Identifying and benchmarking polar research - bibliometric issues and implications

Abstract

This paper explores the challenge of identifying Antarctic polar research using conventional keyword searches, and examines the way these approaches can be used for benchmarking the bibliometric impact of polar research.

It reviews existing work to identify bibliometric methodologies for Antarctic research outputs, and examines their effectiveness in extracting Antarctic-specific literature from the field of polar research as a whole. It reviews keywords used, including geographical search terms, and identifies issues that may affect metrics for countries with a strong bi-polar focus, as well as measures taken to refine and constrain the search parameters in specific databases.

Secondly, this paper seeks to benchmark the current impact of Antarctic polar research against global patterns in scientific research. It seeks to establish whether some observed changes in citation patterns are specific to the field or part of a global trend, looking at the citation impact historically compared to recent years.

Part 1 – bibliometric process

Review of past literature for search strategies

This paper builds on previous work presented to the Colloquy in 2016 (Gray 2023), 2018 (Gray 2019), and 2022 (Ager & Gray, in press). In each of those studies, a moderately complex search string was used to identify Antarctic papers as those which explicitly mentioned Antarctica or one of a set of relevant geographic keywords.

While it is widely agreed that a keyword-based search method is the most effective for finding Antarctic papers, the exact keywords used can vary substantially.

The simplest is a keyword search for “Antarctic”/ “Antarctica”, or a stemmed term such as “Antarctic*”. This approach was used by Dastidar (2007); Dudeney & Walton (2012); Kim & Jung (2016); González-Aravena et al (2021); and by Gray (2023). Aksnes & Hessen (2009) used *antarct* (thus including sub- and peri- Antarctic research) as well as a number of specific geographic keywords such as “Weddell Sea”, most notably including “south pole”, an issue which will be discussed later. Ji et al (2014) used an “Antarctic*” search and manually screened out false positives (such as papers on the yeast *C. antarctica*).

Stefenon et al (2013) and Fu & Ho (2016) searched for specific variants of “Antarctica”, likely to generate similar results to using “Antarctic*”. Hua et al (2014) added a number of additional geographic keywords, as well as station names; Zhang et al (2023) used a similar approach.

An “antarct*” search adding additional geographic terms and correcting for false positives was used in Gray & Hughes (2016), as well as papers at the Colloquy in 2018 and 2022 (Gray 2019,

Ager & Gray, in press). This approach omitted “south pole” due to potential issues around false positives.

Using a different approach, Jang et al (2020) used a two-step process, with an “Antarct*” search to identify an initial set of papers, then generating a detailed list of several hundred relevant keywords from those to bootstrap a second search. However, this approach may have issues due to the way Web of Science keywords are assigned, as will be discussed later in this paper.

Keywords and Keywords Plus

These searches tend to use a combined search on title, abstract, and keywords. In Scopus, this is labelled as such; in Web of Science, it is described as a Topic Search. Both databases, however, include a range of material under “keywords”. In Web of Science, it is a combination of the Author Keywords and Keywords Plus fields. The former are the keywords assigned by an author or by a journal publisher. The latter was developed by Clarivate and has been a feature of the Web of Science platform since 1991. Based on an algorithm, it takes words or phrases from an article’s listed references and searches across disciplines for common references or clusters of co-citation to create relevant subject index terms. It includes areas of overlap with the results produced from searching the other fields within Topic Search, but expands upon it, returning an increased number of – what should be – relevant results. In comparison to the author keywords, Keywords Plus *should* provide a precis of the subject matter of the paper concerned (Garfield, 1990). However, close examination shows that it has some issues which can be particularly challenging when using geographic keywords. Below is an example record returned by the Antarctic search:

> Title: Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice
>
> Author keywords: marine bacteria; phages; psychrophilic; sea ice
>
> Keywords Plus: MICROBIAL COMMUNITIES; BACTERIAL PRODUCTION; **SOUTHERN-OCEAN**; **MCMURDO-SOUND**; **WEDDELL SEA**; VIRUSES; ABUNDANCE; BACTERIOPHAGE; **ANTARCTICA**; DYNAMICS

Based on the title and author keywords, we can see that this is clearly an Arctic paper. However, in the Keywords Plus entry, it has been assigned four distinct terms indicating that it refers to Antarctica (and, notably, has not been assigned any Arctic-specific terms).

Overall comparison of Keywords Plus results

Unless stated otherwise, figures given are for searches run on a full version of the Web of Science Core Collection inclusive of all ten citation indices (Science; Social Sciences; Arts & Humanities; Conference Proceedings Science; Conference Proceedings Social Science & Humanities; Books Science; Books Social Sciences & Humanities; Emerging Sources; Current Chemical Reactions; Index Chemicus).

In order to isolate the effect of the Keywords Plus (KP) search terms in the Topic Search (TS), it was necessary to construct searches which only identified the TI, AB, AK elements – title, abstract, and author keywords. To ensure the searches were evaluated in the same way as the TS search required a slightly complex format:

((TI=(antarc*) OR AB=(antarc*) OR AK=(antarc*)) NOT (TI=(candida OR "except antarctica" OR "not antarctica" OR "other than Antarctica") OR AB=(candida OR "except antarctica" OR "not antarctica" OR "other than Antarctica") OR AK=(candida OR "except antarctica" OR "not antarctica" OR "other than Antarctica"))) OR (TI=("transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "southern ocean") OR AB=("transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "southern ocean") OR AK=("transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "southern ocean"))

This was compared to a search for papers matching a KP element:

KP=((antarc* NOT (candida OR "except antarctica" OR "not antarctica" OR "other than Antarctica")) OR "transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "southern ocean")

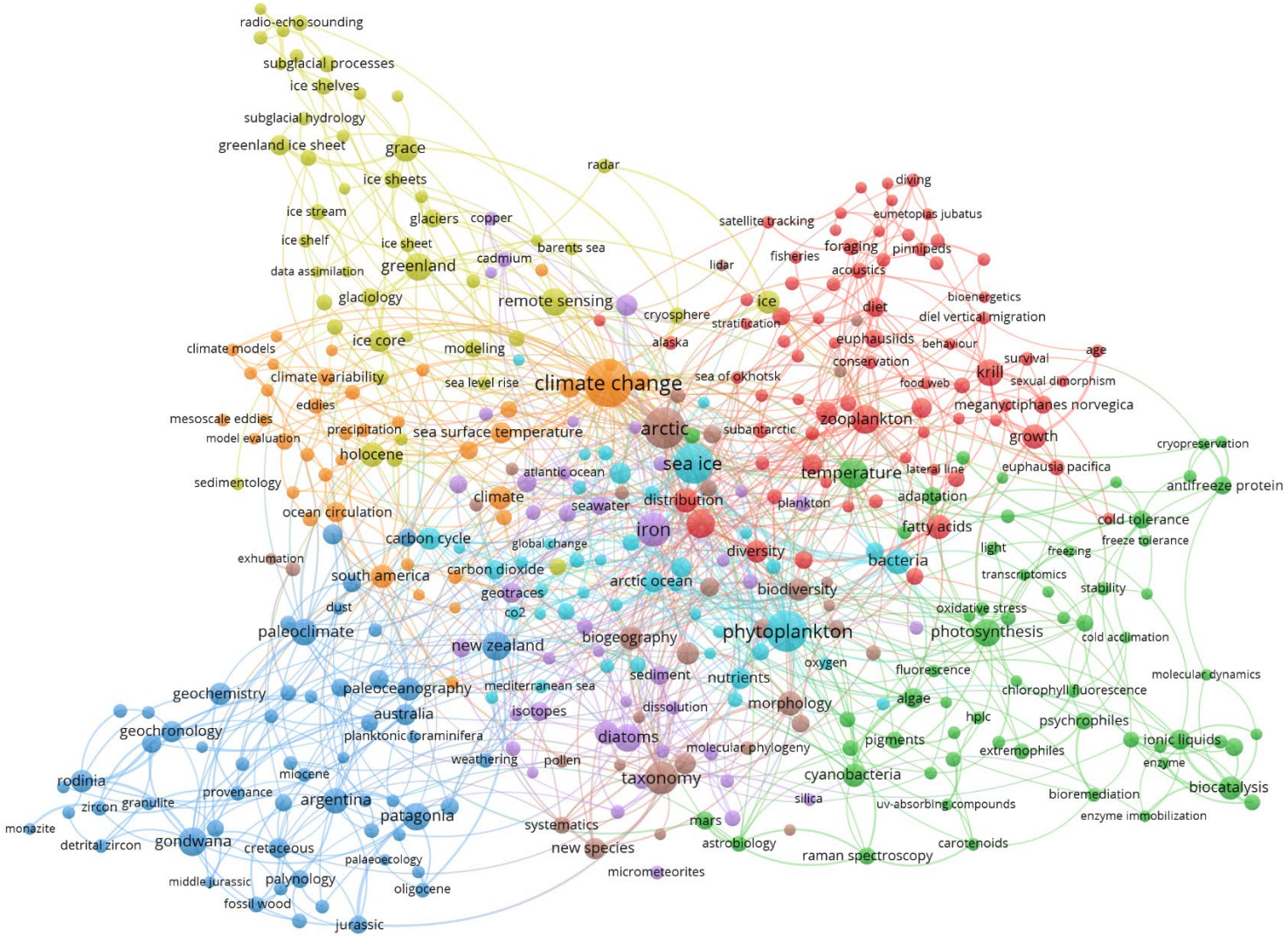
The two were combined with NOT in order to return papers that were only found using the KP search. This returned 87822 papers for an overall TS search, 74964 for TI/AB/AK, 27936 for KP, and 13083 solely matched by KP – around 15% of the original hits.

Reliability of Keywords Plus identification

136x randomly chosen papers were sampled from the KP-only set (around 1% of the total) and assessed as to whether or not they were "Antarctic" papers, and would have been expected to be returned by a search. The majority (75-80%) were clearly not, and only a small minority (10%) were clearly Antarctic-related. The remainder (10-15%) were papers that were not particularly Antarctic-focused but where it would not be surprising had a relevant keyword appeared in the abstract – eg. space weather analysis using data gathered in Antarctica, or global climate studies that touched on the Antarctic ice sheets.

Among the "clearly Antarctic" papers, there were a number that used Antarctic-relevant keywords which was not in our initial search strategy, as well as several that discussed Antarctic species. As an example, a 2019 paper dealing with the vulnerability of Emperor penguins at Halley Bay to climate change was absent from the TI/AB/AK search, as it omitted any of the larger geographical terms. "Emperor penguin" and "Halley Bay" are both unambiguously Antarctic, but were not covered by the search. It was only identified through a "Ross Sea" Keywords Plus entry – which was itself a false positive, as this paper was not studying the Ross Sea area.

To illustrate the high level of non-relevancy produced by KP-only searches, VoS Viewer – a software used for visualisations of bibliographic data – was used to create the cluster diagram below. This shows groupings of author keywords extracted from the KP-only search papers. Assuming that the author keywords from these papers are a more reasonable indicator of subject matter than Keywords Plus, it seems significant that none of the terms featured are specifically Antarctic-oriented.


radio-echo sounding
subglacial processes
ice shelves
subglacial hydrology
greenland ice sheet
grace
ice sheets
ice stream
glaciers
copper
ice shelf
ice sheet
cadmium
barents sea
data assimilation
greenland
glaciology
remote sensing
radar
climate models
ice core
modeling
climate variability
sea level rise
climate change
eddies
mesoscale eddies
precipitation
sea surface temperature
model evaluation
holocene
arctic
sedimentology
atlantic ocean
sea ice
climate
seawater
distribution
ocean circulation
iron
carbon cycle
global change
exhumation
carbon dioxide
south america
arctic ocean
geotraces
biodiversity
dust
co2
paleoclimate
new zealand
biogeography
phytoplankton
sediment
nutrients
geochemistry
paleoceanography
mediterranean sea
morphology
australia
isotopes
dissolution
geochronology
planktonic foraminifera
diatoms
molecular phylogeny
weathering
pollen
miocene
taxonomy
rodinia
provenance
argentina
zircon
granulite
patagonia
monazite
detrital zircon
gondwana
cretaceous
palaeoecology
middle jurassic
palynology
oligocene
fossil wood
jurassic
systematics
new species
micrometeorites
mars
astrobiology
satellite tracking
diving
eumetopias jubatus
foraging
pinnipeds
fisheries
acoustics
lidar
bioenergetics
diet
ice
cryosphere
stratification
diel vertical migration
euphausiids
behaviour
alaska
conservation
survival
age
krill
sea of okhotsk
food web
sexual dimorphism
meganyctiphanes norvegica
zooplankton
subantarctic
growth
cryopreservation
temperature
lateral line
euphausia pacifica
antifreeze protein
plankton
adaptation
fatty acids
cold tolerance
freeze tolerance
light
freezing
bacteria
transcriptomics
stability
oxidative stress
photosynthesis
cold acclimation
oxygen
fluorescence
molecular dynamics
chlorophyll fluorescence
algae
hplc
psychrophiles
pigments
ionic liquids
extremophiles
enzyme
cyanobacteria
bioremediation
biocatalysis
silica
uv-absorbing compounds
enzyme immobilization
carotenoids
raman spectroscopy

VOSviewer

The issue of decreased relevance for the Web of Science Topic Search is not a purely Antarctic phenomenon. For example, a test using a series of Arctic keywords in Chinese publications, using a TS versus TI/AB/AK search, identified that the TS search returned a large number of Himalayan cryospheric papers which were given Arctic keywords.

Known Antarctic sources

Two groups of papers which were presumed to be Antarctic-focused were also assessed. The first of these were papers in *Antarctic Science*, the only purely Antarctic-focused journal currently indexed in Web of Science. Of the 2036 papers indexed in that journal, only 42 (2%) were returned by a KP-only search. These papers were almost entirely dealing with Antarctic or sub-Antarctic regions, with occasional papers looking outside the sub-Antarctic regions (eg. the Falkland Islands).

The second source was papers published by the British Antarctic Survey (BAS). While BAS does carry out and publish non-Antarctic research, its work is heavily focused in the area. Of 9920 BAS papers, 308 were returned by a KP-only search. Of these, there were two relevant groupings, which could be roughly categorised as geographic – eg. Pine Island Glacier, Filchner-Ronne Ice shelf, or biological – eg. Emperor penguins. Others were more broadly related to sub-Antarctic, so would only be relevant in an expanded search. Arctic research also featured strongly as did more generalised/global research impacts eg. global ocean circulation, climate change modelling.

Comparison to Scopus

A comparison search was conducted using Scopus syntax (from Xavier et al 2018):

1. TITLE-ABS-KEY (antarct* OR "southern ocean" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR “transantarctic”) AND NOT TITLE-ABS-KEY (candida OR "except antarctica" OR "not antarctica")
2. TITLE (antarct* OR "southern ocean" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "transantarctic") OR ABS (antarct* OR "southern ocean" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "transantarctic") OR AUTHKEY (antarct* OR "southern ocean" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "transantarctic") AND NOT (TITLE (candida OR "except antarctica" OR "not antarctica") OR ABS (candida OR "except antarctica" OR "not antarctica") OR AUTHKEY (candida OR "except antarctica" OR "not antarctica"))
3. INDEXTERMS (antarct* OR "southern ocean" OR "ross sea" OR "amundsen sea" OR "weddell sea" OR "transantarctic") AND NOT INDEXTERMS (candida OR "except antarctica" OR "not antarctica")

On close examination, some similar results were encountered. The INDEXTERMS search (keywords added by the indexers, not the authors) identified a number of papers which were indexed as “Antarctica” but omitted from the TITLE/ABSTRACT/AUTHKEY search as they used a more specific geographic term. It also identified a large number of sub-Antarctic papers covering South Georgia, etc.

However, it also introduced a new set of false positives. Some papers were assigned the Geobase geographic keyword “Antarctica” seemingly on the basis of it appearing in the abstract, even if that was the phrase “all continents except Antarctica”. Another set of papers were retrieved due to the Embase scientific terms indexing all species mentioned, and including “*X. antarctica*” – the standard search tries to filter out *Candida antarctica*, but there is a long tail

of other species with similar suffixes. While many of these were first identified in Antarctica, they are not endemic and searching for them would entail false positives.

It is clear that the same general issues which apply to Keywords Plus apply to Scopus index terms – albeit for slightly different reasons. Any search strategy using Scopus would need to consider these effects and potentially filter them out in the same way.

Analysis of individual terms

Examination of the papers identified by a Keywords Plus search but not by a TI/AB/AK one suggested that many of the Antarctic-focused papers were using specific keywords identifying them as Antarctic to the reader, which were not found in our search.

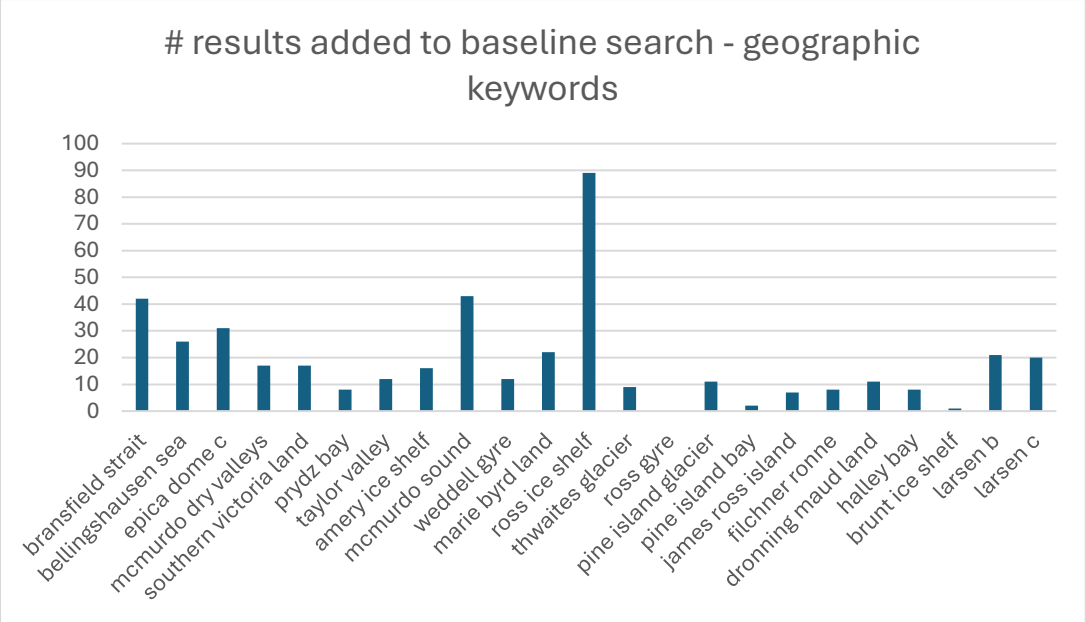
These broadly fell into three groups:

- First, explicitly Antarctic geographical keywords, such as station names or seas, not listed in our research.
- Secondly, biological keywords for Antarctic endemic species – a paper studying Emperor penguins might not see the need to note the continent.
- Thirdly, the special case of "South Pole" / "South Polar", which usually indicated Antarctic work but also included a significant number of astronomical or planetary science papers. This required an additional negative element to correctly address.

Finally, these papers also highlighted ambiguities around the definition of "Antarctic". The two commonly used definitions are either to the Antarctic Treaty area (60S) or to the Antarctic Convergence (the northern boundary of the Southern Ocean, varying between 45S and 60S depending on sector). The broader definition would include some sub-Antarctic regions, such as Kerguelen and South Georgia. Our initial search uses the phrase "southern ocean", which would lead towards a more expansive definition, but did not specifically include sub-Antarctic as a keyword, or any of the named sub-Antarctic regions.

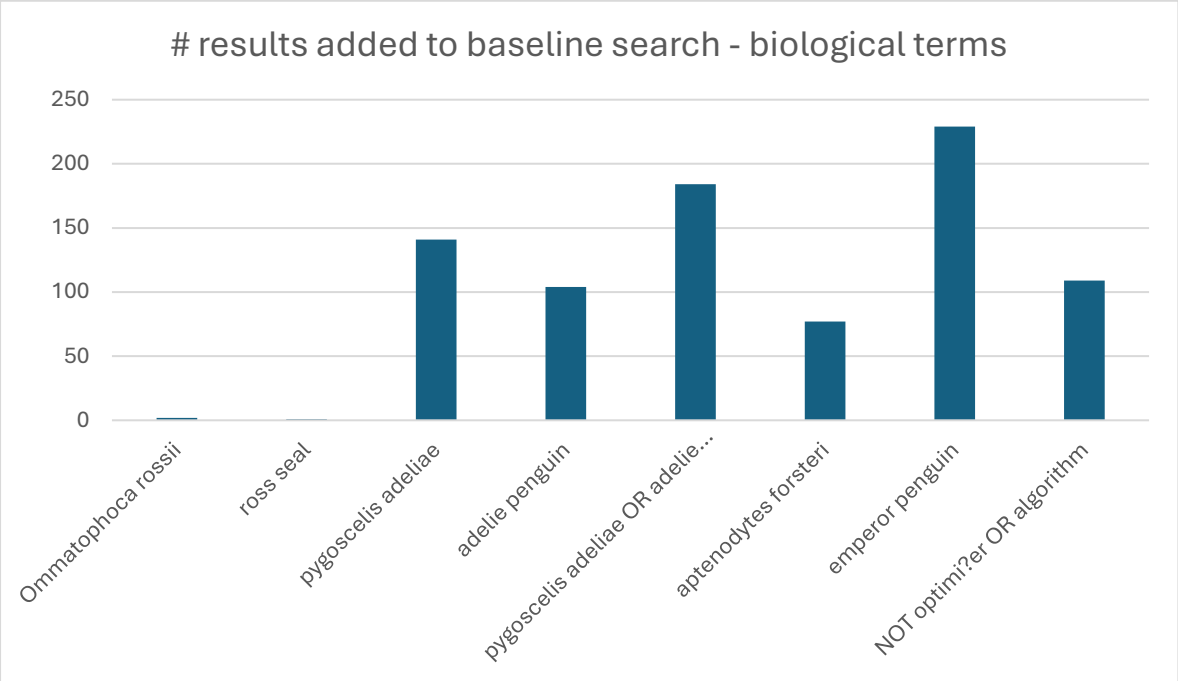
Geographic keywords

As a means of identifying potentially useful keywords, results from both Scopus and Web of Science searches were checked for relevant articles that were identified solely by Index Terms or Keywords Plus respectively. From this a list of potential geographic keywords were picked out for testing, although this approach proved somewhat limited as most relevant results were tagged with just the generic "Antarctic" keyword. As shown by the bar chart below, the addition of each individual geographic term to the standard Web of Science TI/AB/AK search in most cases only produced a limited number of additional results. With the possible exception of Ross Ice Shelf, it was not considered that additional search terms contributed enough to the overall number of the results to be worthwhile including in a metrics-based analysis.



Biological keywords

A number of relevant results not picked up by the standard Antarctic search relate to Antarctic biology – such as the Emperor penguin paper referred to above. To reduce this, a number of sample terms specific to Antarctic fauna were tested. Emperor penguins and Adelie penguins are both endemic to Antarctic, also Ross seals, which apart from some isolated sub-Antarctic examples are generally only found in the Antarctic region. The bar chart below shows the number of results gained by the individual inclusion of each term to the Antarctic search. It was found that both the common and Latin name for each species were required as they weren't consistently used together – eg. Adelie penguin returned 104 results and *Pygoscelis adeliae* 141, whereas both terms added 185 results.



Adding Emperor penguin to the search exposed another false positive, returning a significant number of results related to an optimization algorithm (Emperor Penguin Optimizer or EPO) which would need excluding from the results. Variant terms were identified and tested to check if broadening the keywords eg. optimiz* would exclude too many legitimate papers. The revised sub query constructed was as follows:

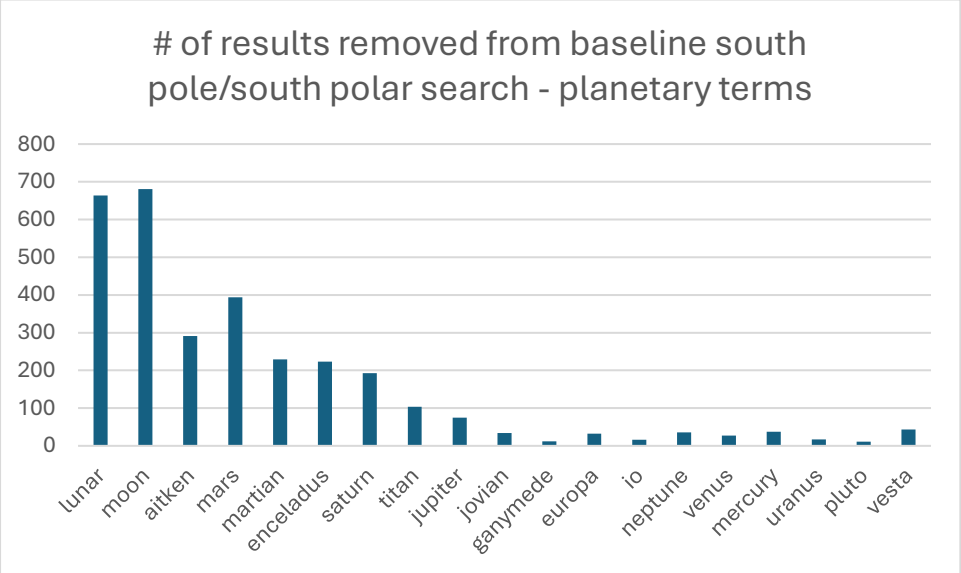
((TI="emperor penguin") OR AB=("emperor penguin") OR AK=("emperor penguin")) NOT (TI=("optimi?er" OR "algorithm") OR AB=("optimi?er" OR "algorithm") OR AK=("optimi?er" OR "algorithm"))

South Pole

The value of adding “South Pole” and “South Polar” to the standard Antarctic search was studied separately to the exploration of geographic terms above. A baseline search of just these two terms, as shown below, gained over 5335 results alone, of which around 77% were not being picked up by the standard Antarctic search:

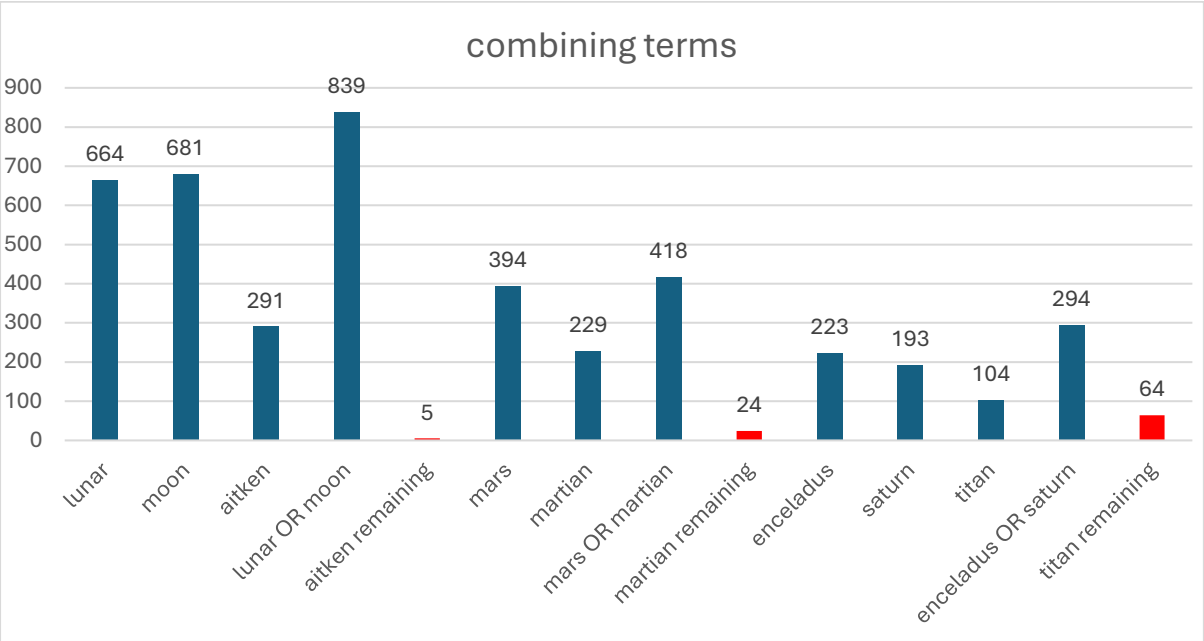
> (TI=("south pole" OR "south polar") OR AB=("south pole" OR "south polar") OR AK=("south pole" OR "south polar"))

However, as anticipated, these results included a high proportion of false positives consisting mainly of other planetary bodies with poles, eg. lunar South Pole, Mars South Pole etc. The bar chart below indicates how many results would be removed by adding each term individually to a NOT clause eg. 681 from “moon” and 664 from “lunar”. However, this approach needs to be weighed against the risk of excluding relevant results, such as research referring to mercury in soils, for example. Also, many experiments use Antarctica’s remoteness to study the impacts of working in isolated environments, as an analogue for space travel or future manned Mars missions. Geographic features named for astronomers and their findings have resulted in place names such as Mars Hills – although sample searches for some of these terms suggested they did not feature in the literature to any significant extent.



To achieve a more economical search structure, searches were run using “umbrella” terms to capture the more granular terms within – eg. lunar OR moon would also remove all but 5 results relating to the South Pole Aitken basin. Running a query using these combined planetary terms removed 69% of the 5335 results from the South Pole baseline search and overall added around

3% to the standard Antarctic search (equating to over 2000 papers) – a reasonable-sized pool that could be further spot-sampled for relevancy.



A significant body of astronomical research is conducted at the South Pole, as evidenced by the level of articles returned that refer to South Pole Telescope or the ICECUBE neutrino observatory. It was considered that this could still reasonably be viewed as "Antarctic research" as it is carried out in Antarctica, managed by Antarctic agencies, and involves significant effort and investment in Antarctic infrastructure. However, it may be desirable to remove this type of research if focussing on the Antarctic environment itself. If so, the best approach would be to add a broader set of astronomical keywords to the filter to remove these papers, or consider filtering out all research identified as astronomy, such as by using a Web of Science category filter WC=("astronomy")).

Recommended search strategy for Antarctic papers

This leads us to three recommendations for search strategies.

Firstly, we recommend the use of **a larger set of Antarctic-related keywords**.

Secondly, we recommend that complex keyword search strategies for Antarctic-related papers, particularly for overall statistical/bibliometric analysis, **only use the Title, Abstract, and Author Keywords fields**, and where possible avoid the Keywords Plus / Indexed Terms / etc fields. This unfortunately requires a more complex search, but significantly reduces false positives.

Thirdly, we recommend making an explicit and clear decision on **what definition should be used for "Antarctic-related"**. This has two aspects: the geographic space and the orientation of the work. Defining the geographic space means reaching a decision on whether work done at South Georgia or in the other sub-Antarctic islands should be considered as "Antarctic" research; this may depend heavily on the context and purpose of the enquiry. Applying the "Southern Ocean" keyword may inevitably mean that a narrow search contains some of these results in any case, but omitting it would lead to a large amount of unambiguously Antarctic

papers being lost. If the sub-Antarctic islands are intentionally included, this may involve adding more geographical keywords as well as more for endemic species.

Considering the orientation of the work brings in the question of work performed in the Antarctic regions that is not actually studying those regions – it is merely more practical to carry out the research there. The most significant element here is astronomical, and we saw that a keyword search for "South Pole" inevitably brought in a large amount of work carried out by the South Pole Telescope, ICECUBE, and other observatories based at the Pole. We would recommend that either these are explicitly excluded (a search for South Pole keywords with planetary and telescope keywords removed) or that they are included (South Pole keywords with only planetary keywords removed). Again, this will depend on the context and purpose of the enquiry. A study looking at "scientific activity in Antarctica", for example, would probably wish to consider these as indicators of work done there; a study on the development of "polar research" might not.

```
( ( TI=(antarc*) OR AB=(antarc*) OR AK=(antarc*) ) NOT ( TI=(candida
OR "except antarctica" OR "not antarctica" OR "other than
Antarctica") OR AB=(candida OR "except antarctica" OR "not
antarctica" OR "other than Antarctica") OR AK=(candida OR "except
antarctica" OR "not antarctica" OR "other than Antarctica") ) )
OR ( ( TI=("transantarctic" OR "ross sea" OR "amundsen sea" OR
"weddell sea" OR "southern ocean" OR "bransfield strait" OR
"bellingshausen sea" OR "mcmurdo sound" OR "ross ice shelf" OR
"adelie penguin" OR "pygoscelis adeliae" OR "aptenodytes forsteri")
OR AB=("transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell
sea" OR "southern ocean" OR "bransfield strait" OR "bellingshausen
sea" OR "mcmurdo sound" OR "ross ice shelf" OR "adelie penguin" OR
"pygoscelis adeliae" OR "aptenodytes forsteri") OR
AK=("transantarctic" OR "ross sea" OR "amundsen sea" OR "weddell
sea" OR "southern ocean" OR "bransfield strait" OR "bellingshausen
sea" OR "mcmurdo sound" OR "ross ice shelf" OR "adelie penguin" OR
"pygoscelis adeliae" OR "aptenodytes forsteri") ) )
OR ( ( TI=("emperor penguin") OR AB=("emperor penguin") OR
AK=("emperor penguin") ) NOT ( TI=("optimi?er" OR "algorithm") OR
AB=("optimi?er" OR "algorithm") OR AK=("optimi?er" OR "algorithm") )
)
OR ( ( TI=("south pole" OR "south polar") OR AB=("south pole" OR
"south polar") OR AK=("south pole" OR "south polar") ) NOT (
TI=("lunar" OR "moon" OR "mars" OR "martian" OR "enceladus" OR
"titan" OR "jupiter" OR "venus" OR "neptune" OR "saturn" OR "vesta")
OR AB=("lunar" OR "moon" OR "mars" OR "martian" OR "enceladus" OR
"titan" OR "jupiter" OR "venus" OR "neptune" OR "saturn" OR "vesta")
OR AK=("lunar" OR "moon" OR "mars" OR "martian" OR "enceladus" OR
"titan" OR "jupiter" OR "venus" OR "neptune" OR "saturn" OR "vesta")
) )
```

The second recommendation is most significant for statistical and bibliometric analysis. For individual literature searching, etc, the false positives are less concerning as it will be straightforward to filter these out by manual review. In addition, some of the complex negative

elements (eg. candida, planetary south poles) could be omitted if a manual review will be taking place. However, for a large scale or statistical analysis, where manual review is not practical, the use of these keywords can potentially lead to significant numbers of false positives appearing in the totals. As such, removing the negative elements which screen out the false positives is risky unless the results are being individually examined.

While we have not looked at searches for the Arctic, it is quite likely that similar issues apply to the use of geographic keywords – papers being incorrectly identified through overclassification. We would encourage users writing search strategies for statistical purposes to consider the effects of Title, Abstract, and Author Keywords only searches rather than more expansive Topic ones.

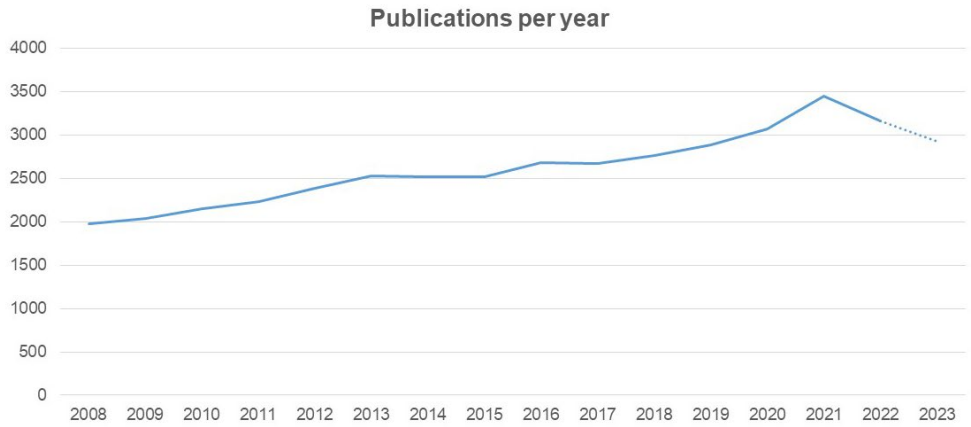
Part 2 – the impact of Antarctic research

Patterns in impact of Antarctic research

The data provided within this section largely builds on what has been presented previously with relation to the profile of Antarctic papers and the bibliometric patterns of Antarctic research worldwide (Ager & Gray, in press). The publications assessed here are those matching the final Web of Science search described above, filtered to only articles, reviews, conference papers and data papers, and matched into the InCites dataset.

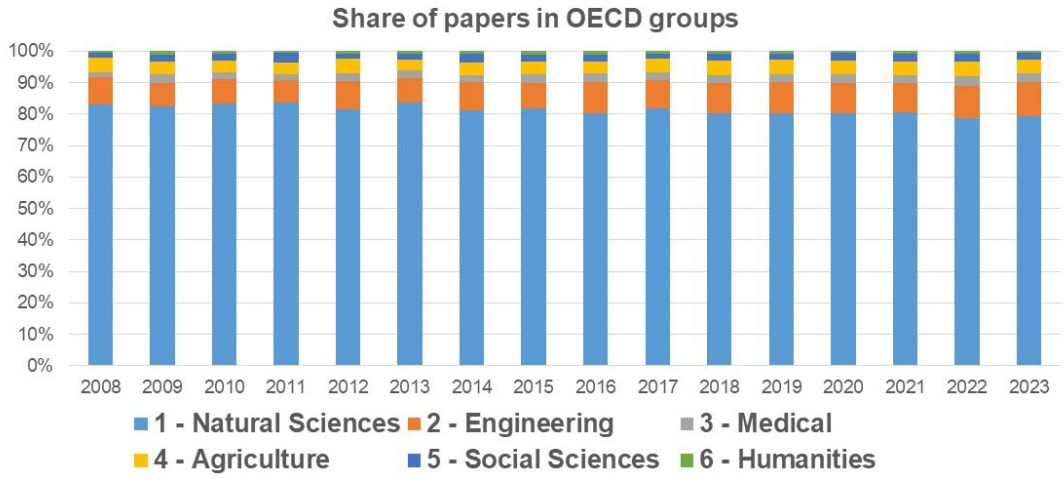
Total output

In terms of annual output of Antarctic papers, total volume continues to rise, at a rate of around 5% per year from 2008 to 2021. The peak of 2021, as shown in the graph below, is found consistently across global research and is not particular to Antarctic science (Ahrabiran et al 2024). It is primarily a result of publishing patterns connected with the pandemic, exacerbated by a change to the way early access papers are accounted for in Web of Science that causes an artificial boost in 2021 (Clarivate 2022).



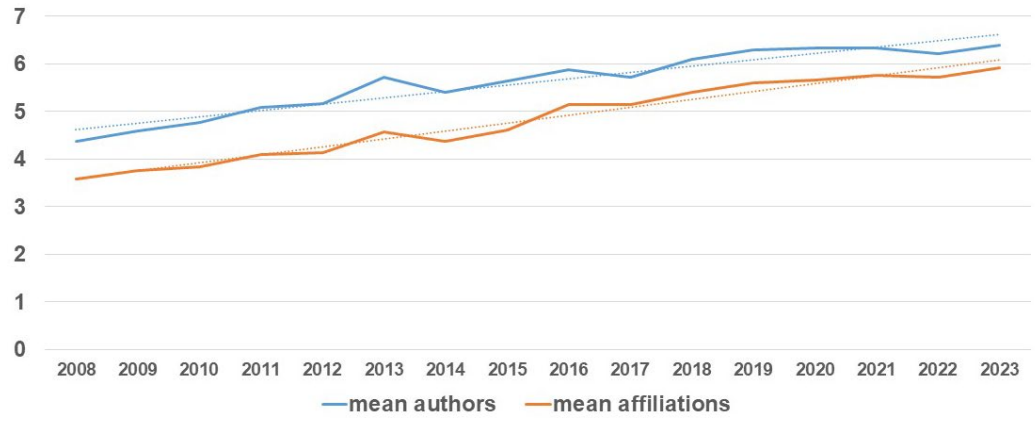
Subject areas

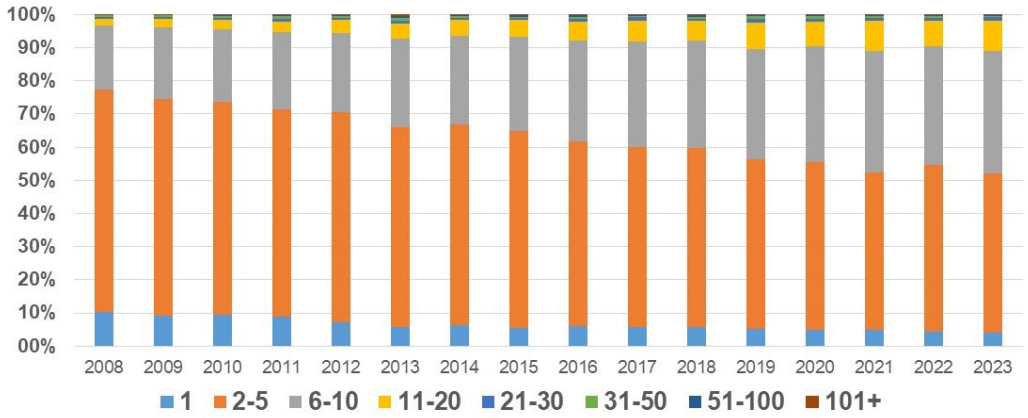
In terms of activity by subject area, OECD classifications have been used to show the distribution of papers across the six main OECD groupings. Note, unlike Web of Science categories, papers are only assigned to one main category within the OECD schema. Most Antarctic research falls under the broad categorisation of Natural Sciences – an average of 81% for the period – followed by Engineering (9%), Agriculture – namely, fisheries research (4%), Medicine (2.5%), Social Sciences (2.2%), with a negligible amount remaining defined as Humanities. The split of these categories would appear to be relatively stable with little evidence of variation across the years. A breakdown of the subjects within these five main groupings is provided in the Appendix.



Author numbers

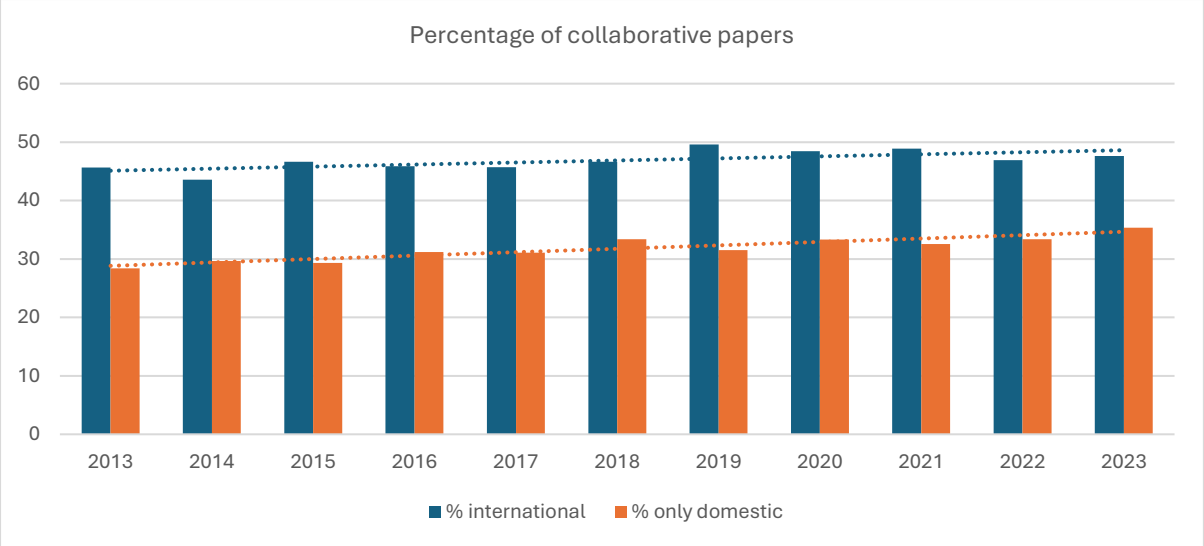
The graph below indicates a rise in the average number of authors per paper, which is consistent with international trends across a range of disciplines (eg. Jakab et al 2024). As indicated in the bar-chart below, there has been a fall in single-author papers (10% to 4%) and papers with 2-5 authors (66% to 50%). Conversely, significant rises in papers with 6-10 and 11-20 authors (20% to 37%; 2% to 9%) – ie. the medium-long author lists are increasing. However, there is little significant change in the small proportion of papers with very large numbers of authors, which follows a similar pattern to that reported previously.





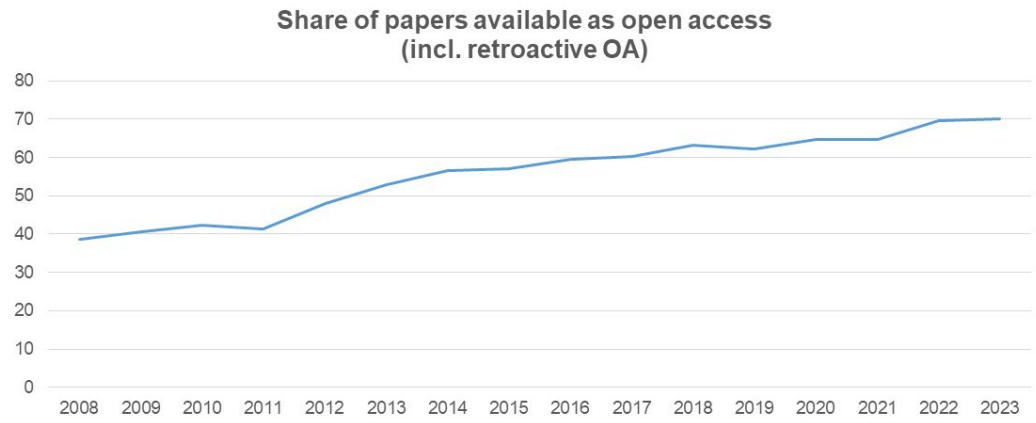
International collaboration

International cooperation is a key criterion of the Antarctic Treaty and collaborations on scientific papers can be used as a measure of this. Country data are recorded in the author affiliation field in Web of Science and multiple-author papers with at least two affiliated countries are termed an international collaboration. As shown below, there is a little variation over the period but no significant increase, however, the number of domestic-only papers seems to be increasing. Domestic-only excludes single-author papers and are defined as papers with at least one other same-country affiliation. The share of non-collaborative papers – those with only one author or from only one institution – continues to diminish steadily.



Open Access

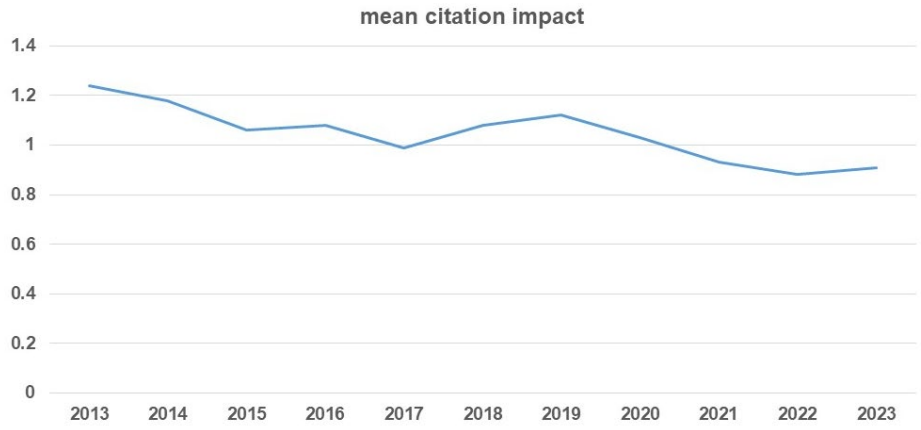
Open Access to Antarctic research increasing and as of 2023 sits at about 70% globally. To note, the graph below includes retrospective open access, not just papers that have been made openly and freely available at point of publication – it indicates the percentage of papers that are freely available as of 2024. The effective open access rates differ significantly between individual countries, as a major driver for this has been local government policy and funder mandates, leading to a high level of local variation.

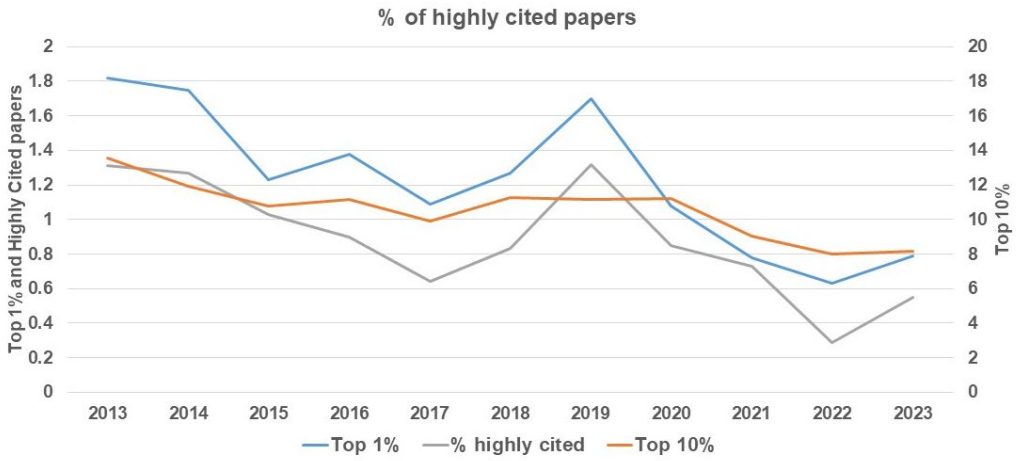


Global impact patterns

The citation impact of Antarctic papers was calculated using InCites Category Normalised Citation Impact (CNCI). The CNCI measure compares each paper to a reference set of comparable papers of the same age, publication type, and subject area; a CNCI value of 1 would indicate the group had the same normalised citation rate as the world average. The data used here includes Web of Science content indexed through to 31 March 2024.

It can be observed that there was an overall decline in average CNCI for the Antarctic pool of papers with a discernible dip in 2022. This pattern is also shown in the Web of Science highly cited categories.





It has proved difficult to identify the drivers of this decline. As shown in the breakdown of papers by subject category, there has not been any significant alteration to how Antarctic research has been categorised that might significantly alter CNCI. We have seen in earlier work (Ager & Gray, in press) that the pattern is consistent in both Arctic and Antarctic papers; it is apparently a pattern of polar science at large. The distribution of papers among CNCI bands remains reasonably consistent year on year.

It is unlikely to be purely an artefact of how papers are classified, but this may have some effect; CNCI is defined by reference to a group of papers in the same subject. The default schema classifies subjects into 250x categories by journal, but the "highly cited papers" metric uses a simplified set of 22 broader subject categories and, as can be seen, closely tracks the standard top 1% metric. Assessing it on the basis of the more granular Citation Micro Topics (2449x groupings assigned by citation clustering) gives a CNCI decline from 1.25 to 1.13 – the trend is the same, but the decline itself is less dramatic.

It is possible that the effect is linked to the dominance of particular subjects. In the Web of Science schema, the largest single category (Geosciences, Multidisciplinary) includes 27.6% of the total papers, and the top five categories 61.3% collectively. In the ESI schema used for highly cited papers, the largest category (Geosciences) includes 50.5% of papers. The overall CNCI patterns closely track that for the geosciences categories. Conversely, in the Citation Topics schema, the largest grouping (8.19.668 Glacier) includes just 12.9% of papers and the top five topics represent 36.7%. Citation underperformance in one particular area will thus have a less dramatic effect.

It is likely that there is no single explanation for this shift, and it is driven by a combination of various factors, some of which are not immediately obvious.

Discussion and analysis

Creating a reliable dataset is crucial for carrying out bibliometrics-based studies on Antarctic research. As shown in the first part of this paper, the Keywords Plus feature – part of the Web of Science Topic Search – is not recommended for this purpose as it will also identify Arctic papers as relevant. However, this can be avoided by using a more precise field search structure limited

to Title, Abstract and Author Keywords only – and vice versa, with the same method used for bibliometric research on Arctic papers.

Designing an expanded Antarctic search has highlighted issues with false positives, and shows the complexity involved in establishing what exactly constitutes "Antarctic" science, and what should be included (or excluded) as relevant. This is particularly an issue if expanding the search further to include sub-Antarctic areas.

It also highlights the tension between research "on Antarctica" – which may now be done by analysing third-party data, or from satellite observations, without leaving an office in the northern hemisphere – and research "in Antarctica", which may be done there for practical reasons but have a completely different focus – for example, the astronomical observations carried out at the South Pole. Traditionally, we would assume "Antarctic research" fulfilled both criteria; whether we should include work that only fulfils one is a more subjective question and it is challenging to provide an unambiguous answer for all cases.

The second part of the paper updates data presented in earlier PLC papers. The long-term trends continue – an increase in author numbers, an increase in open access rates – though the long-term increase in numbers may have stalled somewhat after 2021. It will remain to be seen if this is a long-term shift or just a temporary anomaly after the pandemic.

The most interesting factor, however, is the continued decline in relative citation impact rates. This has been noted before – as has a corresponding decline amongst Arctic papers – but without any clear explanation for why this is the case. We have suggested some potential explanations for why this might be the case, but this will need further and more in-depth examination to determine which – if any – of those factors are having a noticeable impact, and how they interact.

References

Ager, BJ & Gray, AD (tbc). Collaborations in polar research: a bibliometric study. In: xx (eds.) Polar Libraries Colloquy 2022, *Proceedings*. *Sharing Polar Cultures and Knowledge: Perspectives from Libraries and Archives* [in press]

Ahrabiran, K et al (2024). Surprising Resilience of Science During a Global Pandemic: A Large-Scale Descriptive Analysis. [preprint]. https://arxiv.org/abs/2409.07710v1

Aksnes, DW & Hessen, DO (2009). The Structure and Development of Polar Research (1981–2007): a Publication-Based Approach. *Arctic, Antarctic, and Alpine Research*, 41:2, 155-163, https://doi.org/10.1657/1938-4246-41.2.155

Clarivate (2022). Journal Citation Reports: Online-Early, Ahead of Print, and In-Press Articles, May 20, 2022. https://support.clarivate.com/ScientificandAcademicResearch/s/article/Journal-Citation-Reports-Online-Early-Ahead-of-Print-and-In-Press-Articles?language=en_US

Dastidar, PG (2007). National and institutional productivity and collaboration in Antarctic science: An analysis of 25 years of journal publications (1980–2004). *Polar Research*, 26:175–180. https://doi.org/10.1111/j.1751-8369.2007.00017.x

Dudeney, JR, & Walton, DWH (2012). Leadership in politics and science within the Antarctic Treaty. *Polar Research*, 31:1. https://doi.org/10.3402/polar.v31i0.11075

Fu, HZ, & Ho, YS (2016). Highly cited Antarctic articles using Science Citation Index Expanded: A bibliometric analysis. *Scientometrics*, 109(1):337–357. https://doi.org/10.1007/s11192-016-1992-4

Garfield, E (1990). Current Comments, 32:3-7. https://garfield.library.upenn.edu/essays/v13p295y1990.pdf

Garfield, E & Sher, IH (1993). KeyWords Plus™Algorithmic Derivative Indexing*, Journal of the American Society for Information Science*, 44(5):298-299. https://garfield.library.upenn.edu/papers/jasis44(5)p298y1993.html

González-Aravena M, Krüger L, Rebolledo L, et al (2023). Antarctic science in Chile: a bibliometric analysis of scientific productivity during the 2009–2019 period. *Antarctic Science*, 35(1):46-59. https://doi.org/10.1017/S0954102022000487

Gray, AD (2019). Bibliometrics for studying polar research. In: Hallikainen, Liisa; Parikka, Susanna, (eds.) Polar Libraries Colloquy 2018, *Proceedings. Developing polar networks: Ideas and possibilities for the future*. Rovaniemi, University of Lapland. https://lauda.ulapland.fi/handle/10024/63992

Gray, AD (2023). Changing patterns of polar research. In: Carle, Daria, (ed.) Polar Libraries Colloquy 2016, *Proceedings. Mapping Change: Chitduug Gha Nen' Ch'etnetl'edz*. https://scholarworks.alaska.edu/handle/11122/13220

Gray, AD & Hughes, KA (2016). Demonstration of "substantial research activity" to acquire consultative status under the Antarctic Treaty. *Polar Research*, 35, 34061. https://doi.org/10.3402/polar.v35.34061

Hua, W, Li, Y & Yuan, S (2014). A quantitative analysis of Antarctic related articles in humanities and social sciences appearing in the world core journals. *Scientometrics*, 100, 273–286 (2014). https://doi.org/10.1007/s11192-013-1190-6

Jakab, M, Kittl, E & Kiesslich, T (2024). How many authors are (too) many? A retrospective, descriptive analysis of authorship in biomedical publications. *Scientometrics*, 129, 1299–1328. https://doi.org/10.1007/s11192-024-04928-1

Jang, D, Doh, S, & Choi, Y (2020). Networks of international co-authorship in journal articles about Antarctic research, 1998–2015. *Polar Research, 39*. https://doi.org/10.33265/polar.v39.3647

Ji, Q, Pang, X & Zhao, X (2014). A bibliometric analysis of research on Antarctica during 1993–2012. *Scientometrics*, 101(3):1925-1939. https://doi.org/10.1007/s11192-014-1332-5

Kim, H & Jung, W-S (2016). Bibliometric Analysis of Collaboration Network and the Role of Research Station in Antarctic Science. Industrial Engineering and Management Systems, 15(1), 92-98. https://doi.org/10.7232/iems.2016.15.1.092

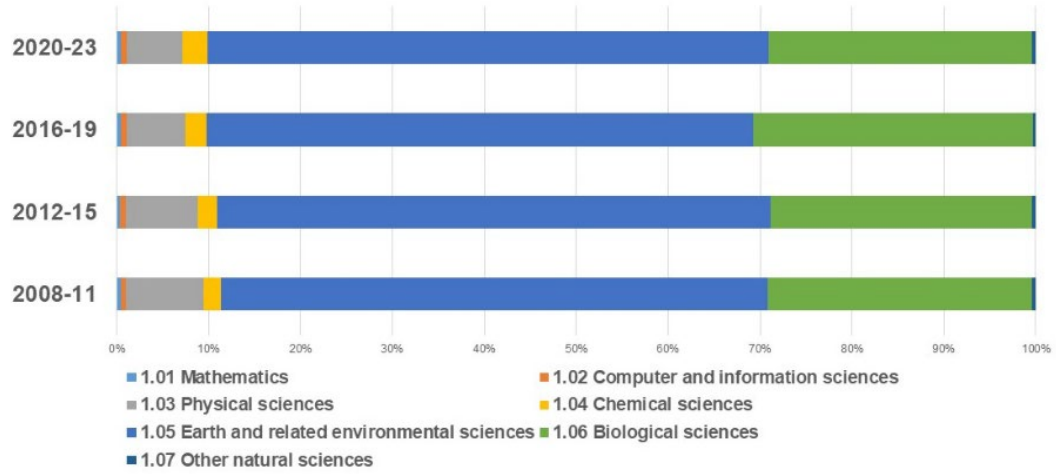
Stefenon, VM, Roesch, LFW & Pereira, AB (2013). Thirty years of Brazilian research in Antarctica: ups, downs and perspectives. *Scientometrics* 95, 325-331. https://doi.org/10.1007/s11192-012-0809-3

Xavier, JC, Gray, AD & Hughes, KA (2018). The rise of Portuguese Antarctic research: implications for Portugal's status under the Antarctic Treaty. *Polar Record*, 54(1):11-17. https://doi.org/10.1017/S0032247417000626

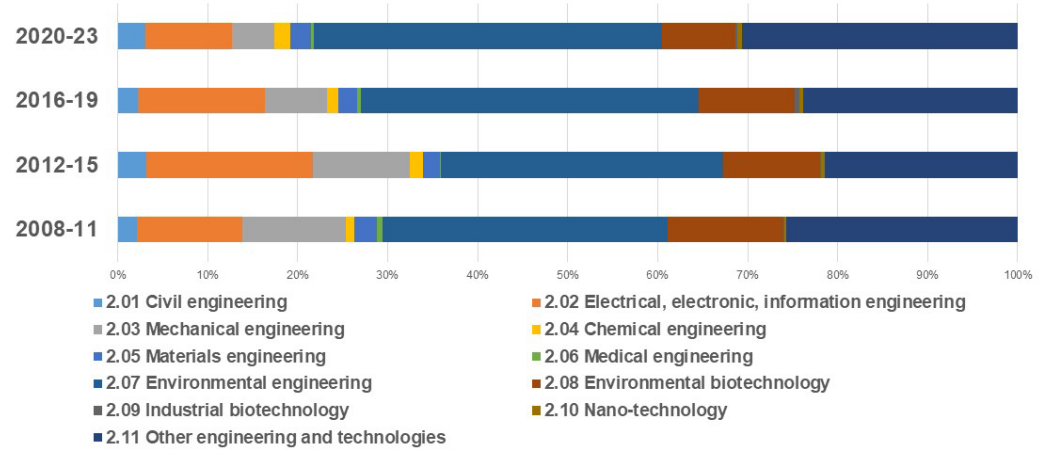
Zhang, Y, Zou, C, Peng, C, Lan, X & Zhang, H (2023). Geophysics in Antarctic Research: A Bibliometric Analysis. *Remote Sensing*, 15(16):3928. https://doi.org/10.3390/rs15163928

Appendix

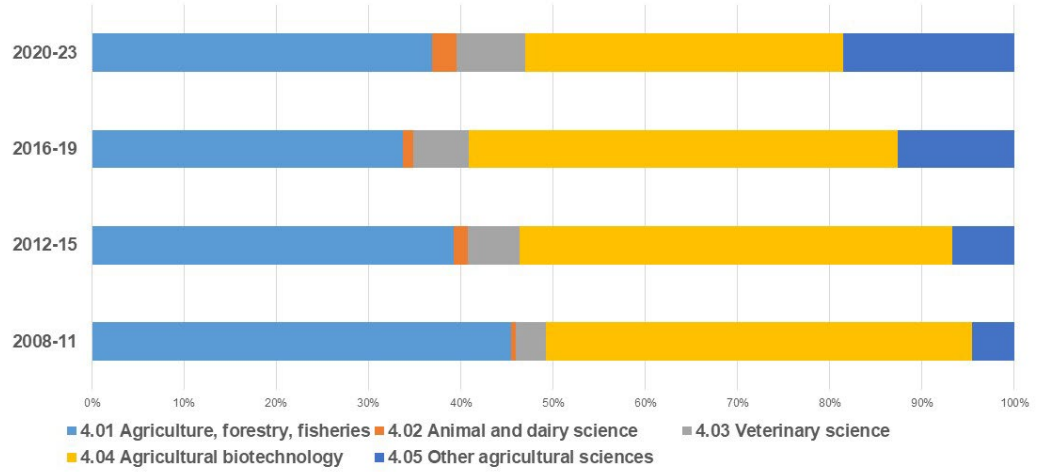
Share of papers – natural sciences (81%)



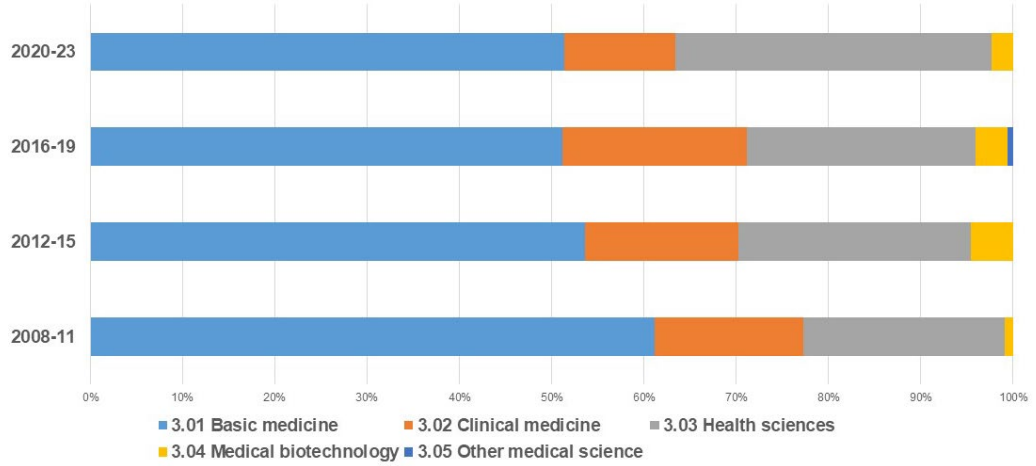
Share of papers – engineering (9%)



Share of papers – agriculture (4%)



Share of papers – medicine (2.5%)



Share of papers – social sciences (2.2%)