

A Bibliometric Study to Manage a Journal Collection in an Astronomical Library: Some Results

Monique Gomez

*Instituto de Astrofísica de Canarias, Library, C/Vía Lactea s/n,
E-38205 La Laguna, Tenerife, Spain*
mog@ll.iac.es

Abstract. The evaluation of the use of library collections is a fundamental tool for the development of a relevant and cost-effective collection. Bibliometrics offers several methods to measure the level of use of collections. In this paper, we present the first results of a bibliometric analysis of the references cited in the articles published by authors of the Instituto de Astrofísica de Canarias during the last decade. We have determined which kinds of publications and which journals are mainly used. We also have studied the ages of the cited journals and their cost-effectiveness. Additionally, to see if the advent of the WWW in the mid 1990s has influenced the reading or citing practices of the authors, we have compared two sets of data: articles published over the period 1991–3 that we define as the pre-WWW period and articles published over the period 1997–9 defined as the WWW period.

1. Introduction

The management of journals in libraries is a difficult task, as subscription costs increase year after year, library budgets are tight, new journals appear and new modalities of subscriptions are becoming available. Small and medium size astronomical libraries that do not belong to big consortia may have difficulty in maintaining a relevant collection of journals. In this context, the evaluation of collections is becoming even more necessary.

Many quantitative as well as qualitative parameters must be taken in account to perform an appropriate evaluation of a library collection. Nevertheless, one of the most important parameters is the level of use of collections. As direct measurements are not always possible, a librarian can use a bibliometric method such as the analysis of references cited in articles published by the patrons, in order to obtain an estimate of the use of the collection.

In recent years at the library of the Instituto de Astrofísica de Canarias (IAC), we have tried to improve our acquisitions policy and especially to better manage our journal collections. We have conducted several surveys among our patrons to determine the level of use of a specific journal collection. Although these surveys have always allowed us to drop some subscriptions, the results were usually insufficient to clearly define a subscription policy because patrons consider everything important even if they do not use it. So, in order to obtain

more objective data about the use of the collections, we decided to carry out an analysis of the references cited in the articles published by IAC authors during the decade 1990-1999. In this article, we present the first results of this analysis.

2. Methodology

From a list of articles published by IAC authors in research journals during the period 1990–9, we selected one set that included the articles published from 1991 to 1993 (the pre-WWW period) and a second that included the articles published from 1997 to 1999 (the WWW period). For each article we had the following abbreviated reference: title of publication, volume number and page number.

2.1. Extraction of the Cited References

To collect the references cited in IAC authors' articles, we had to search a bibliographic database that provides the cited references such as the NASA–ADS database or the Science Citation Index (SCI) database produced by ISI. Although the NASA–ADS database is more accurate as cited references are checked and facilitates further processing with the bibcode format, we finally used the SCI database because it is supposed to be more exhaustive. We used the CD-ROM version of the SCI available at the library of the CINDOC-CSIC in Madrid.

Each IAC author's article was searched in the corresponding file of SCI. Each retrieved article was collected and saved in a user-defined profile in order to obtain (in a comma-delimited format) the following information: "IAC author article", "cited ref 1/ cited ref. 2/cited ref. 3/...".

2.2. Processing of the Collected References

The data obtained with the SCI database were arranged to be included in an Excel spreadsheet with the following structure:

Publication year of IAC Art./Ref. IAC Art./Cited Author/Cited Year/Cited Publication/Cited Volume/Cited Pages.

As the cited publications were not always abbreviated in the same way, each cited publication had to be identified and written in a standard form. This was relatively easy for journals, but more difficult for books, as they could be cited by abbreviated title or by abbreviated series title, etc. Each cited publication was then classified by category, i.e. Book, Journal, Data, Not published and Unidentified. The "Book" category was subdivided in the subcategories "Proceedings", "Monographs" and "Theses". The "Data" category included the subcategories "Catalogues" and "Circulars". The "Not published" category was subdivided in the subcategories "Unpublished", "In press", "Communication", "Pre-print" and "E-print". Once arranged and classified uniformly, quantitative data could be easily obtained.

Table 1.

	1991-3	1997-9	Total
Art. searched	205	459	664
Art. retrieved	189	395	584
Cited ref. in Art.	6,351	12,937	19,288
Art. with no cited ref.	4	7	11
Cited refs./Art.	34	33	33

3. Analysis Results

As summarised in Table 1, for the period 1991-3, 205 articles were searched, 189 (92.2%) were retrieved and 6,351 cited references were obtained. For the period 1997-9, 459 articles were searched, 584 (86.1%) were retrieved and 12,937 cited references were obtained. The average of cited references per article is 33 and 32 respectively.

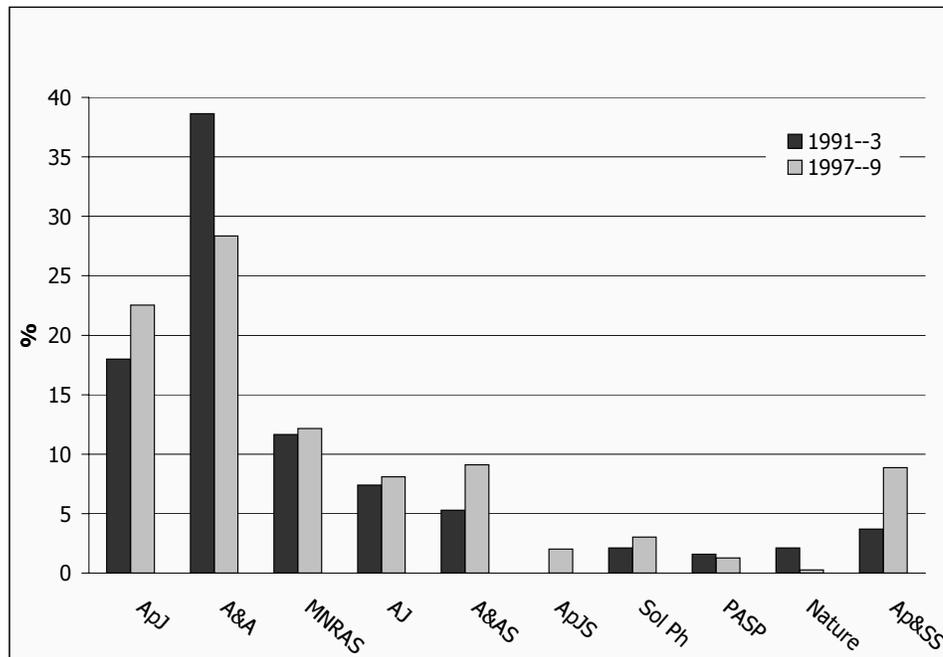


Figure 1. Distribution of IAC authors' articles by journal. Only journals with more than 1% of the articles are represented. Journal titles are abbreviated according to the NASA-ADS list.

3.1. The Cited Articles

IAC authors' articles were published in 27 different journals. More than 70% were published in the three core journals in astronomy: *Astrophysical Journal*

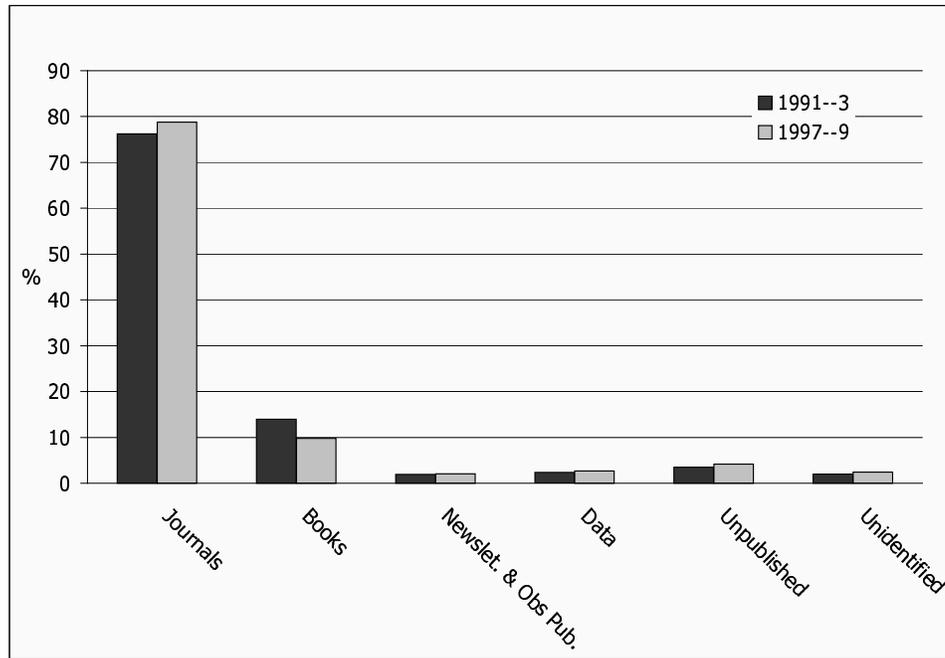


Figure 2. Distribution of cited articles by type of publication

(*ApJ*), *Astronomy & Astrophysics (A&A)* and *Monthly Notices of the Royal Astronomical Society (MNRAS)*.

The distribution of IAC authors' articles by journals is illustrated in Figure 1. We observe that IAC authors publish mainly in *A&A*, as is usually the case among European astronomers. In the WWW period, the apparent decrease in *A&A* is related to the increase in *A&AS* because at the end of the studied decade, two sections of *A&A* were moved to *A&AS* in order to complete the volumes of *A&AS* that were due before this journal ceased publication by the year 2000. So some articles that should have been published in *A&A* were finally published in *A&AS*. The increase observed for *ApSS (Astrophysics and Space Science)* is because of the publication of the proceedings of a conference hosted by IAC and to which many IAC authors contributed.

3.2. The Cited Publications

In this part of the work, we wanted to know which kinds of publications IAC authors cite in their articles, which journals they mainly cite, how far back they cite them and if the advent of the WWW had changed their reading or citing habits.

Which Publications Are Cited? As shown in Figure 2, IAC authors cite mainly journals (78%) and then books (11.2%), newsletters & observatory publications (2%), data (2.6%), unpublished material (4%) and unidentified sources (2%). In 1997–9, a small increase in journals, and a small decrease in books is observed.

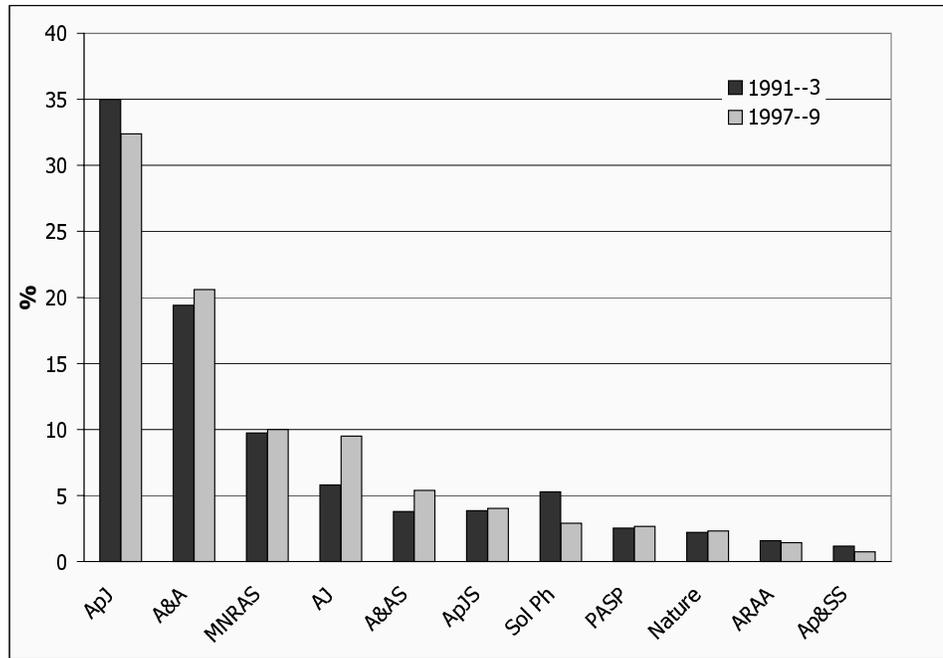


Figure 3. The top cited journals

These differences may be caused by the fact that journals are available online, whereas generally books are not. Nevertheless, as the observed differences are small, we should study the data over a longer period to confirm this tendency.

For both periods, the books cited are mainly proceedings of conferences (64%), then monographs (24%) and theses (12%). In the Data category, 51% are catalogues and 49% are circulars. In the Not published category, 53% are in press, 33% are unpublished, 8% are private communication and 6% are preprints or e-prints. In 1991-3, 4% of the Not published material were preprints, and there were no e-prints. In 1997-9, preprints dropped to 2%, and e-prints appeared and represented 6%.

Which Journals Are Cited? Among the cited journals, 95% are astronomical journals and 5% are a variety of journals in physics, general science, geophysics and other related sciences. More than 90% of the cited references were published in only 11 astronomical journals, and the rest was published in 142 different journals for the period 1991-3 and 174 different journals for the period 1997-9.

As shown in Figure 3, where the top cited journals are represented, IAC authors cite mainly *ApJ* then *A&A* and *MNRAS*. These three journals account for 63% of the total citations. Between the pre-WWW and the WWW-period we observe important differences in the patterns of citation for two journals: *AJ* increases from 5.8% of the cited journals in 1991-3 to 9.5% in 1997-9 and *Solar Physics* decreases from 5.3% to 2.9%.

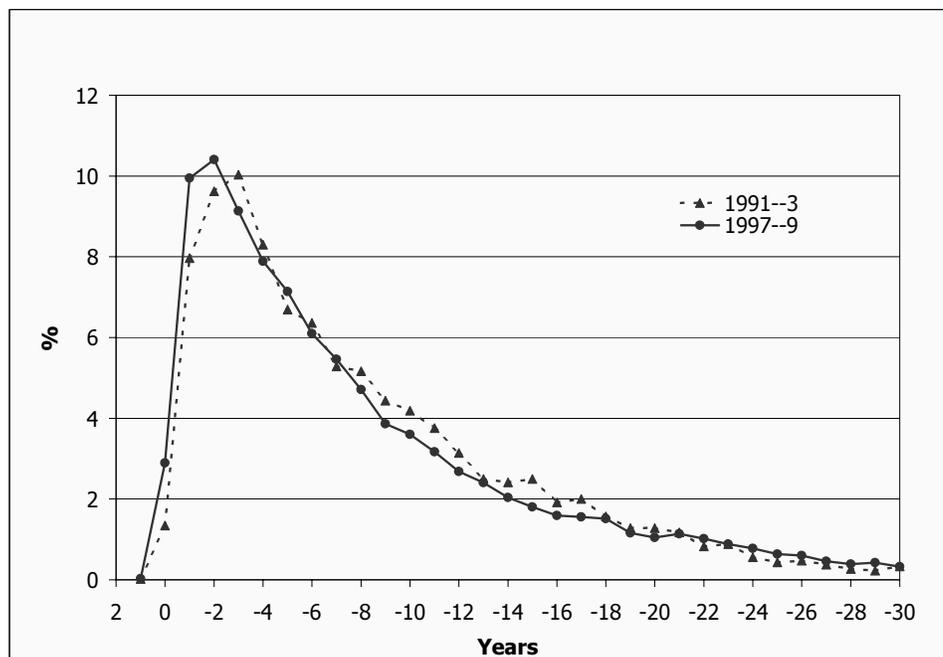


Figure 4. Distribution of cited journals by age

Although there may be various reasons to explain the increase in *AJ*, we think that in our library, *AJ* has become more visible since the advent of the WWW because of the NASA-ADS databases (Abstract and Article Services) and the Online edition of *AJ*. In fact, in the early 1990s, new issues of *AJ* were received with long delays. Patrons often did not read or check it on a regular basis, as they did with the three core journals. At present, with the NASA-ADS Abstract Service, patrons retrieve easily bibliographic references of *AJ*, and they can access the full text of the online edition through the publisher or the scanned articles through the ADS. In the case of *Solar Physics* the observed decrease can be explained at least partly by the fact that abstracts and scanned articles have not been available through ADS until 1998 and 2001 respectively, and our access to the online edition started only in year 2000. So this journal has become more visible only recently. Although based on our observations of the habits of our patrons, to confirm these explanations, we should conduct a survey.

How Old are the Cited Journals? IAC authors have cited journals that were published between 1858 and the year of publication of their own articles. However, they mainly cite journals that are less than 10 years old (70% of the cited years).

In Figure 4, we have represented the “age” of the cited journals, where the age is defined as the difference between the year of publication of the IAC article and the year of publication of the cited article. It is usually negative because authors normally cite articles that have been previously published. Although the

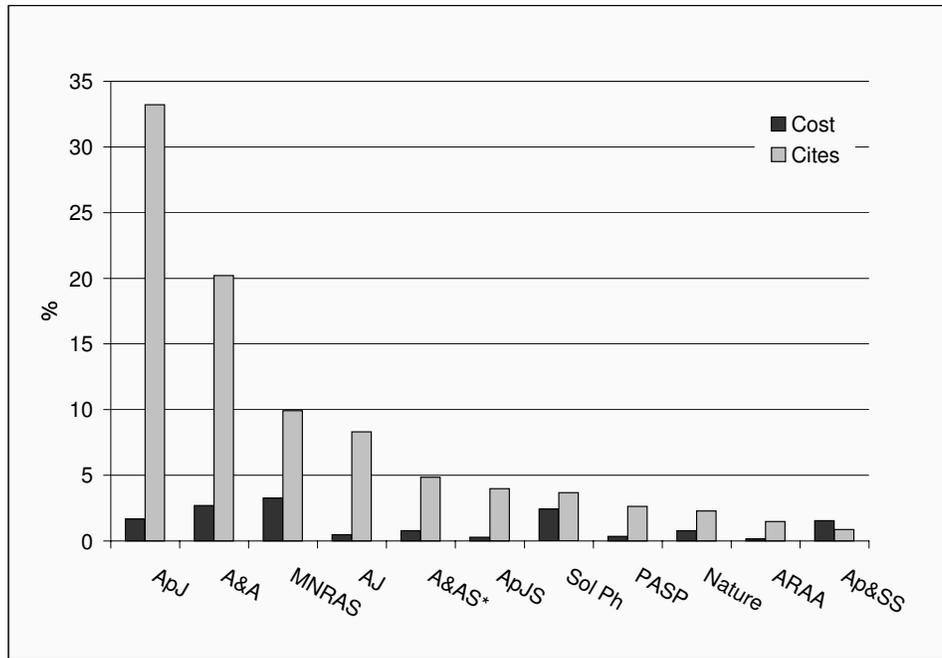


Figure 5. Cost-effectiveness of the top cited journals. The cost of *A&AS* is the cost in year 2000.

pattern is similar for both periods, we observe that more recent publications are cited in the WWW period. In fact, the maximum is at -2 years for the WWW period and at -3 for the pre-WWW period. We think that this difference reflects the fact that journals are available faster now with the online edition than in the past when we had to wait for the paper edition. But by our understanding, it does not reflect a change in the citing habits of the authors. They still refer to older articles when necessary and they read the literature as soon as it is available.

4. Cost-effectiveness of the journals

In this section, we have studied the cost-effectiveness of the top cited journals, on the one hand, and of our physics journals collection on the other, as this latter makes heavy inroads into our annual subscription budget.

The costs, expressed as a percentage of the annual subscription budget and the percentage of citations received by each top cited journal, are represented in Figure 5. If we exclude *A&AS* (which ceased publication in the year 2000), these journals cost 13.5% of the annual subscription budget and account for 87% of the citations received by journals. So we can say that they are very cost-effective, *ApJ* especially so.

In Figure 6, we have represented the same information as in Figure 5 but for the physics journals (excluding optics) cited. These journals represent 24%

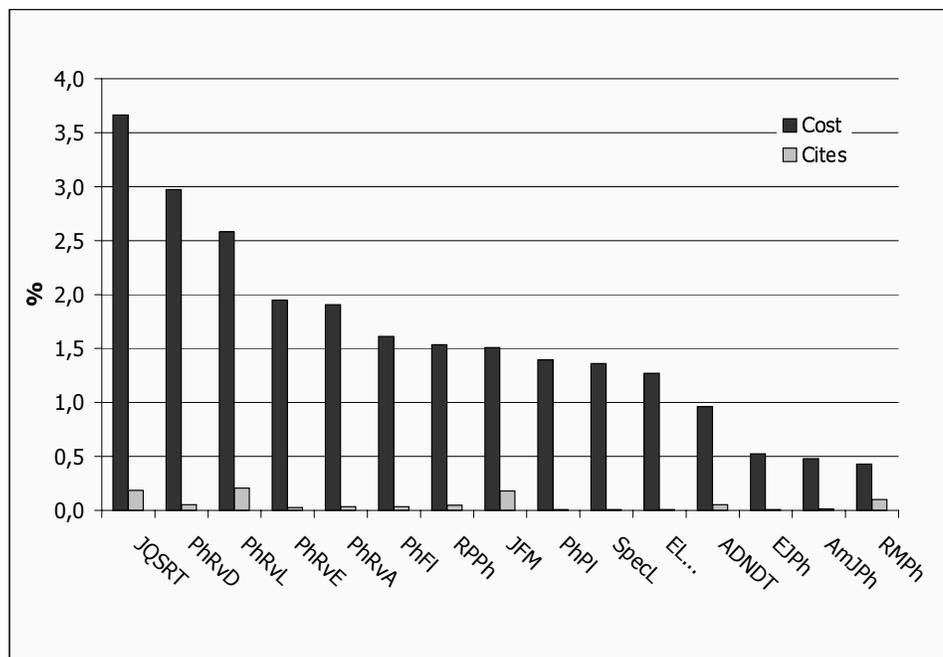


Figure 6. Cost-effectiveness of the physics journals

of our annual subscription budget and account for less than 1% of the citations over the whole period. The cost-effectiveness of these journals is really very low. With this study we have also detected that among our physics collection four journals have never been cited, although they cost us 8.3% of the annual subscription budget.

Since 1985, the IAC library has spent a lot of money to maintain a good collection of journals, not only in astronomy, but also in physics and other related sciences. We subscribed to many physics journals, because at the time the physics section of the University of La Laguna library did not exist, inter-library loan was very slow and access to bibliographic databases was difficult and expensive. So we developed collections just in case our astronomers needed them, although we knew that the physics journals would never be heavily used. Now, with the automation of libraries and the advent of WWW, access to bibliographic databases or to the table of contents of journals is easier and even free in some cases, and interlibrary loan is faster. In addition, the University of La Laguna has now a physics library, and in Spain the research council (CSIC) has several physics libraries with which we maintain good relations. So we are probably ready to change our subscription policy especially for the journals that are not much used.

5. Summary

With our analysis of the references cited in IAC authors' articles, we have obtained several interesting quantitative parameters about the use and the cost-effectiveness of our library collections. The results have confirmed in a more objective way what we already had detected through observation of the patrons' habits and the surveys conducted previously.

We have seen that our patrons publish mainly in the core astronomical journals, with a preference for *A&A*, and that above all they cite astronomical journals, *ApJ* being the most frequently cited. Most of the journals cited are less than 10 years old. The cost-effectiveness of astronomical journals is high whereas that of the physics journals is very low.

Although there are no dramatic differences between the pre-WWW and the WWW period, it seems that fewer books and more journals are being cited in the WWW period. We have seen that in our library at least one journal (*AJ*) has become more visible because of the advent of WWW. Also, we have detected that the more recent literature seems to be cited more quickly because it becomes available earlier through the WWW.

The next step, if we want to improve our subscription policy, will be to complete the evaluation of the collection taking in account not only the quantitative data obtained with this study but also some other parameters that must be defined in collaboration with our Library Committee.

Acknowledgments. I would like to acknowledge the help received at the library of the CINDOC-CSIC in Madrid in collecting the data. I thank very much Dr. Luis Corral and Dr. Jorge Sánchez for their invaluable advice with the data analysis. I am very grateful to Dr. Luis López who taught me enough LaTeX to enable me to prepare this article and to Terry Mahoney who revised and corrected my English and my LaTeX file. Finally, a very special thank you to my "boss", Dr. Carlos Martínez who let me attend LISA IV and prepare this article.