White Paper ID Number W014

Title of White Paper Canada's astronomy performance based on bibliometrics

ID of Associated Expression of Interest E002

Topic Area of White Paper state of the profession

Executive Summary of White Paper (5000 character limit)

Using bibliometrics this white paper shows:

- The observing facilities that Canada supports financially are amongst the highest performing, both in productivity and impact, of facilities world-wide.
- For most observing facilities, Canadian any author papers have higher impact per paper than that of the full bibliography of the facility.
- However, Canadian first author papers on most facilities have a lower impact per paper than that of the full bibliography of a facility
- Almost 80% of Canadian first author papers based on telescope data are authored by either students or PDFs.
- Using ADS numbers, for the period 2014 – 2017 Canada ranks second in the world in mean impact per paper when considering any author papers.
- However, when considering only Canadian first author papers, Canada ranks only 8th in the same metric.
- The number of Canadian first author papers has declined slightly since 2010 (based on ADS numbers), while the number has increased for most of our ‘competitors’.
- Based upon the publication history of Canadian faculty members, a person’s first author paper production peaks during their postdoc years and then declines once they have a faculty position.

Lead author and affiliation Dennis Crabtree - NRC Herzberg

Email address of lead author Dennis.Crabtree@nrc.ca

Canada’s Astronomy Performance Based on Bibliometrics

Dennis R. Crabtree – NRC Herzberg

Introduction

Bibliometrics in its most basic form is the simple counting and measurement of publications. Its core application is the quantification of scientific output (usually refereed publications) and using the results to measure the performance of people and countries and any aggregation in between these two extremes.

The use of bibliometrics became a standard evaluation tool with the explosion of scientific output following the Second World War. Today, with publication and citation data available online, the use of bibliometrics has exploded into almost all evaluations of the scientific enterprise. While the use of bibliometrics has weaknesses as an evaluation tool, it is a time-tested tool that can provide a quantitative measurement of research performance.

In this white paper I will evaluate various aspects of Canadian astronomy and the Canadian community though the use of bibliometrics. The NASA ADS has been used to gather all citation information in this work.

Canadian Offshore Observatories

Canada’s offshore facilities include the Canada-France-Hawaii Telescope (CFHT), the Gemini Observatory (Gemini) and the Atacama Large Millimeter Array (ALMA). For many years I have maintained a database of refereed publications based on data from the world’s major telescopes. In this section I compare the performance of Canada’s three facilities with the others for the most recent 5-year period in my database. Note that it takes time for the citation information to be usable, so 2018 papers are not included.

Figure 1 shows the productivity of each telescope in my database for the period 2013 – 2017. Each individual year is shown (left axis) as well as the total over the 5-year period (right axis). This shows the *productivity per telescope*. The raw productivity numbers for multi-telescope facilities such as Keck and Gemini are divided by the number of telescopes associated with the facility.

The three facilities in which Canada is a partner are all in the upper half of the distribution. The rapid yearly increase in the number of papers from ALMA is evident. The ramp-up in the productivity of a new facility is normal. Based upon previous telescopes it takes 8-10 years before productivity reaches a plateau.

CFHT is the 4th most productive facility – not bad for a 40-year old facility that no longer is amongst the largest aperture in the world. The high productivity of CFHT is due to its wide-field imager, MEGACAM, and the availability of archival data through the CADC. Approximately 50% of CFHT papers include the use of archival data, mostly MEGACAM data. The CADC provides efficient and easy access to the archival data, and the wide-field imaging data is very useful for archival research.

Gemini’s productivity is lower than most of the 8-m class telescopes. Many factors can affect a telescope’s productivity including, but not limited to, the quality of the site, quality of the telescope, quality of the instruments, alignment of the instruments with the users' interests, and the makeup of the user community. It is likely that all of these factors are contributing to Gemini's current productivity level.

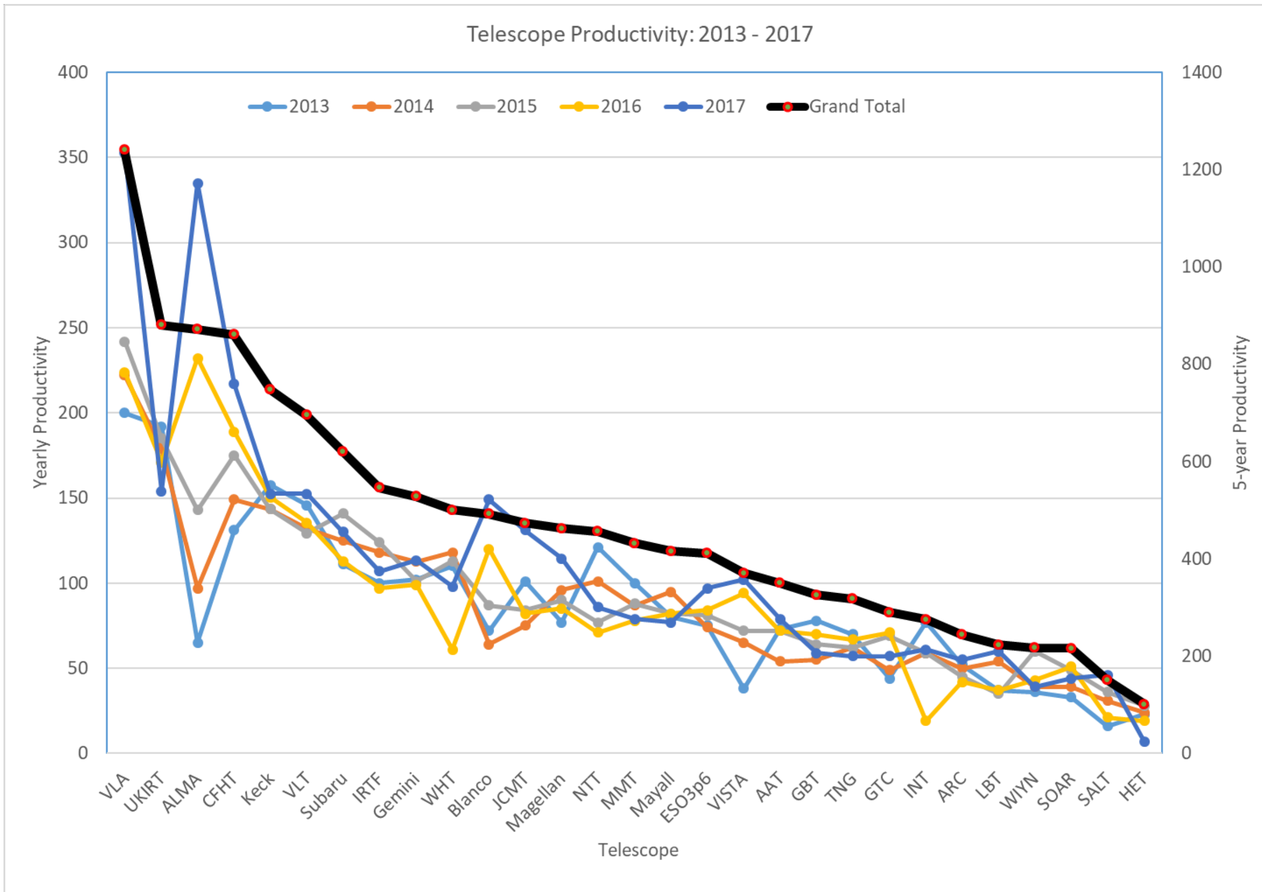

Figure 1 Telescope Productivity 2013-2017

Productivity is one useful metric, but it does not measure the impact. The impact of a paper is usually measured by the number of citations that it receives – more citations means higher impact (relevance). Citation numbers for a paper increase with the age of a paper. To properly combine papers over a period of years one needs to correct for this citation age effect.

In this work I have chosen to normalize each paper's citation count by the citation count for the median paper published in ApJ in the same year. The median ApJ paper acts as a standard measuring rod that increases with age. I refer this normalized citation count as the *impact* of a paper.

I use two measures of the aggregate impact of papers for each telescope; the average impact per paper (AIPP) and the median impact per paper (MIPP). (The distribution of impacts of individuals is very non-normal with a long tail towards high impact papers.)

Figure 2 shows *a Box and Whisker* plot of the impact of each facility's papers for the 5-year period. The MIPP is indicated by the horizontal line inside each box. The AIPP is indicated by the line in the plot. The ends of the box indicate the 1st and 3rd quartiles of the distribution. The difference between the lower quartile and upper quartile (the box) is called the inter-quartile range. The whiskers are defined as 1.5 times the inter-quartile range. The data points outside the whiskers are considered outliers and are not shown in this plot. The positive skew in each telescope's impact distribution is evidenced by the longer top tail compared to the bottom tail. In terms of MIPP, ALMA ranks 1st, CFHT 5th and Gemini 15th.

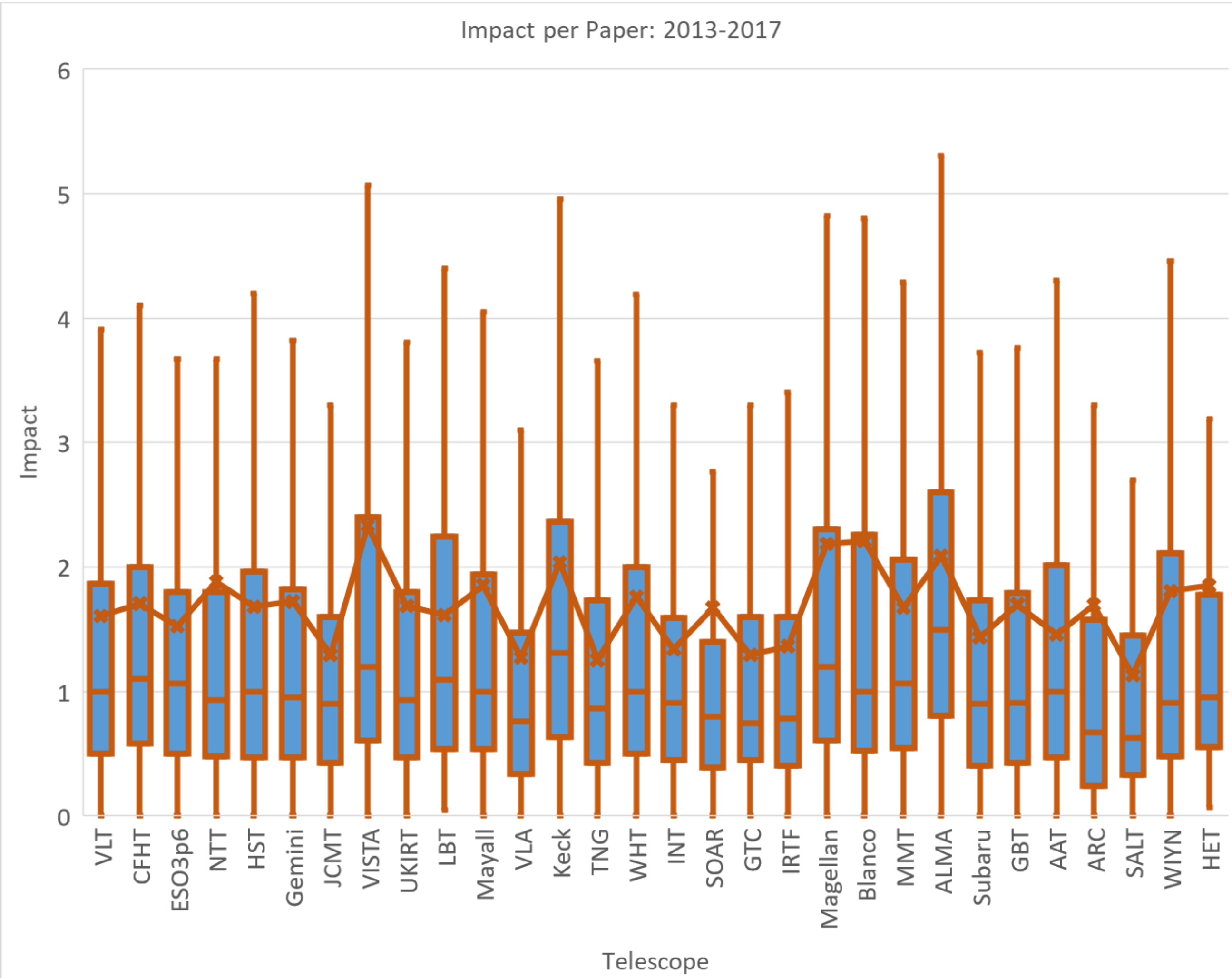

Figure 2 Box and Whisker Plot of Impact

Another measure which is useful is the number, or fraction, of high impact papers (HIPs). I define a HIP as one which is in the 90th percentile of the complete impact distribution. The percentage of HIPs are shown in Figure 3. ALMA has the highest fraction of HIPs which is no surprise given its recent completion and unique capabilities. CFHT and Gemini and firmly in the top plateau of the distribution.

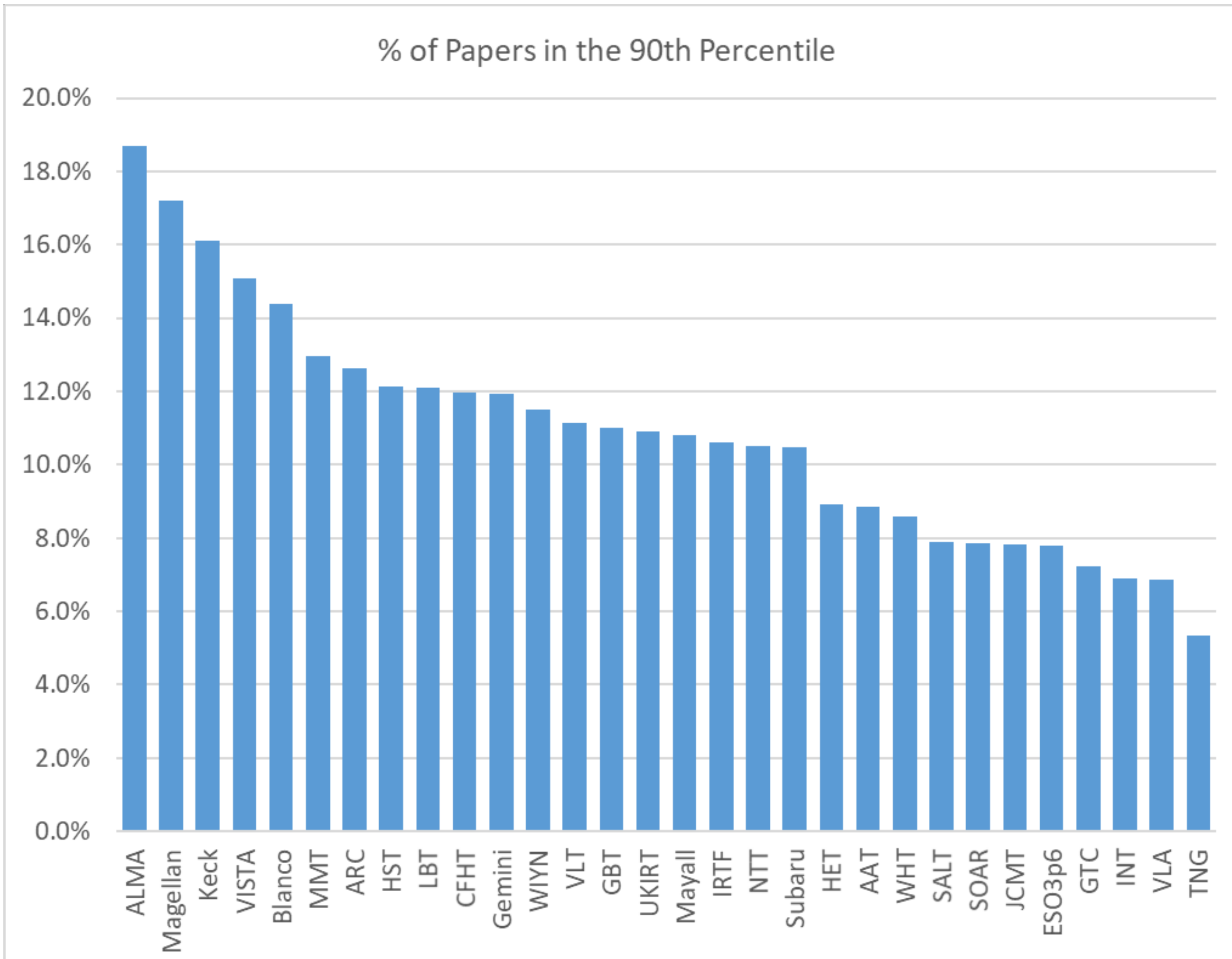

Figure 3 Percentage of high impact papers

Canadian Use of All Telescopes

While Canada invests in ALMA, CFHT and Gemini, Canadian researchers utilize the vast array of observing facilities available world-wide. It Is interesting to explore the bibliometrics of papers published by Canadians based on data from the portfolio of telescopes available to them.

Canadian papers, defined as having at least one author based in Canada, can be identified using NASA ADS's Bumblebee interface which allows one to search the affiliation field. Canadian papers are identified by matching "Canada" in the affiliation field. In addition, the positional search in Bumblebee can produce of a list of Canadian first author papers by matching "Canada" in the affiliation field of the first author.

The list of Canadian papers from each telescope can be identified by cross correlating the list of Canadian papers and the complete list of telescope papers using the unique *bibcode* identifier to match papers.

Figure 4 shows the percentage papers from each telescope that can be considered Canadian papers, both any and first author, for the period 2013 – 2017. As expected, there is generally a higher percentage of Canadian papers for those facilities in which Canada was a significant partner during the period (this includes JCMT). More surprising is the high percent of papers (both and first author) based on GBT papers. The Canadian community is making heavy use of this facility under NRAO's "open skies" policy. The fact that Canadian any author papers represent at least 10-15% of papers from all telescopes indicates the highly collaborative nature of Canadian astronomy. It is also interesting to note that almost 3% of ALMA papers have Canadian first authors, which is close to our nominal share of ALMA time. However, for CFHT and Gemini, where Canada has a 42.5% and 18% share, Canadian first author papers are 13% and 11% respectively.

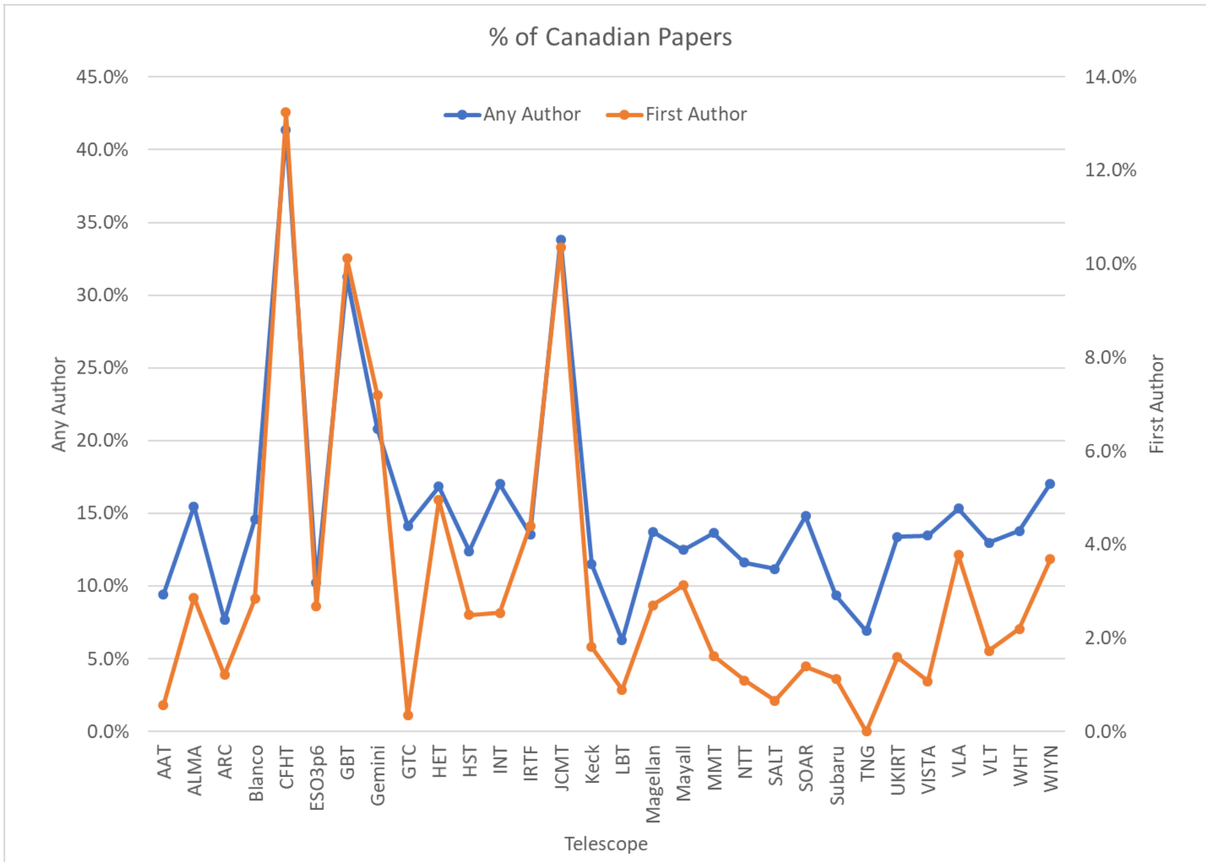

Figure 4 Percentage of Canadian Telescope Papers

It is instructive to look at the impact of Canadian papers compared the impact of all papers for each telescope. Figure 5 shows the AIPP of Canadian papers compared to the AIPP of all papers. Note that the AIPP of Canadian any author papers is higher than the overall AIPP for almost all telescopes. However, when one looks at Canadian first author papers the result is reversed, i.e., the AIPP for Canadian first author papers is lower than the overall AIPP for most telescopes. Across all papers in this sample, the AIPP is 1.7 while the Canadian any author AIPP is 2.15 and the Canadian first author AIPP is 1.39. **Papers with Canadian first authors have, on average, lower impact that for the complete sample of observatory papers.**

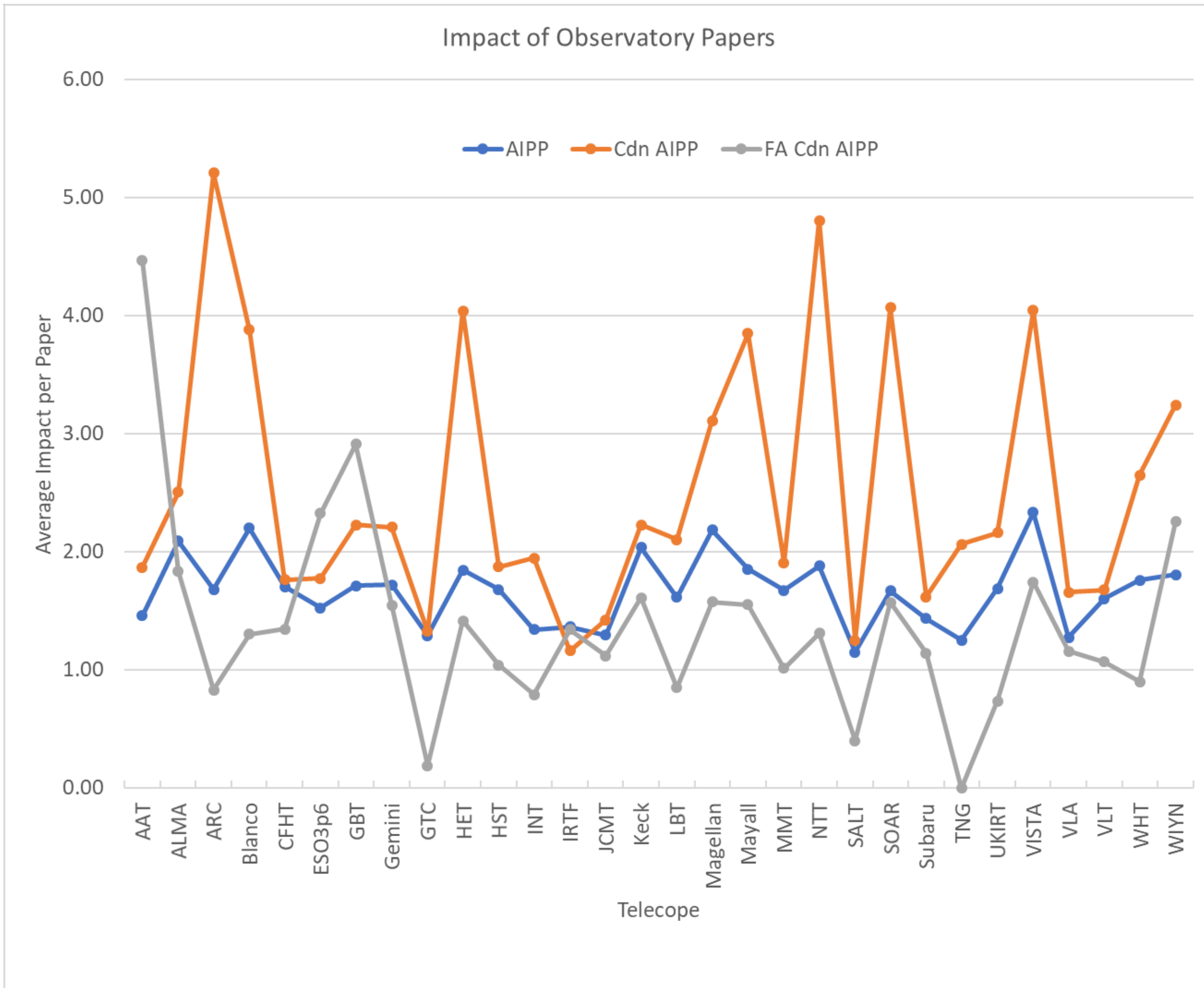

Figure 5 The AIPP for observatory papers separating our Canadian papers

One can investigate Canadian first author observatory papers more closely by identifying the status of the first author – faculty, PDF or student. For each 2013 – 2017 Canadian first author paper (467 papers) the status of the first author was determined. Figure 6 shows the number of first author papers, and the AIPP, for each of the three groups; faculty, PDF, student. Students are first author on 44% of these papers, twice the number of faculty first author papers. PDFs are first authors on 32% of the papers.

While students are first authors on the largest share of first author papers, these papers have the lowest mean number of citations. PDF first author papers have the highest number of citations, slightly more than faculty numbers.

Ideally one would want to compare the status distribution of first author papers for other countries, but this would require a large investment of time identifying the status of each person at the time the papers was published.

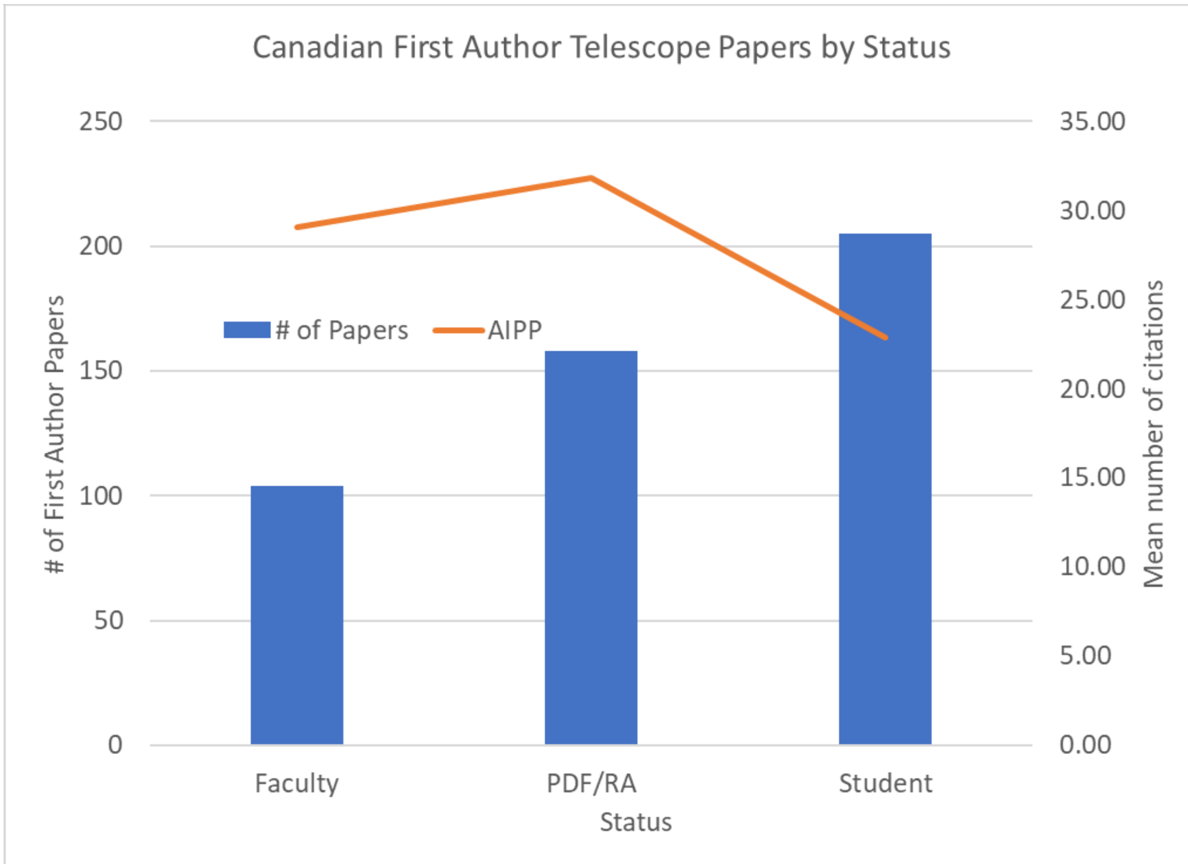

Figure 6 The number of papers and mean number of citations for Canadian first author papers based on observatory data published between 2013 – 2017.

It is also instructive to look at the number of authors on observatory papers. For the collection of 2013 – 2017 papers the mean # of authors is 12.15 and the median # of authors is 7. For Canadian any author papers the mean number of authors is 28.26 and the median is 12, well above that of the complete collection. For Canadian first author papers the mean is 11.55 and the median is 7, almost identical to the numbers for the complete sample of papers. It seems that Canadians do not lead large collaborations (assuming leading a collaboration results in being the first author), but are participants in large collaborations.

A box and whisker plot of the distribution of # of authors for Canadian any author and first author papers is show in Figure 7. The any author distribution is much more highly skewed towards a higher # of authors on each paper.

The AIPP (and MIPP) is correlated with the number of authors. This is not a result of more self-citations because of the higher number of authors. The AIPP of papers with 7 authors is 1.15 while the AIPP for papers with 12 authors is 1.6, an increase of 40%. If one takes the AIPP of all observatory papers ((1.7) and corrects by 40% difference in the median # of authors of the complete sample vs. the median for Canadian any author papers (7 authors per paper vs 12), one calculates an AIPP of 2.36, quite close to the AIPP of Canadian any author papers of 2.15. **One can account for the higher AIPP of Canadian any author observatory papers by the increase in the median # of authors on Canadian any author papers.**

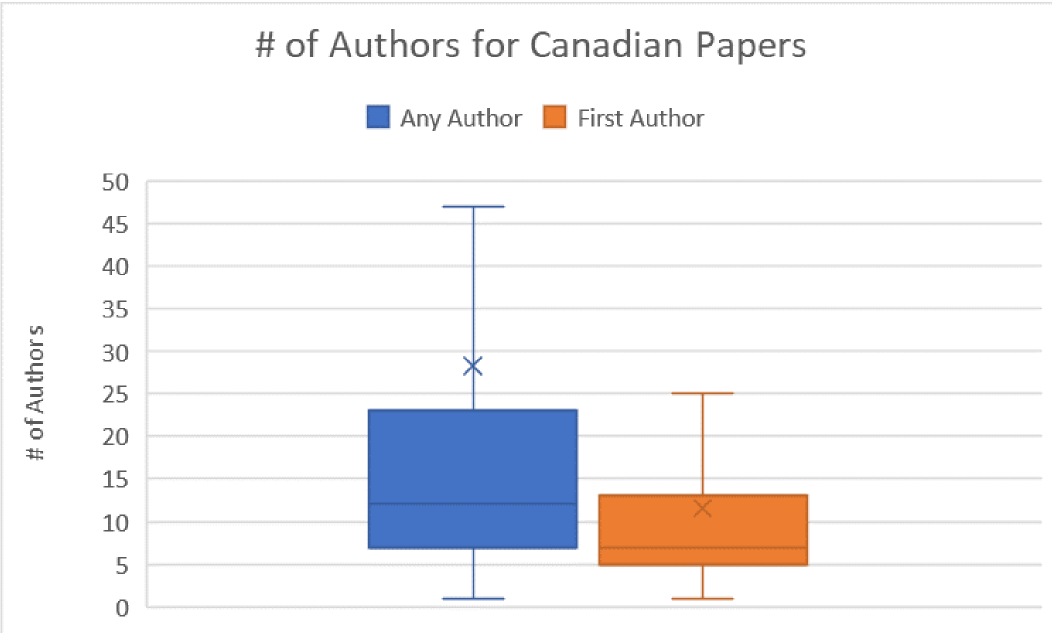

Figure 7 Box and whisker plot showing the distribution of the # of authors for Canadian observatory papers.

Canada's Overall Bibliometric Performance

Canada's ranking in astronomy and astrophysics research has long been recognized as being at or near the best in the world when measured as the mean citations per paper (Section 2.3 of LRP 2010). This metric is a measure of the relevance of the work published and is not influenced by the size of the research community.

Country ranking data from the SCImago Journal & Country Rank website is displayed in Figure 8. Their metric is average citations per paper. Their ranking data is based on data in Elsevier's SCOPUS database. As with all previous studies of Canada's international stature, this is based on counting any paper with at least one author based at a Canadian institution as a Canadian paper.

Figure 8 show that Canada ranked as one of the top 3 countries between 2006 and 2013. Their was a wider range in Canada's ranking prior to 2006. Possibly concerning is that for the 4 years following 2013, Canada has ranked as fifth or sixth.

Our field is fortunate is that we have a source of bibliometric information independent of Elsevier. Complete bibliometric information for astronomy & astrophysics papers is available through the NASA ADS (ADS) system. ADS's new system, Bumblebee, allows for more advanced searching techniques than in Classic ADS. In Bumblebee is it possible to search for papers by the affiliation of the authors. This makes it possible to identify papers with at least one author at a Canadian institute. (There will be a very small number of false positives due to places such as La Canada in Spain). This search, which can be limited to a specific year or range of year, can be restricted to refereed papers. The search will also indicate the total number of citations to these papers and one can easily calculate the mean number of citations per paper. This can be repeated for other countries allowing us to rank countries on their average citations per paper. Figure 9 shows the mean citations/paper for the period 2014 – 2017 for several countries based on any author affiliations. In this plot Canada ranks second behind the Netherlands.

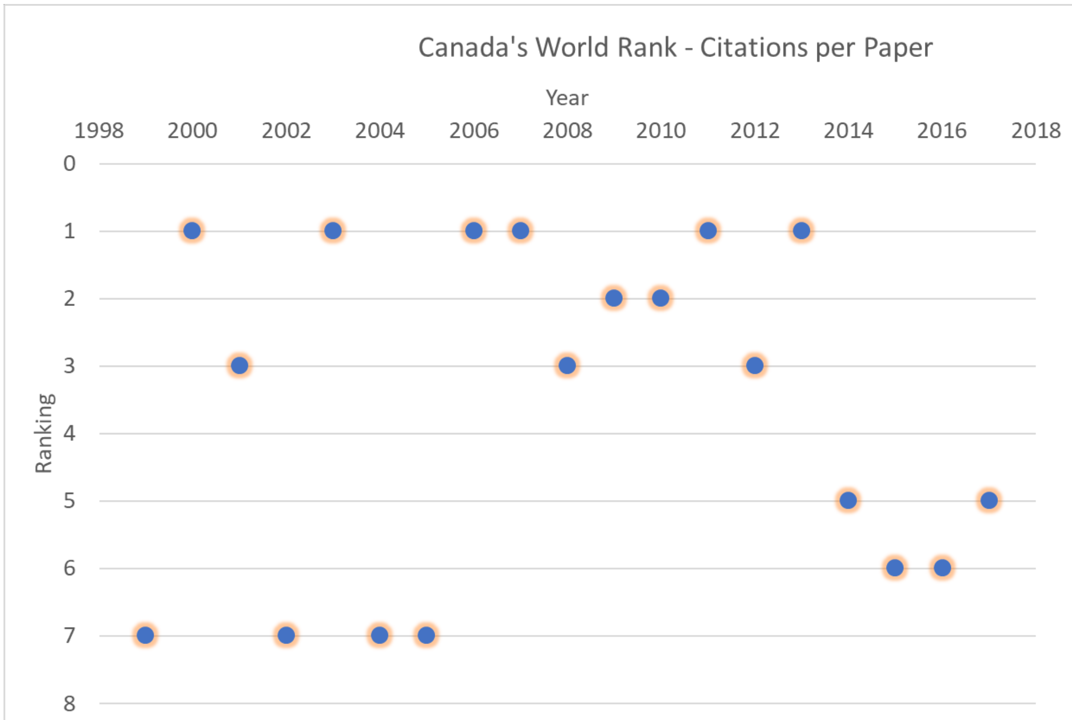

Figure 8 Canada's Ranking Astronomy & Astrophysics by Year (from SCImago Journal & Country Rank)

This is different from the SCImago results. This may come down to the which journals are considered to publish astronomy papers, what is considered an astronomy paper, and the completeness of citation information.

The NRC has been tracking publication information for its Research Centres using data from SCOPUS (the same source as SCImago) and I've been sent the data for Herzberg. There are differences in the SCOPUS papers associated with Herzberg and those I've identified as being Herzberg papers in the ADS. This leads me to conclude that there is likely to be differences between results based on SCOPUS data and data from ADS.

One important feature of ADS Bumblebee is that one can identify Canadian papers based on the affiliation of the first author. Unlike any previous analyses of country rankings, which have been based on any author identifications, a country ranking based on first author affiliation can be produced.

Figure 10 shows the mean citations/paper for the period 2014 – 2017 for several countries based on first author affiliations. In this plot Canada ranks eighth, significantly different from a ranking based on any author affiliation

This change of country ranking between any and first author affiliation is consistent with the results obtained when looking at the impact of observatory papers. Canada performs much better when considering any author Canadian papers compared to first author Canadian papers.

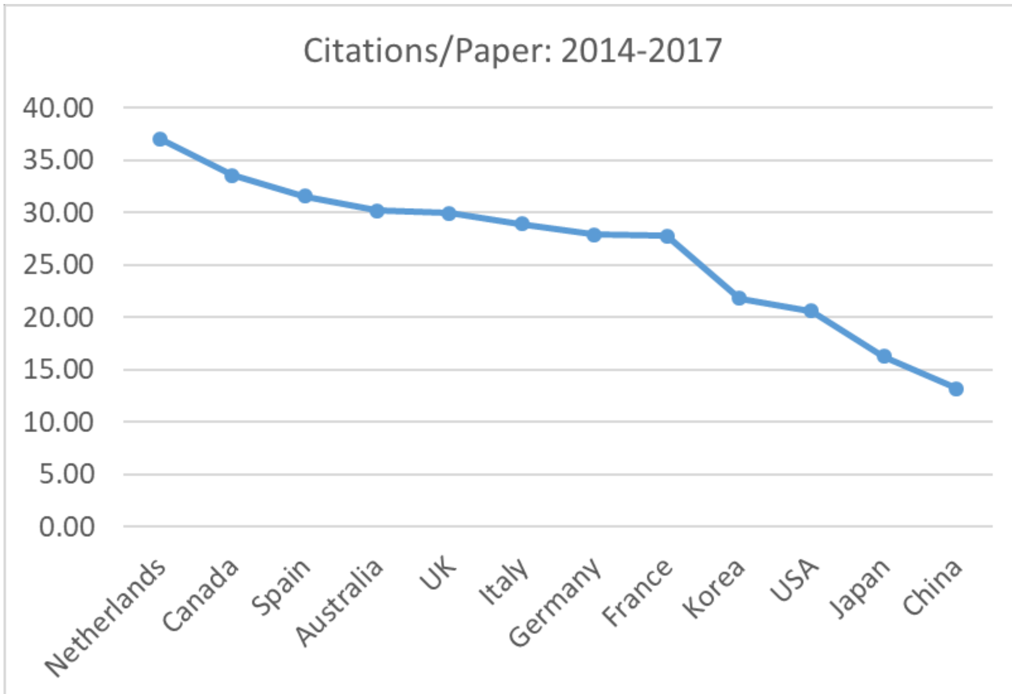

Figure 9 Citations per paper for selected countries based on ADS Bumblebee results based on any author country affiliation

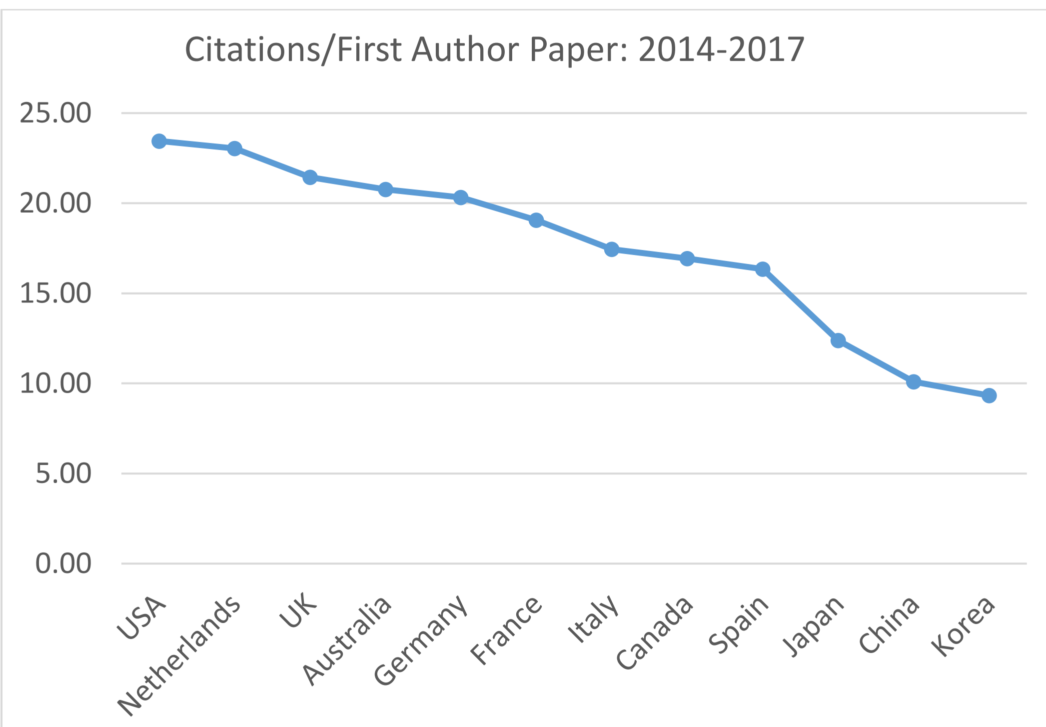

Figure 10 Citations per paper for selected countries based on ADS Bumblebee results based on first author country affiliation

The trend in the number of first author papers by Canada, compared to other countries, is also interesting. Figure 11 shows the number of first author papers for the same set of countries between 2010 and 2018. Each country's paper counts have been normalized to the average of their 2010 and 2011 numbers.

Canada's astronomical community is very similar to Australia's. While Australia's first author paper production has increased by over 30% from 2010, Canada's production has declined. France is the only other country whose first author paper production has declined. The increase in the number of first authors for the UK, China and Japan is quite noticeable.

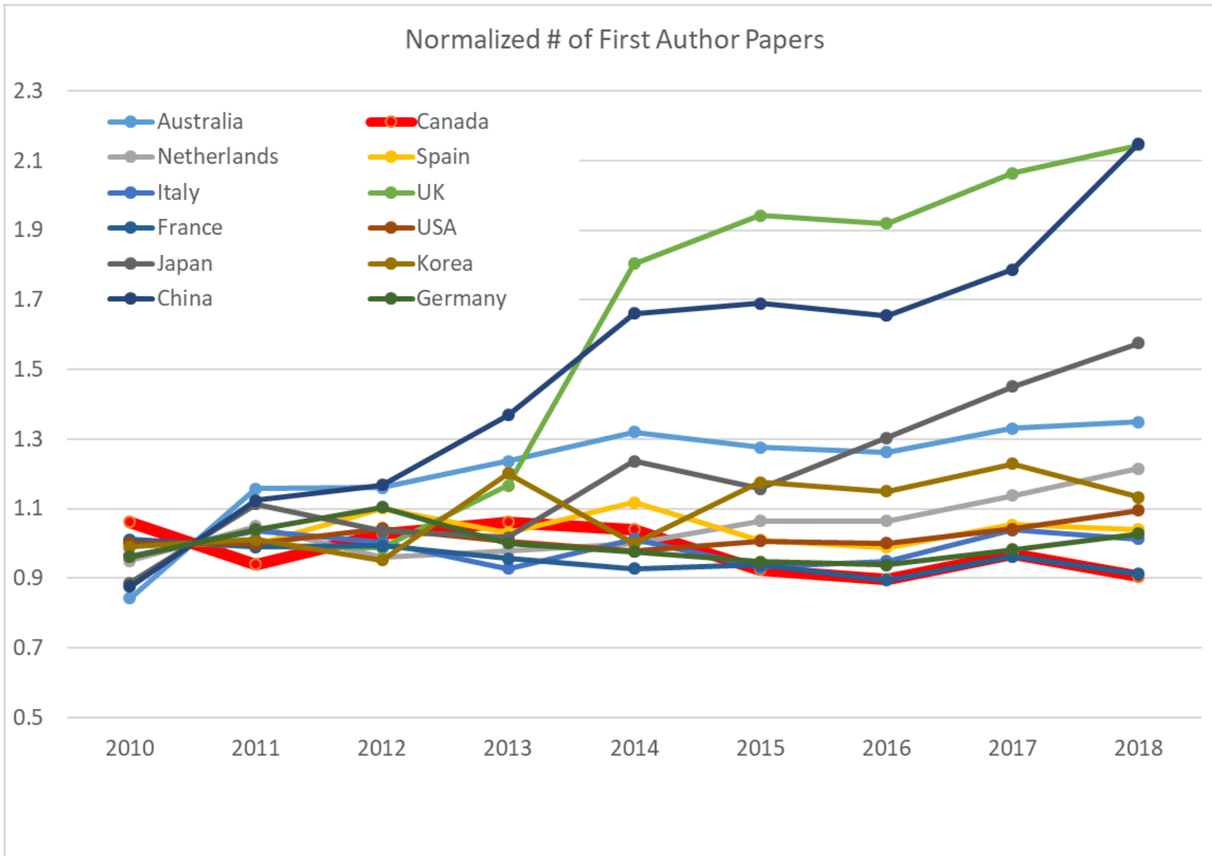

Figure 11 Normalized number of first author papers by affiliation country

In trying to understand the cause of the trend of Canadian first author papers in Figure 11, one needs to understand who writes first author papers. Looking at the publication history of Canadian faculty members provides some insight.

I maintain a database of Canadian faculty that includes a variety of information including the year in which they received their PhD. ADS's API provides access to detailed information on the publication history of an author, i.e., the number of papers published per calendar year. These numbers can be shifted, for each person, to the number of papers published by 'years since PhD'. This can be done for any author and first author papers.

Figure 12 shows the mean number of first papers published each year, relative to the year of PhD, for Canadian faculty. Also shown is the cumulative number of first author papers. The peak period for first author paper production is when a person is a PDF and the early years of holding a faculty position. In fact, the median number of first author papers drops to zero early in a person's career as a faculty member.

The number of first author Canadian papers is heavily influenced by the number of PDFs in the community and one explanation for Canada's decreasing number of first author papers compared to other countries is the lower number of PDFs in Canada.

The number of PDFs in Canada can be approximately obtained by exploring each astronomy department's website. This reveals that Canada has approximately 0.55 PDF positions for each faculty position. The UK published a report on the demographics of their astronomy community in 2016. Using the numbers in that report one finds that at that time the UK had approximately 1.04 PDF positions for each faculty position, almost double the rate in Canada.

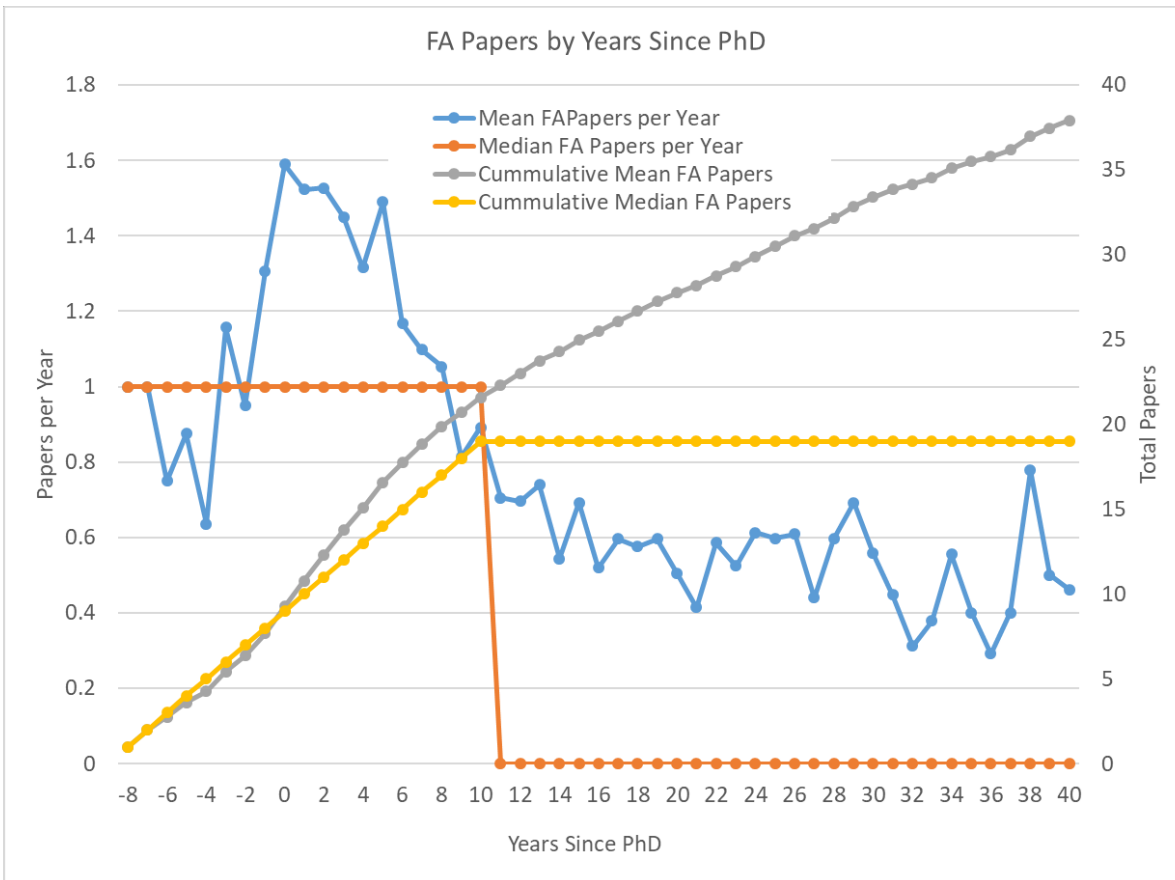

Figure 12 Publication history of Canadian faculty members

Summary

- The observing facilities that Canada supports financially are amongst the highest performing, both in productivity and impact, of facilities world-wide.
- For most observing facilities, Canadian any author papers have higher impact per paper than that of the full bibliography of the facility.

- However, Canadian first author papers on most facilities have a lower impact per paper than that of the full bibliography of a facility
- Almost 80% of Canadian first author papers based on telescope data are authored by either students or PDFs.
- Using ADS numbers, for the period 2014 – 2017 Canada ranks second in the world in mean impact per paper when considering any author papers.
- However, when considering only Canadian first author papers, Canada ranks only 8th in the same metric.
- The number of Canadian first author papers has declined slightly since 2010 (based on ADS numbers), while the number has increased for most of our ‘competitors’.
- Based upon the publication history of Canadian faculty members, a person’s first author paper production peaks during their postdoc years and then declines once they have a faculty position.