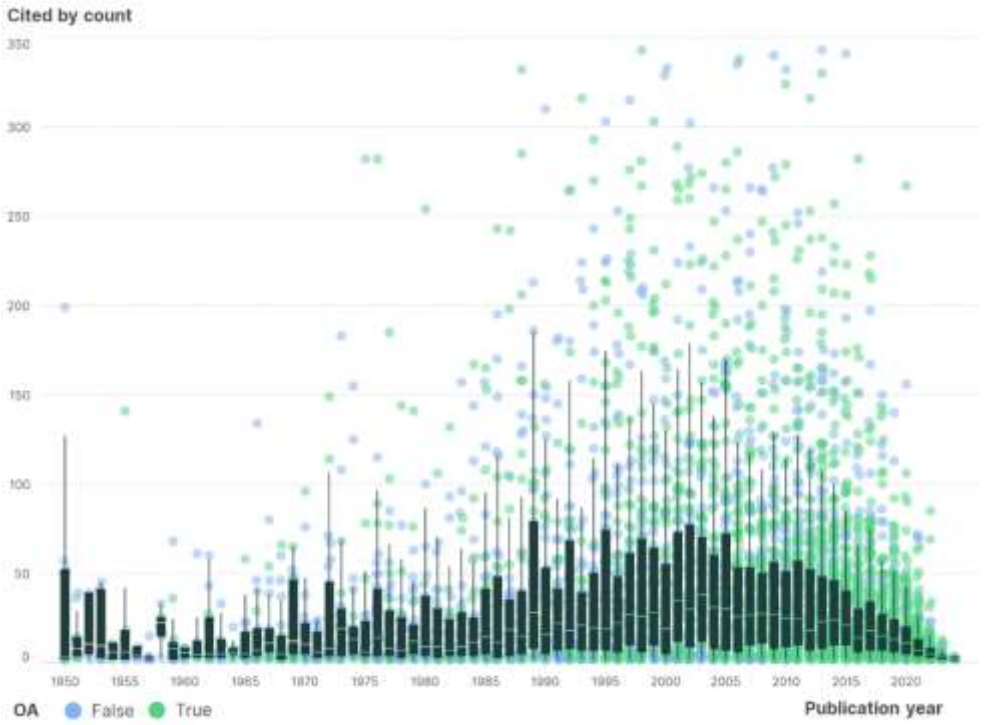

Fig. 1. Scatter diagram of the data sample documents by year relative to the number of citations with colour division by access type.
* To keep the visual clarity of the chart, only documents published since 1950, cited no more than 350, have been reflected.

Although it cannot be argued that Open Access guarantees more citations, two conclusions can be drawn. Firstly, Climatology has not always been dominated by 'closed' access publications, and in recent years the number of open access publications has been increasing, thus increasing the accessibility of research. Secondly, statistically, for authors who have chosen open access publication sources, this decision has paid off in terms of attention from the scientific community in the form of citations. This may be another reason for authors to choose open access resources and disseminate knowledge in climate science.

On the OpenAlex platform, each publication is assigned a specific concept or a number of them. Concepts are abstract ideas that are discussed in scientific papers. It can be a particular theory, method, or any other aspect that is the subject of the publication. OpenAlex indexes about 65,000 such concepts, which makes it possible to group and describe a number of studies without delving into the study of individual units.

In order to analyse the concepts attributed jointly to particular studies, the concepts were grouped into several clusters. Since climatology is not an isolated science, in most cases each cluster consists of one dominant concept, corresponding to classical climatology, and a number of minor concepts. These secondary concepts serve either to broaden or narrow down the main concept by the authors in their choice of research directions. Thus, the secondary concepts represent the directions in which climate science is evolving.

1. Geography cluster: brings together concepts from various fields, including data sciences (computer science, mathematics, statistics, artificial intelligence), humanities

(archaeology, history, philosophy, political science, sociology, law), applied sciences (engineering, aeronautics, composite materials) and medicine.

2. Climatology cluster: includes concepts related to climate research and meteorology (sea surface temperature, tropical cyclone, monsoon) as well as concepts from oceanography and thermodynamics.
3. Biology cluster: combines concepts from biology and ecology, and includes concepts from geography, palaeontology, cartography, geotechnical engineering, chemistry, agricultural hydrology and materials science.
4. Atmospheric Sciences cluster: includes concepts related to atmospheric sciences, physics, remote sensing, astronomy, satellites, aerospace engineering, troposphere, and aerosols.
5. Quantum Mechanics cluster: combines concepts from quantum mechanics, geodesy, geophysics, ionosphere, astrophysics and condensed matter physics.
6. Meteorology cluster: includes concepts related to meteorology, telecommunications, convection, radar, thunderstorms, physical power, boundary layer and tornadoes.
7. Geology cluster: includes concepts related to geology, snow, and the Arctic.
8. Environmental Science cluster: includes concepts related to environmental science, precipitation, climate models, and downscaling.

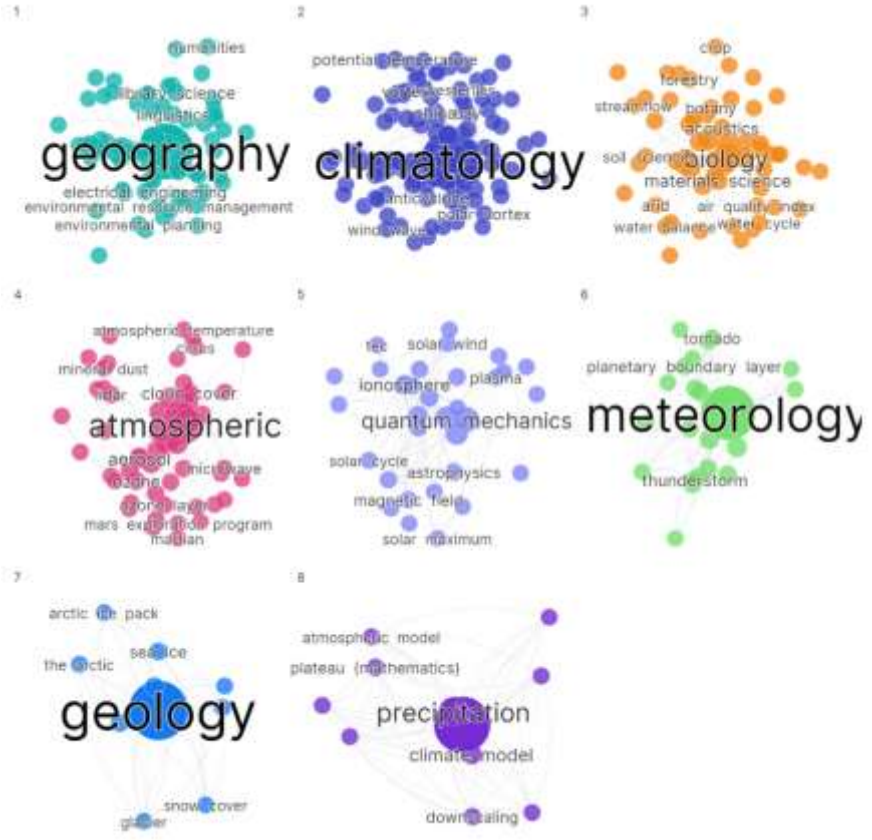

Fig. 2. A network of concept clusters to which the documents included in the analysed data sample belong. The size of the points is proportional to the number of references.

Among the dominant concepts, the concepts of geography, climatology and geology, which can be considered basic for this area, are the most frequently mentioned (according to

the average number of references for the whole period). However, it is worth noting that these are top-level concepts representing general thematic areas. For a more detailed understanding of the thematic landscape of the research, it is recommended to refer to specific terms and keywords that appear in the research.

The average year of mentioning a term or keyword can serve both as an indicator of the popularity peak and the average value between the beginning period when the term was mentioned and the moment when its popularity decreased. The key words and terms mentioned in the analyzed publications were grouped according to the average year when they were mentioned, which allows us to draw conclusions about the approximate chronology in development the climate science area: 1900-1910: Early Associations and Data, Climatological Societies and Profile Associations; 1910-1930: Sections and Sections; 1940-1950: Medical Climatology; 1950-1960: General Data; 1960-1970: Agricultural climatology and atlases; 1970-1980: General and physical climatology, meteorology; 1980-1990: Tropical and dynamic climatology; 1990-2000: Oceanography and records, water balance, precipitation and weather in general; 2000-2010: Climatological models, approaches and their significance, pollution and air quality, solar radiation, ozone layer and global warming; 2010-2020: Current research and trends, their impact. Selected natural phenomena and new technologies.

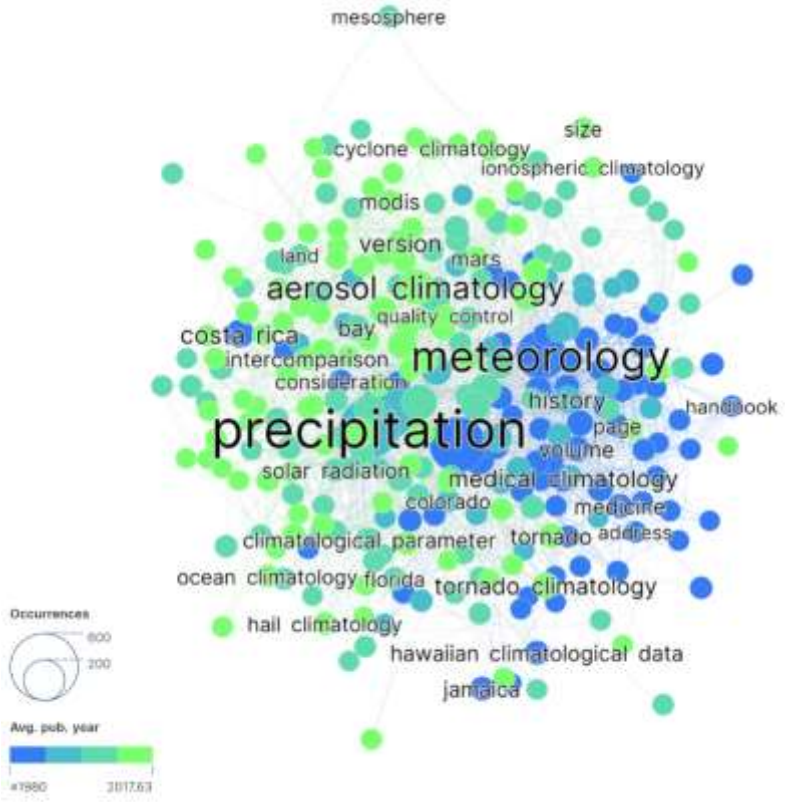

Fig. 3. A network of terms and keywords occurring in the analysed data sample.

Several terms and keywords are highlighted by the average number of citations, both normalised and normalised. The topics that received the most citations, especially ‘global ocean’, ‘Global Precipitation Climatology Centre’ (GPCC) and ‘Global Precipitation Climatology Project’, imply a significant focus on global climate processes and their analysis.

Subjects dealing with specific aspects of climate, such as 'cyclone climatology', 'dust climatology' and 'climatological drought', show an interest in more specialised areas of the study. In addition, the presence of topics related to methodologies and research tools such as GCM (General Circulation Model) and MISR (Multi-angle Imaging SpectroRadiometer) can be highlighted.

4 Conclusions

A quantitative analysis of climate science publications, including papers published since 1851, reveals a clear increase in publication activity over time. Currently, one-third of all publications are open access, but this proportion is increasing. The study indicates a moderate positive correlation between the type of publication distribution and the number of citations. Furthermore, it reveals that open access publications are statistically more prevalent in publications with a high number of citations.

Furthermore, the methodologies employed by the OpenAlex platform for the categorisation and description of scientific research were subjected to analysis. Climatology is an interdisciplinary science, and within its study, different concept clusters are formed that reflect major and minor research areas. The constructed network of concept clusters, to which the documents included in the analysed data sample belong, is entitled major concepts such as geography, climatology, biology, atmospheric sciences, quantum mechanics, meteorology, geology and environmental sciences. The secondary concepts included in the clusters serve to expand or narrow the main concept, thereby enabling authors to select research directions and to anticipate future developments in climate science.

A chronological chain of data field development was constructed based on the retrospective of the average year of keyword mentions, from early associations and data in 1900-1910 to the introduction of modern technologies relevant in 2010-2020. This indicates the content of the topic, its relevance and continuous development. The highlighted most demanded directions allow you to evaluate the current global issues that concern the scientific community of the topic "Climatology".

References

1. J. Culbert, A. Hobert, N. Jahn, N. Haupka, M. Schmidt, P. Donner, P. Mayr, Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus. arXiv preprint arXiv:2401.16359 (2024)
2. J. P. Alperin, J. Portenoy, K. Demes, V. Larivière, S. Haustein, An analysis of the suitability of OpenAlex for bibliometric analyses. arXiv preprint arXiv:2404.17663 (2024)
3. M. A. Simard, I. Basson, M. Hare, V. Lariviere, P. Mongeon, The open access coverage of OpenAlex, Scopus and Web of Science. arXiv preprint arXiv:2404.01985 (2024)
4. J. Priem, H. Piwowar, R. Orr, OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint arXiv:2205.01833 (2022)
5. Y. Hassan-Montero, F. De-Moya-Anegón, V. P. Guerrero-Bote, Profesional de la información/Information Professional **31(5)** (2022)
6. H. Abich, The Journal of the Royal Geographical Society of London **21**, 1-12 (1851)
7. B. Z. Yuan, J. Sun, Theoretical and Applied Climatology **150(1)**, 347-362 (2022)
8. P. Singh, B. Gargi, P. Semwal, S. Verma, Theoretical and Applied Climatology, 1-15 (2024)

9. R. Haunschild, L. Bornmann, W. Marx, PloS one **11(7)**, e0160393 (2016)
10. Z. Li, L. Guo, Y. Sha, K. Yang, Environmental Science and Pollution Research **28(36)**, 49755-49773 (2021)
11. S. H. Zyoud, D. Fuchs-Hanusch, Environmental Science and Pollution Research, **27(3)**, 3523-3540 (2020)
12. P. Alex, B. Preedip Balaji, *Mapping climate change research in India: A bibliometric approach* (2010)
13. G. Saravanan, V. R. Rajan, S. Prasad, G. Muthusankar, Climate change research (1991–2012): comparative scientometric study of Argentina, Brazil, China, India and Mexico. Library Philosophy and Practice, Paper-1134 (2014)