Bibliometrics for studying polar research

ANDREW GRAY
British Antarctic Survey,
Cambridge, United Kingdom

Abstract

Bibliometrics and scientometrics, the quantitative study of scientific and publishing activity, is a rapidly growing field. It can offer valuable insights into how scientific activity is structured, and the way knowledge develops over time. However, the application of bibliometric approaches to polar science can be challenging, as it is difficult to clearly identify as a fixed discipline. This paper reviews the particular challenges of polar bibliometrics and the ways in which we can best get accurate information on the field. It then sets out a short bibliometric study of recent developments in polar studies, comparing activity in the Arctic and Antarctic, which builds on a brief survey presented to the Colloquy in 2016.

Challenges of polar bibliometrics

A key problem in any bibliometric analysis is identifying the set of papers to be examined. This is particularly challenging for polar science, which is a broad, fuzzily-defined, and heterogenous field cutting across many formal disciplines, and so it is rare to see it explicitly identified by subject headings within a database.

The first method to do this is manual selection. This was essentially the approach used by the Bibliography on Cold Regions Science and Technology from 1951 onwards, and the Antarctic Bibliography from 1965 onwards; however, this approach is inherently expensive and time-consuming. With the growing availability and scope of general non-topical publication databases in the 1990s, it became possible to try and identify polar papers by searching within these, rather than relying on curated lists. Doing this well is more challenging and has a higher risk of false positives, but had the potential to produce more valuable results and to scale in a way that curated bibliographies cannot.

Using a database, a particularly simple approach would be to examine all papers published in ”polar publications” – this approach was used by eg Aksnes & Hessen 2009, to supplement a keyword search. However, there are only around a dozen identifiably polar titles, with the vast majority of polar science now published in non-polar disciplinary journals. This approach would only ever find a small fraction of the published literature, though it has some promise as a way of validating other search approaches.

A third approach is to examine the affiliation of papers – it might be a reasonable approach to assume that all papers published by ”an Antarctic institute” are relevant. However, this is not the case, with many seemingly specialised institutions producing a substantial amount of non-polar work. In addition, the majority of polar research is produced by researchers at universities or ”nonpolar” research institutes, who would of course not be identified by this method.

A fourth approach is to rely on topic indexing in the database. The Antarctic & Cold Regions bibliographies, of course, managed this by default. Web of Science and Scopus index papers by subject, which is generally quite high-level and derived from the topic of the journal. This, as noted above, will not work for our purposes. Dimensions contains a subject index derived from the contents of the paper, but this is still high-level and does not contain polar science.

This may improve in the future, however. A recent project has algorithmically clustered papers in Scopus into around 96,000 ”topics of prominence”, stable groups based on citation connections and each presumed to represent a discrete narrow field of enquiry – perhaps of around 100 researchers and a few hundred papers. These can be summarised by notional names such as ”role of nursing in clinical trials” or ”properties of olive extracts”. (Klavans & Boyack, 2017) It is highly likely that a number of these groups are identifiably Arctic or Antarctic in orientation – perhaps 25-50 of them – and this would offer a new avenue for identifying relevant research for analysis. However, these groupings are not yet available in Scopus, but they can be accessed through the (separate) SciVal research assessment tool. This is a very promising development and it will be interesting to see what emerges in future.

In general, however, in the absence of reliable indexed topic grouping, we have to fall back on the fifth option, keyword searching. This is overwhelmingly the most common approach used for Antarctic research and is likely to dominate Arctic research as well.

Search terms

At the 2016 colloquy, I presented a survey on Arctic and Antarctic bibliometrics based on minimal search keywords (Gray 2016). Later work (Gray & Hughes 2016) indicated that these terms could be misleading in some circumstances, and that more work was needed to establish an accurate set of keywords. In a review of keyword techniques used for Antarctic bibliometrics – currently in preparation – I identified a hopefully comprehensive search term (this using Web of Science syntax):

```
TS=((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")
```

In comparison to this relatively straightforward search, the complexity of assigning Arctic search terms is well known; see, eg, Campbell (2014) which identified a search term requiring around 200 terms purely to look at Indigenous subjects in northern Canada. In the absence of detailed subject knowledge with which to build such a query from first principles, a broad search was derived from the keyword list in Aksnes & Hessen (2009):

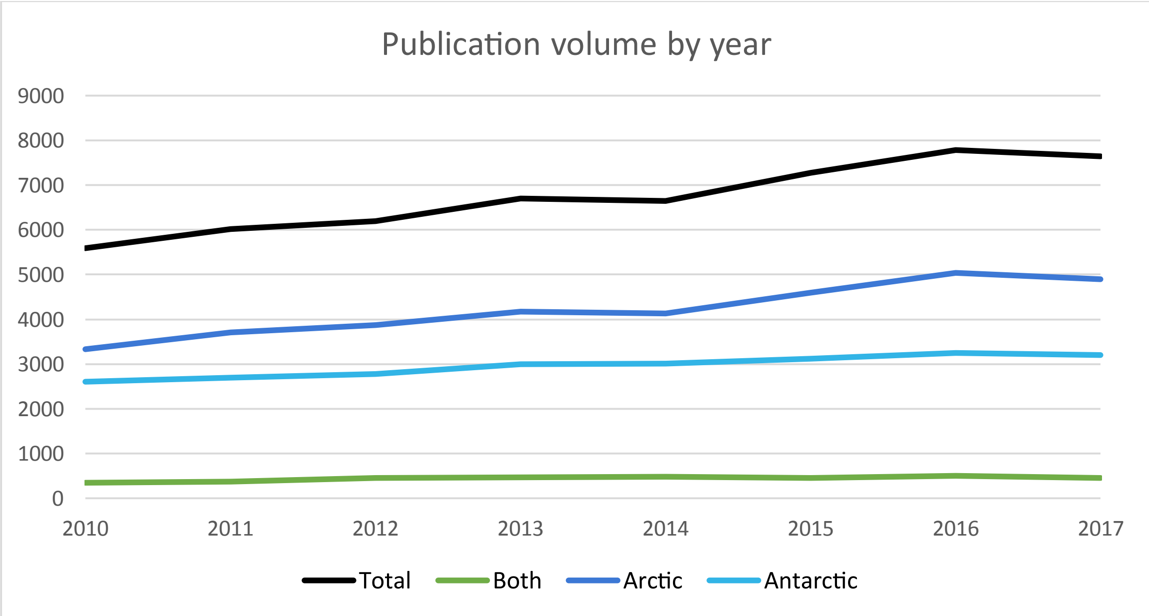
```
TS=(("Arctic" NOT "arctic bramble" NOT "sub-Arctic") OR "Svalbard" OR "Spitsbergen" OR "Longyearbyen" OR "Ny-Alesund" OR "Hornsund" OR "Barentsburg" OR "Kongsfjord" OR "Hopen" OR "Bjornoya" OR "Bear Island" OR "Greenland" OR "Baffin Island" OR "Queen Elizabeth Islands" OR "Ellesmere Island" OR "Devon Island" OR "Somerset Island" OR "Prince of Wales Island" OR "Banks Island" OR "Ellef Ringnes Island" OR "Amund Ringnes Island" OR "Bathurst Island" OR "Axel Heiberg Island" OR "Prince Patrick Island" OR "King William Island" OR "Prince Charles Island" OR "Bylot Island" OR "Bathurst Island" OR "Southampton Island" OR "Brooks Range" OR "St Lawrence Island" OR "St Matthew Island" OR "Seward Peninsula" OR "Nunivak Island" OR "Novaya Zemlja" OR "Severnaja Zemlja" OR "Novosibirskije Ostrova" OR "Jan Mayen" OR "Victoria Islands" OR "Nunavut" OR "Fram Strait" OR "Beaufort Sea" OR "Davis Strait" OR "Barents Sea" OR "Kara Sea" OR "Storfjorden" OR "Baffin" OR "Hudson Bay" OR "Siberian Sea" OR "Laptev Sea" OR "Chukchi Sea" OR "Bering Strait" OR "Bering Sea" OR "Karskoje Sea")
```

As can be seen, the larger Arctic search reflects the much more fragmented nature of Arctic geography. It could in theory be substantially more detailed – eg Aksnes et al (2016) reported the use of 350 geographical terms and 225 names for indigenous groups – but it is likely that this query covers most target papers. Further work to produce and publish a more detailed Arctic search term to serve as a baseline for bibliometric analysis would be worthwhile.

All data was gathered from Web of Science with one of the two search terms above, filtered to only articles or reviews, published between 2010 and 2017 inclusive. Initial assessment and data analysis was carried out through InCites; this caused a small number of papers to be omitted as the databases used for Web of Science and InCites are not completely identical.

Yearly changes

It is apparent that the total number of papers published in polar science continues to rise, with an average increase of around 5-6% year-on-year in Arctic research and 4% in Antarctic research. This is a substantial and respectable growth, although perhaps lower than the overall increase in scientific activity (estimated at 8-9% per year; Bornmann & Mutz 2015). The volume of research which is identifiably "bipolar" is strikingly low, less than 10% of the overall total, although it appears to be growing at a similar rate.



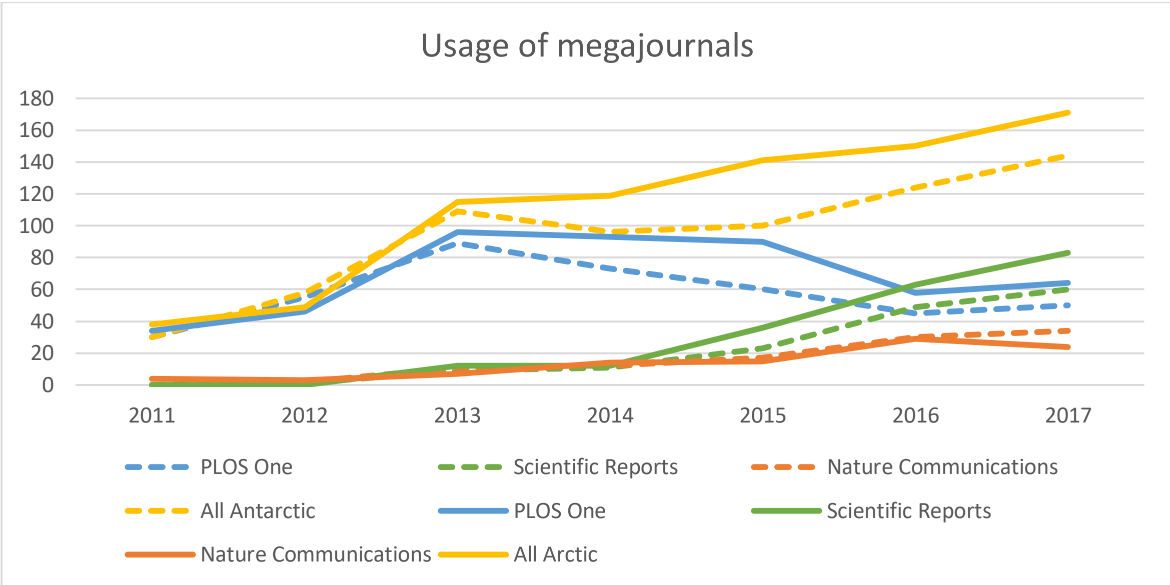
Distribution by journal

3115 titles are represented in the Arctic data, and 2325 in the Antarctic. Both are highly clustered in the most heavily-used journals; the top 20 Antarctic titles cover 30% of publications, and the top 20 Arctic titles cover 25%. However, as noted earlier, few of these are purely "polar". The majority of highly-used papers are published in general disciplinary journals such as Geophysical Research Letters or the various parts of the Journal of Geophysical Research, but there are some exceptions - Polar Biology is the most commonly used Antarctic title and the second most common Arctic title, and Antarctic Science is the third most commonly used Antarctic title. Arctic, Polar Research, Polar Record, Polar Science, and Polish Polar Research all have over a hundred papers and appear in the top fifty for one or both of the regions. The table below shows the top 20 journals, 2011-2017, with specialist "polar" titles italicised.

	Arctic		**Antarctic**	
1	Geophysical Research Letters	770	*Polar Biology*	748
2	*Polar Biology*	638	Geophysical Research Letters	632
3	JGR Atmospheres	620	*Antarctic Science*	490
4	JGR Oceans	578	JGR Oceans	474
5	Quaternary Science Reviews	548	PLOS One	426
6	Atmospheric Chemistry & Physics	547	Deep Sea Research II	396
7	Journal of Climate	526	JGR Atmospheres	360
8	PLOS One	499	Quaternary Science Reviews	346
9	Cryosphere	490	Earth and Planetary Science Letters	339
10	Biogeosciences	325	Journal of Climate	332
11	Climate Dynamics	318	Cryosphere	309
12	Journal of Glaciology	300	Journal of Glaciology	298

13	Marine Ecology Progress Series	296	Atmospheric Chemistry & Physics	275
14	Deep Sea Research II	288	Journal of Physical Oceanography	273
15	Climate of the Past	265	Marine Ecology Progress Series	262
16	*Arctic*	263	Biogeosciences	234
16	Environmental Science & Technology	263	Annals of Glaciology	213
18	Earth and Planetary Science Letters	261	Paleoceanography	209
19	Environmental Research Letters	219	Climate of the Past	202
20	Palaeogeography, Palaeoclimatology, Palaeoecology	216	Climate Dynamics	178

It is interesting to note that the number of papers published in the non-disciplinary ”megajournals” has increased about fivefold between 2011 and 2017 in both regions, with a similar pattern in both – the early dominance of PLOS One was eroded in 2015-17, with Scientific Reports becoming the most popular megajournal. As of 2017, these three megajournals now represent 3.5% of Arctic papers and 4.5% of Antarctic papers; Scientific Reports was the fourth most heavily used journal for both Arctic and Antarctic science, and PLOS One the sixth.



All three journals are, of course, fully open access. The most heavily used polar journals in 2017 included two others which are fully open access (Cryosphere, Atmospheric Chemistry & Physics), and almost all the heavily-used journals offer optional “hybrid” per-article open access, allow self-archiving of author’s manuscripts in institutional repositories, or both.

Across all journals, the situation for open access looks positive. The article-level open-access data available through Web of Science is not immensely reliable, but it appears to indicate that around 30-40% of polar material is available as open access (freely available, immediately after publication, on the journal website) or “bronze open access” (freely available from the publisher, after an embargo period or with no clear license). This has increased slightly over the past few years but has not had dramatic changes. Data for green open access (provided through an institutional repository, possibly after a short embargo period) is patchy but a reasonable estimate might be that somewhere over 50% of material, in total, is freely available in some way.

The heavy role of the geosciences in polar research gives an added boost to free access via the "bronze" route, though it should fairly be noted that there is some dispute over whether this properly constitutes "open access" in the strict sense of the term. Several of the most commonly used journals are published by either the American Meteorological Society or the American Geophysical Union. These two publishers make all journal articles free to read within two years (1 year for the AMS, 2 years for the AGU), in a way that is relatively common in medical research but rare in other fields.

Distribution by field

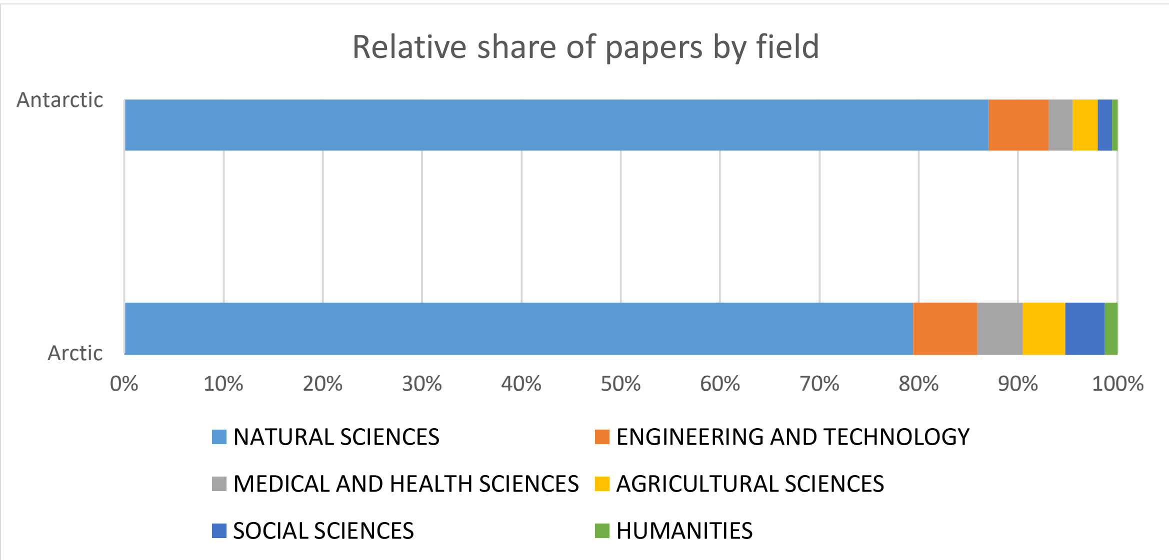
Overall, Arctic and Antarctic science have a broadly similar distribution. Both are heavily skewed towards the natural sciences, representing 79% of Arctic research and 87% of Antarctic research. Applied research in engineering and technology represents about 6% of both. (Fields here are drawn from the OECD definitions; papers are assigned to one field only).

Medical research, and agricultural research, each represent about 4.5% of Arctic work but only 2.5% of Antarctic work. Similarly, the social sciences and humanities are more pronounced in the Arctic – together they are around 5% of papers as opposed to 2%.

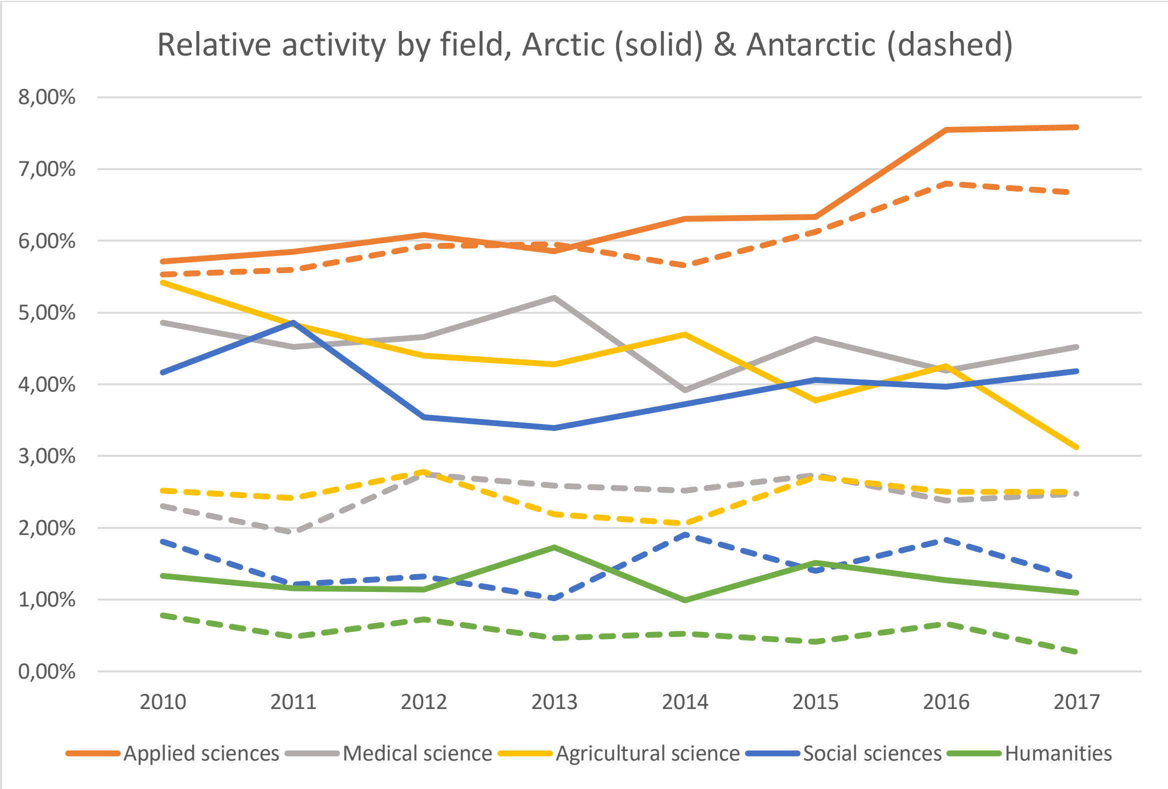
Breaking down the broad heading of the natural sciences, about 60% of research in both areas is on the earth and environmental sciences, but twice as much Antarctic research is on fundamental science such as chemistry, physics, or astronomy, and around 30% on biological sciences as opposed to 25% in the Arctic.

Arctic applied science & engineering is substantially skewed towards environmental engineering, while Antarctic science includes a substantial amount of environmental biotechnology not found in the Arctic. Basic medical research is equally common in both, but applied clinical medicine is three times as common in the Arctic, and health sciences four times as common.

In the social sciences and humanities, an unusual discipline is psychology, where the share of papers in Antarctic research is almost as high as in the Arctic. This can perhaps be attributed to the specialist human-factors research that continues to be done in Antarctica, looking at the dynamics of small and highly isolated groups.



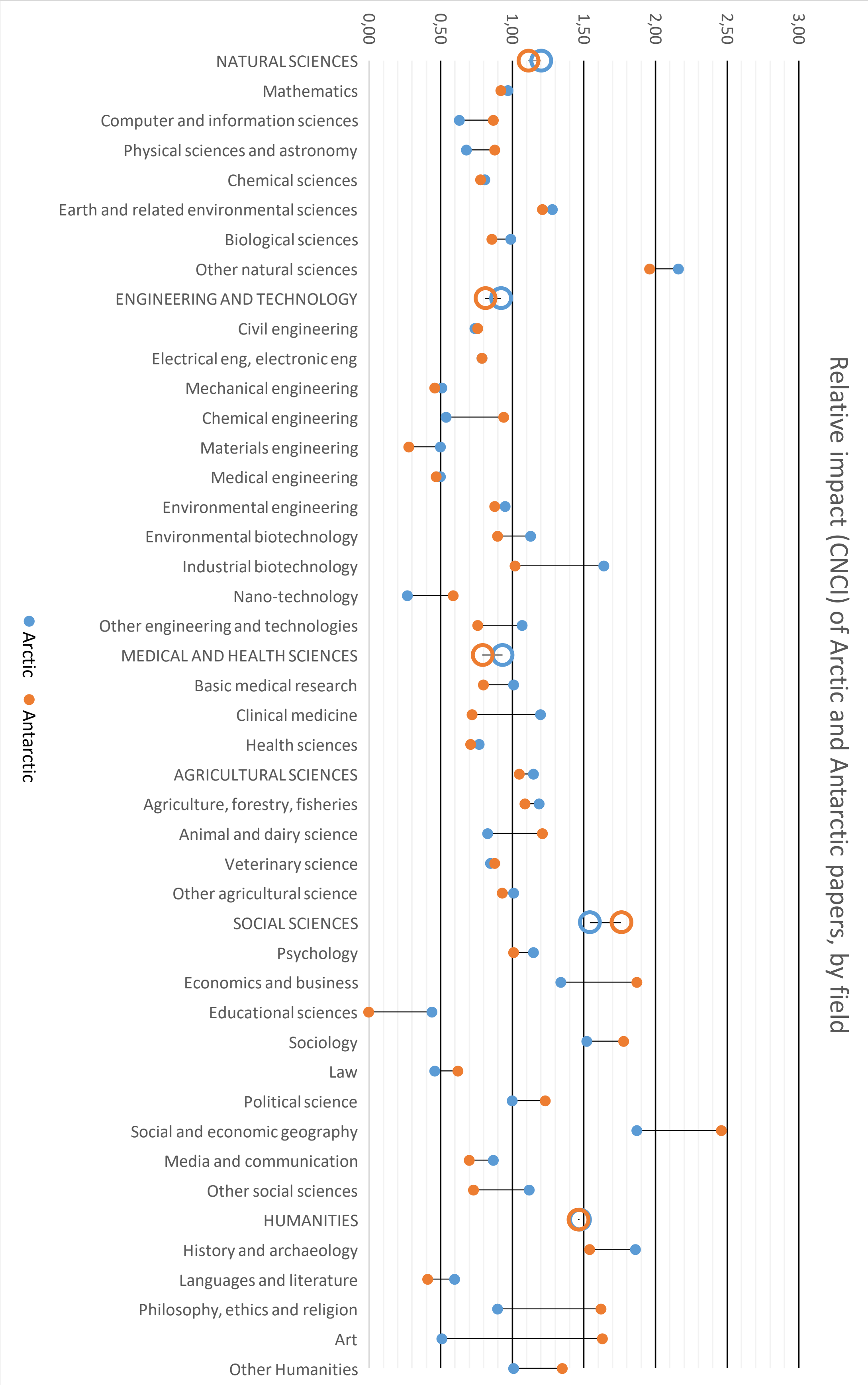
A more detailed set of topic breakdowns can be found in the Appendix.



This graph omits the natural sciences, which remain approximately equal throughout the period, but shows the relative change in different disciplines. A distinctive factor is the falling off of agricultural sciences in the Arctic, and the general growth of the applied sciences (engineering and technology) in both regions. The volume of Antarctic humanities work continues to decline.

In terms of research quality, both Arctic and Antarctic science are in general somewhat above-average. Arctic science has an averaged category-normalised citation impact (CNCI) of 1.18 - that is to say, identifiably Arctic papers have an average of 1.18 times as many citations as the average for papers in that discipline published in the same year. Antarctic papers have a mean CNCI of 1.09. (These figures are taken from the CNCI of all papers by discipline, weighted by the number of papers in each).

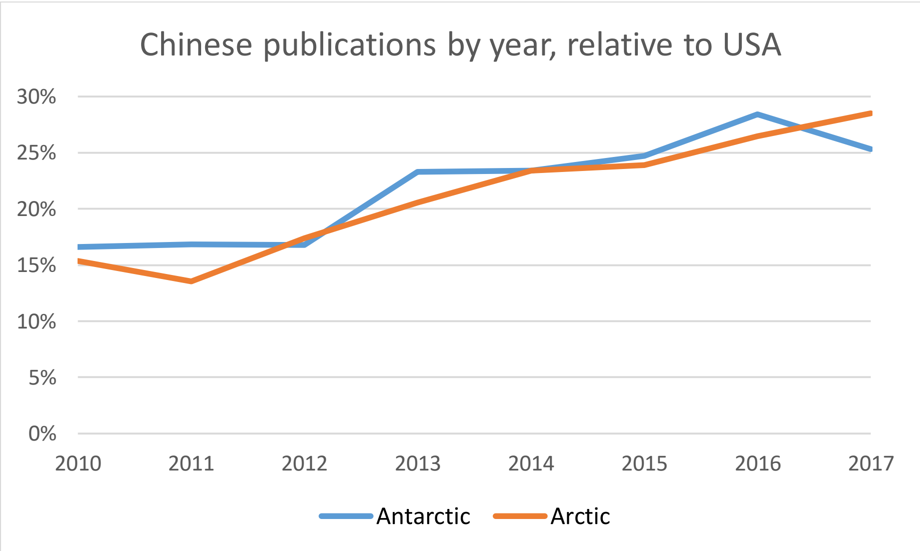
The earth sciences is strong in both regions, while the biological sciences is about average (Arctic) or below average (Antarctic). One of the strongest Arctic topics is social and economic geography, which has a CNCI of 1.87 – almost twice the average for the field. Perhaps unsurprisingly, this compares sharply with 0.87 and substantially fewer papers in the Antarctic.


Relative impact (CNCI) of Arctic and Antarctic papers, by field
0,00
0,50
1,00
1,50
2,00
2,50
3,00
NATURAL SCIENCES
Mathematics
Computer and information sciences
Physical sciences and astronomy
Chemical sciences
Earth and related environmental sciences
Biological sciences
Other natural sciences
ENGINEERING AND TECHNOLOGY
Civil engineering
Electrical eng, electronic eng
Mechanical engineering
Chemical engineering
Materials engineering
Medical engineering
Environmental engineering
Environmental biotechnology
Industrial biotechnology
Nano-technology
Other engineering and technologies
MEDICAL AND HEALTH SCIENCES
Basic medical research
Clinical medicine
Health sciences
AGRICULTURAL SCIENCES
Agriculture, forestry, fisheries
Animal and dairy science
Veterinary science
Other agricultural science
SOCIAL SCIENCES
Psychology
Economics and business
Educational sciences
Sociology
Law
Political science
Social and economic geography
Media and communication
Other social sciences
HUMANITIES
History and archaeology
Languages and literature
Philosophy, ethics and religion
Art
Other Humanities
Arctic
Antarctic

Country of origin

The lists of most prolific countries begin predictably. The United States heads both lists. In the Antarctic, it is followed by a series of countries with historically strong Antarctic programs (UK, Germany, Australia, France) before the first "new" countries, China and Canada. The substantial growth of Chinese polar science in recent years has been noted before, but the high showing for Canada is quite striking. It had been identified by earlier work (Gray 2016) as having unusually high levels of Antarctic activity for a country with no fixed infrastructure and not closely engaged with the Antarctic Treaty system; it seems to be maintaining this high level of output.

In the Arctic, the United States is followed by Canada, the United Kingdom, Norway, Germany, and Russia. All have historically strong Arctic research interests. China, again, shows up close behind these, with a rapid growth of publications. The graph below, which shows Chinese research output in comparison to the United States, demonstrates the rapid relative growth in both Arctic and Antarctic research – in comparative terms, output has almost doubled over the eight years.



This dramatic growth in output appears to have some corresponding cost in quality. The weighted average CNCI for Chinese papers in 2010-12 was 1.08 (Antarctic) and 1.32 (Arctic). In 2015-17, the weighted average was 0.8 and 1.03 respectively, a significant drop in both fields. By comparison, the impact for Western scientific output was much more stable, showing a much less marked decline (1.3 to 1.29 and 1.38 to 1.29 for all OECD countries). This suggests that the Chinese program, while publishing large amounts of research, is doing so with reduced overall scientific impact – quantity over quality?

It should, however, be noted that there are substantial known gaps in the major citation databases for non-Western publications. This may serve both to reduce the number of Chinese publications, but also to reduce the number of citations to them – assuming that Chinese-language publications are more likely to cite other Chinese-language publications, omitting a significant number from the database could have disproportionate effects on the citation counts of those that remain. A similar problem affects Russian-language publications, which are not consistently covered by Web of Science (Moed et al, 2018) – it is noticeable that Russia is the other productive country with unusually low CNCI values. These should, thus, be treated with a grain of salt.

From a regional perspective, European Union member states account for 45.4% of all Antarctic papers, and 41.2% of Arctic papers; Nordic countries are responsible for 7.8% and 24.1% respectively.

Leading institutions

The Russian Academy of Sciences is notionally the largest Arctic research institution, with 5.57% of papers published, but in practice this reflects a large number of individual institutions being grouped together. Barring such large composite bodies, the biggest individual institutions are the University of Alaska Fairbanks (3.96% of papers), NOAA (3.54%), the Arctic University of Tromso (3.39%), the Alfred-Wegener Institute (3.37%), and the University of Copenhagen (2.89%).

Among those with more than 0.5% of all Arctic papers, the highest-impact institutions by CNCI are the (US) National Centre for Atmospheric Research, the University of California Irvine, the University of Exeter, the University of California Berkeley, and the Ohio State University. All have a mean CNCI of 2.7 or above, and in addition have a substantial share of very highly-cited papers – 7-10% of their publications are in the top 1% of papers by citation in their field.

In the Antarctic, the largest individual institution is the British Antarctic Survey, with 6.05% of all papers, followed by the Alfred-Wegener Institute (4.25%), the University of Cambridge (3.39%, predominantly but not completely from the Scott Polar Research Institute), the University of Tasmania (3.18%, closely linked to the Australian Antarctic Division), and NASA (2.79%). The concentration of papers at BAS is remarkable and does not have a close Arctic analogue.

Among those institutions with more than 0.5% of all Antarctic papers, the highest impact by CNCI was the University of California Irvine, then the Jet Propulsion Laboratory, the (UK) Met Office, the University of East Anglia, and the (US) National Centre for Atmospheric Research. Again, 8.5-11% of their publications are in the top 1% of papers by citation in their field, showing both broadly high-quality work as well as a strong share of the most significant research.

The Antarctic work has a distinctive group of institutions focused on modelling and climate work, reflecting the prominence of this field within Antarctic research.

National focus

A useful measure to consider the significance of polar research is to look at the intensity of it within a given country – the proportion of a country's research which is identifiably oriented towards that topic. This allows us to distinguish between a very large country and a smaller one which has a much stronger national focus on polar science; an assessment based simply on publication numbers could mask the importance of the subject in smaller countries.

By far the country with the highest Arctic research intensity is Greenland, (around 72% of papers were Arctic-related) which can be in part explained by very small numbers of publications, around 100/year, but mostly by the fact that Greenland itself was in fact one of the keywords used in our searches. It is perhaps unsurprising that so many papers were returned.

Leaving Greenland aside as a special case the next most prominent Arctic countries are Iceland, Norway, Denmark, Canada, Russia, Finland, and Sweden - all of the Arctic nations save the United States, which is a special case and less explicitly Arctic-oriented than the others. The intensity for the Arctic nations ranges from 5.17% in Iceland to 0.96% in Russia. The first non-Arctic state is Estonia (0.85%), then the United Kingdom (0.50%).

(If we were to consider only papers with an affiliation within Alaska, we do find an Arctic intensity of around 23%. However, Alaska represents only 0.28% of the scientific output of the United States, so the results are dominated by non-Arctic regions; the national figure is 0.31%)

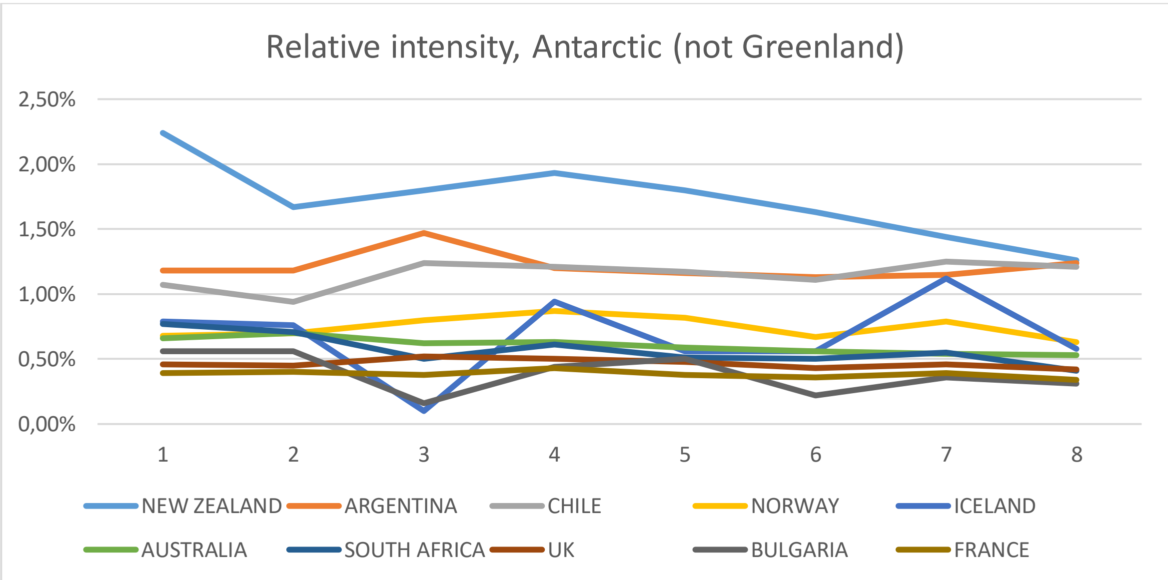
There is no dramatic change in intensity over the period of study for any of the major Arctic countries, though further down the list there are some interesting signs of change – Poland is steadily increasing (up from 0.36% to 0.72%), as is the Czech Republic (0.21% to 0.4%)

In the Antarctic, again Greenland is one of the highest, with an intensity of between 1.9% and 9.6%. Leaving this aside as a special case, the highest intensity is New Zealand (1.7%), Argentina (1.2%), Chile (1.15%), Norway (0.75%), Iceland (0.68%), Australia (0.60%), South Afica (0.57%), and the UK (0.47%).

As noted in previous work, this shows a strong investment in Antarctic research by New Zealand, Argentina, and Chile in particular, but also by a group of other nations with long-term commitments to the continent. Iceland is an unusual appearance here; this may well represent Icelandic collaboration on general polar science rather than explicit Icelandic commitment to Antarctic research. (The same is likely true of Greenland)

Russia (0.26%), the United States (0.23%), and China (0.1%), all identified as active Antarctic nations, are substantially lower. It is interesting to highlight a surprisingly high intensity on the part of Bulgaria, with an overall intensity of 0.39%, just below the UK, despite an Antarctic program

There are some interesting signs here of systemic change. New Zealand has dropped from an intensity of 2.24% (2010) to 1.26% (2017). South Africa and Australia are also reducing their focus on Antarctic research, albeit more slowly. Bulgaria, a small nation with an unexpectedly strong Antarctic program, is also reducing its intensity. For Argentina it remains generally stable, and for Chile may be increasing slightly.



Changes in intensity may reflect a shift in national priorities or an explicit change in funding for a particular research topic, but may also simply be driven by broader changes. In the case of New Zealand, for example, the total level of Antarctic research has remained approximately steady, but the overall level of national research output has increased by about 35%. Likewise, the level of Chinese research on both Arctic and Antarctic topics has grown at much the same rate as Chinese science overall, meaning that a dramatic growth in real terms has not represented a noticeable shift in priorities.

References

AKSNES & HESSEN (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

AKSNES ET AL (2016). *Arctic Research Publication Trends: A Pilot Study*. https://dspace.spbu.ru/bitstream/11701/5555/1/Arctic%20Research-PublicationTrends%20Final%20version.pdf

BORNMANN & MUTZ (2015). *Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references*. Journal of the Association for Information Science & Technology, 66(11), 2215-2222 https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.23329

CAMPBELL (2014). *Analyzing Published Health Research Related to Indigenous Peoples in Northern Canada: explorations using a sensitive search filter*. Proceedings of the 25th Polar Libraries Colloquy. http://nora.nerc.ac.uk/id/eprint/513471/

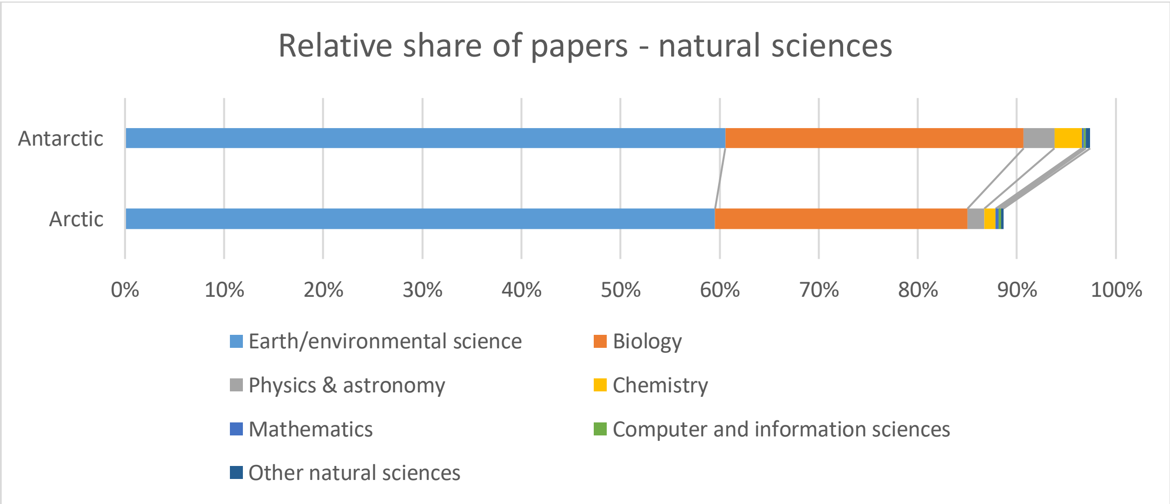
GRAY (2016). *Changing patterns of polar research*. Proceedings of the 26th Polar Libraries Colloquy [in press]

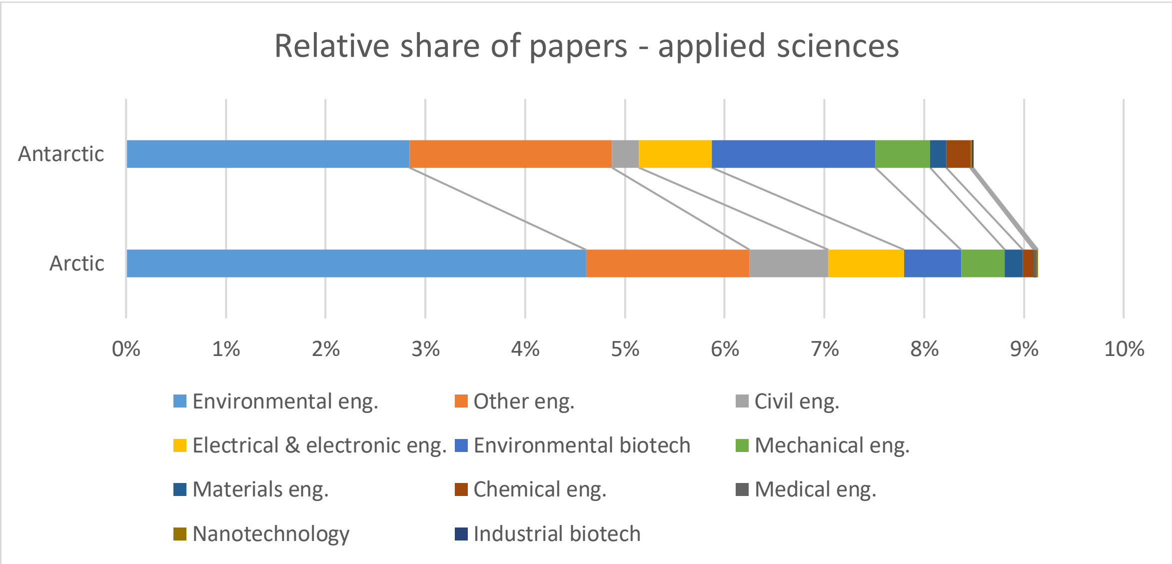
GRAY & HUGHES (2016). *Demonstration of "substantial research activity" to acquire consultative status under the Antarctic Treaty*. Polar Research, 35, 34061. https://doi.org/10.3402/polar.v35.34061

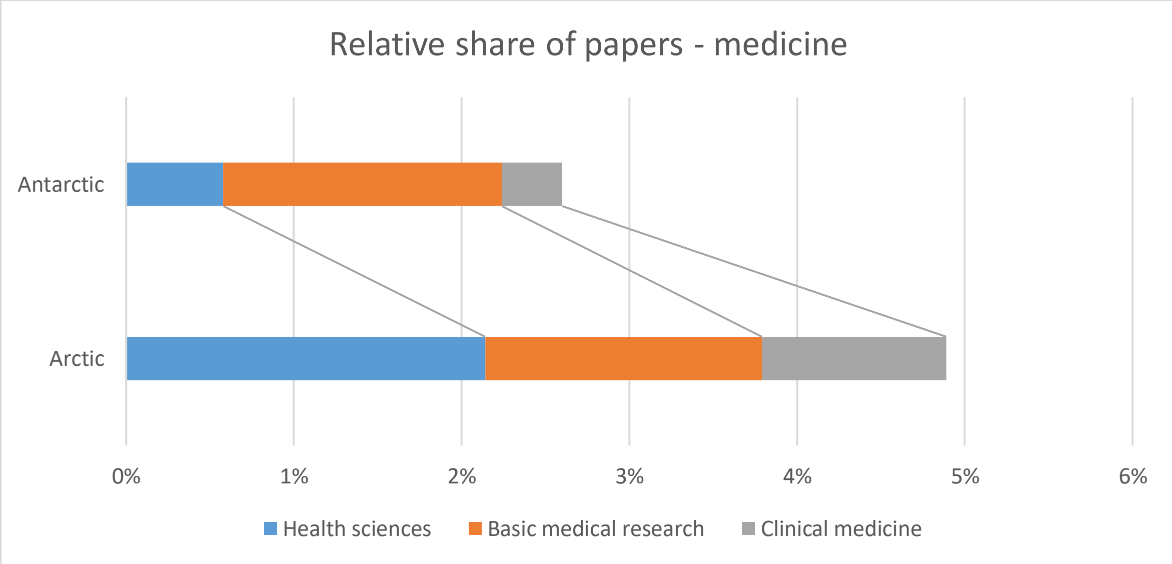
MOED ET AL (2018). *Trends in Russian research output indexed in Scopus and Web of Science*. [arXiv preprint] https://arxiv.org/abs/1805.02434

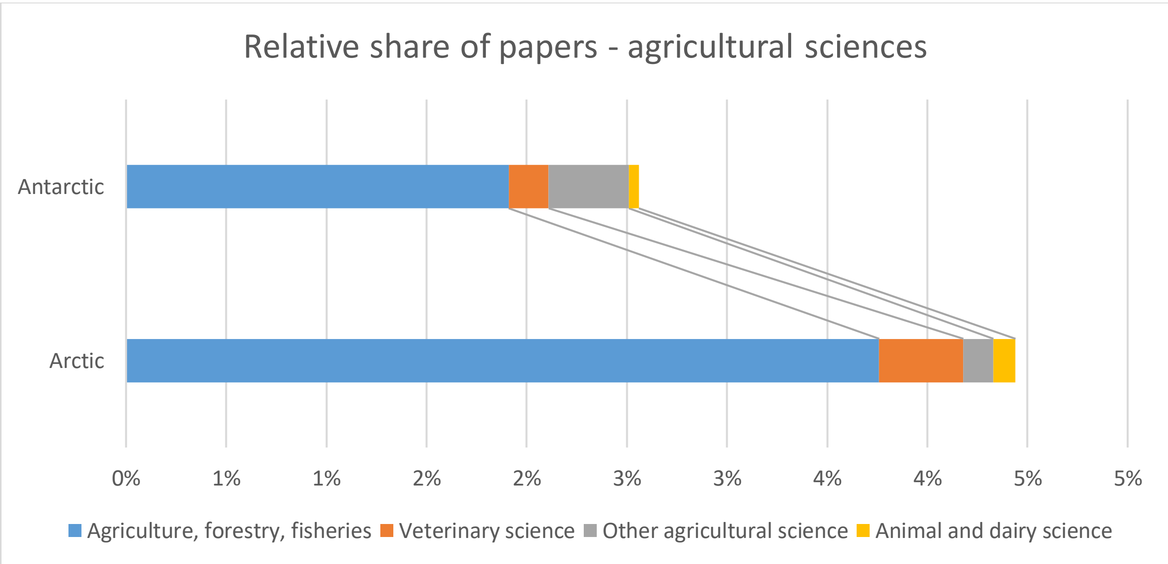
Appendix

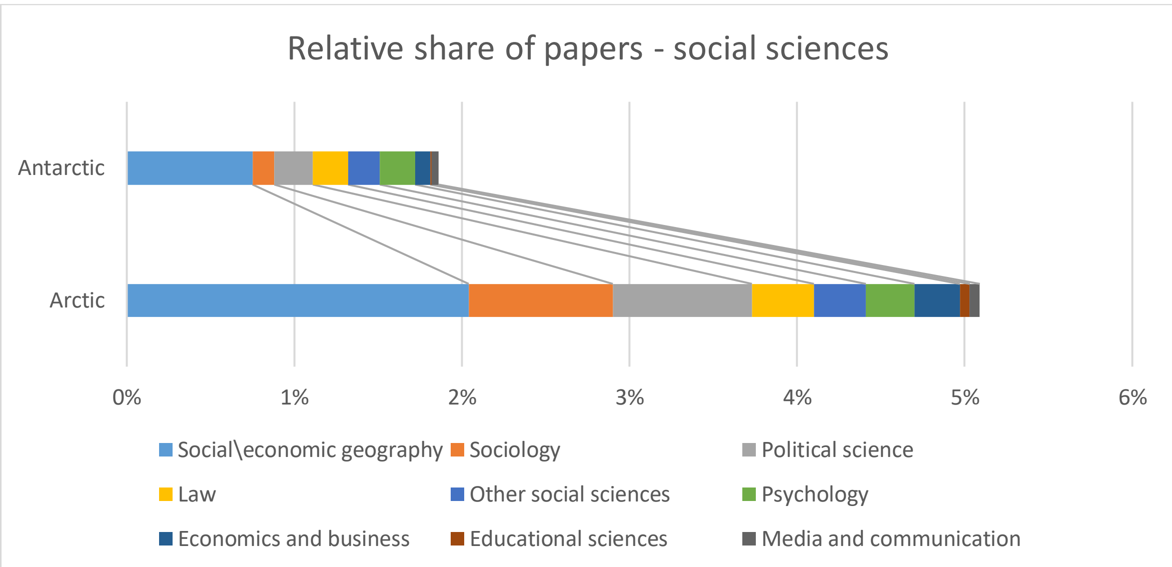
Relative breakdowns of research by field, 2010–17.

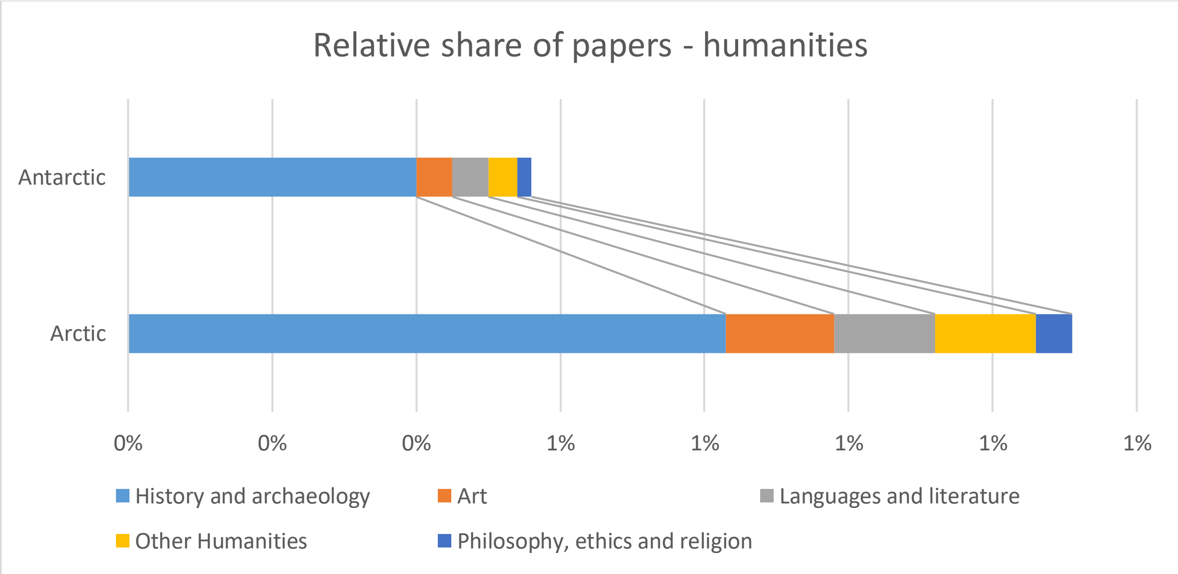





Relative share of papers - medicine
Antarctic
Arctic
0%
1%
2%
3%
4%
5%
6%
Health sciences
Basic medical research
Clinical medicine


Relative share of papers - agricultural sciences
Antarctic
Arctic
0%
1%
1%
2%
2%
3%
3%
4%
4%
5%
5%
Agriculture, forestry, fisheries
Veterinary science
Other agricultural science
Animal and dairy science


Relative share of papers - social sciences
Antarctic
Arctic
0%
1%
2%
3%
4%
5%
6%
Social\economic geography
Sociology
Political science
Law
Other social sciences
Psychology
Economics and business
Educational sciences
Media and communication


Relative share of papers - humanities
Antarctic
Arctic
0%
0%
0%
1%
1%
1%
1%
1%
History and archaeology
Art
Languages and literature
Other Humanities
Philosophy, ethics and religion