New Metrics for Observatory Publication Statistics

Arnold H. Rots, Sherry L. Winkelman, Glenn Becker

CXC/SAO Harvard-Smithsonian Center for Astrophysics

Abstract

We are proposing several new publication metrics that are more meaningful and less sensitive to observatory-specific characteristics than the traditional ones. They fall into three main categories:

- **Fraction of observing time published**: percentage of observing time presented in refereed journal articles, taken after approximately three times the median time.
- **Speed of publication**: median time it takes for observations to get published; time it takes to reach the stable percentage above.
- **Archival usage**: median time after observation for publications subsequent to the first; percentage of observing time presented more than twice in refereed journal articles; percentage of available exposure time published in a given year.

Citation of results is a fourth category, but it lends itself less well to definite statements.
Applied to the bibliography of the Chandra X-ray Observatory, the median time from observation to publication is 2.36 years; after about seven years 90% of the observing time is published; the total annual publication output of the mission is 60-70% of the cumulative observing time available, assuming a two year lag between data retrieval and publication; and after seven years approximately 60% of available exposure time is published more than twice.

This work has been supported by NASA under contract NAS 8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center.

Background

There has always been a desire to be able to measure the success or scientific value of our observatories – not in the least on the part of funding agencies – and this has taken on more urgency in times of increased budgetary pressures. Metrics based on publications (as is not uncommon for judging the scientific efficacy of individual researchers) are obvious candidates, but the problem is to design them in such a way that they are not sensitive to specific observatory characteristics and allow some degree of comparison. There have been various attempts based on numbers of publications, numbers of citations, as well as more complicated algorithms. See, e.g., Trimble & Ceja (2010), Madrid & Macchetto (2009), and Crabtree (2008). As an example, Crabtree (2008) uses an Impact Distribution Function which he extended in a recent private communication to a purportedly objective "Performance Factor": the ratio of high-citation-rate papers to low-citation-rate papers. This is problematic: does an observatory with 10 papers in both categories perform significantly better than one with 100 in the former and 500 in the latter category? Moreover, one should also keep in mind that publication and citation practices are not uniform across the astronomical community, but vary significantly between our sub-communities.

The Chandra X-ray Observatory has one of the most complete and comprehensive bibliographic databases (see also Winkelman & Rots 2010). We used the records associated with refereed journal articles that present identified Chandra observations to experiment with various metrics and we propose specific metrics in three areas that we feel are fairly insensitive to individual observatory or mission characteristics – or those of their user communities.
The bibliography's use is currently being extended by its integration into semantic search tools (see Winkelman & Rots 2012).

Proposed Bibliographic Metrics

In the following metrics we use refereed journal articles that can be identified as presenting specific observations. We assume that we do know the exact exposure time of each observation; for total (annual) exposure time we use the net exposure time, excluding calibrations, lost time, slew time, etc.

Fraction of observing time published

This is the percentage of observing time that is presented in refereed journal articles. In our (admittedly limited) experience, this percentage is stable when taken after approximately three times the median time to publication (see next paragraph).

Speed of publication

There are two metrics that provide information on the speed of publication: the median time it takes for observations to be published; and the time it takes to reach the stable percentage in the previous paragraph. We estimate the latter to be approximately three times the former.

Archival usage

There are at least three potential metrics for archival usage of observational data: the median time after observation for publications subsequent to the first one; the percentage of observing time presented more than twice in refereed journal articles; and the percentage of available exposure time published in a given year.
We need to emphasize two points here. First, it is extremely difficult to classify papers as containing purely archival research. If the data are published during a proprietary period, with the proposal's PI as first author, it is clearly not an archival paper. But if the first publication is dated two years later, with the PI as the 26th author, it becomes rather murky. Moreover, a significant number of papers may combine the publication of new observations with a re-analysis of older data. Consequently, any statistics on archival research tend to be unreliable. It is far easier, and more meaningful, to flag papers as "having archival content". Second, if one sets out to determine the percentage of available exposure time that is published, one needs to take into account the lag between extracting the data from the archive and the actual publication of the paper. In addition, this lag tends to be shorter for a new observatory, instrument, or mission than for a well-established one. In other words, the lag gets longer as a mission ages.

Caveat

We offer these metrics as ones that have greater potential for representing an objective picture of an observatory's performance than metrics that have been in use in the past. However, we still need to caution strongly against a blind use of these numbers that ignores their context. Different observatories will produce different results for these metrics. The nature of the observatory, its data, and/or its user community may explain such differences and indicate that they are not necessarily indicative of a difference in performance. Taking these factors into account while considering the numbers will enable one to make a sensible judgment on the value of an observatory's scientific impact, but one should avoid striving for a measure of dollars/euros per canonical science quantum – that would be foolishness.

Chandra X-ray Observatory Statistics

As an example, we present an application of the metrics proposed on the left to the Chandra X-ray Observatory. It is based on the contents of its bibliographic database on 10 August 2011.

Fraction of observing time published

- The percentage of exposure time that is eventually published is approximately 90% (Fig. 1)

Speed of publication

- The median time between observation and first publication is 2.36 years (Fig. 2)
- The stable percentage of 90% is reached after approximately 7 years, or three times the median (Fig. 1)

Archival usage

- The median time for (all) publications subsequent to the first one is 5.77 years; this figure is undoubtedly related to the age of the mission: as it happens it is currently roughly half the mission's age
- We estimate that approximately 60% of all exposure time is eventually published more than twice (Fig. 1)
- During any given year, approximately 40% of the exposure time available in the archive is extracted to be published one to two years later; the total annual publication output of the mission in publications, using that metric but including multiple use of the same data, is 60-70%, amounting to six times the annual exposure time in 2010 (Fig. 3)

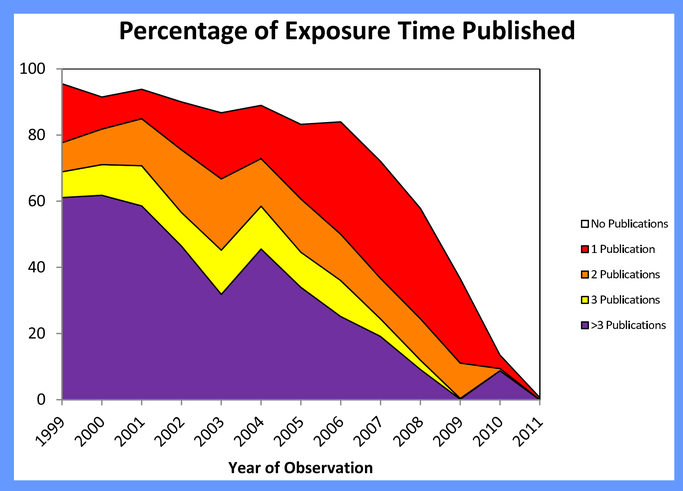

Fig. 1 Percentage of exposure time for each year of operation of the observatory that was published once, twice, three times, more than three times, or that remained unpublished.

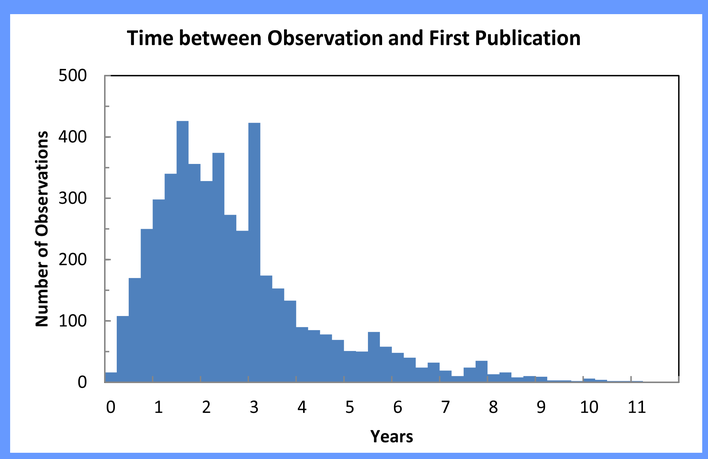

Fig. 2 Distribution of lag between observation and first publication.

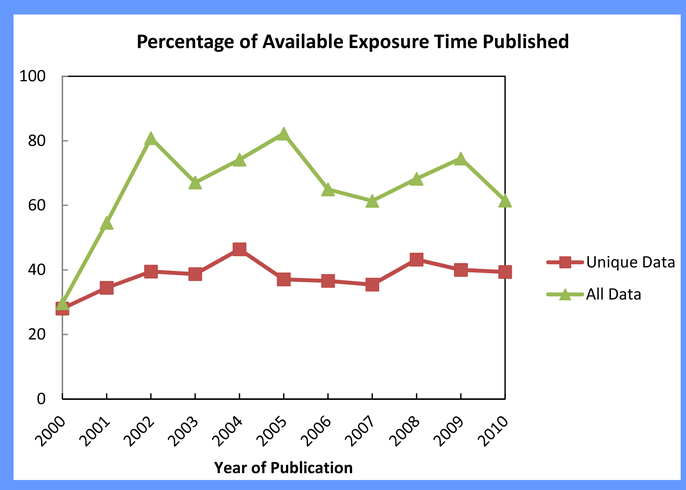

Fig. 3 Percentage of available exposure time (total and unique) published in a given year.

References

Crabtree, D. R. 2008, Proc. SPIE 7016, 70161A
Madrid, J. P. & Macchetto, D. 2009, arXiv:0901.4552v1
Rots, A. H., Winkelman, S. L., & Becker, G. 2011, arXiv:1110.1212
Trimble, V. & Ceja, J. A. 2010 Astron. Nachr. 331, 338
Winkelman, S. L. & Rots, A. H. 2010, ADASS XIX
Winkelman, S. L. & Rots, A. H. 2012, **Poster P163**

arots@head.cfa.harvard.edu
http://cxc.cfa.harvard.edu/cgi-gen/cda/bibliography
http://arxiv.org/abs/1110.1212