Volume Title??
ASP Conference Series, Vol. ??, 2002
???, eds.

How to Succeed in Astronomy without Having to Use a Telescope, or a Librarian's Guide to High Impact Papers

Sarah Stevens-Rayburn

Space Telescope Science Institute 3700 San Martin Drive, Baltimore, MD 21218; library@stsci.edu

Ellen N. Bouton

National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903-2475; library@nrao.edu

Abstract. We have undertaken a study of High Impact Papers in astronomy from a subset of ISI's 2001 database, covering papers published 1991–1999 and cited 1991–2000. The intent of the study was to measure the relative impact of papers based on data gathered from multiple telescopes as opposed to those from a specific telescope. The results have been somewhat surprising in that high citation rates are not as frequently tied to direct observational material as they are to theories, data compilations (either laboratory astrophysics or observational surveys), and reviews. We explore these observations, and suggest caveats in the use of citation information, based on differing counts for the same papers from ISI and the ADS.

1. Introduction

ISI, formerly the Institute for Scientific Information, is the world's best known resource for citation information in the sciences, both physical and social. They provide a huge number of products, at a correspondingly huge price, that allow one to slice, dice, dissect, and in general examine to the nth degree the publication patterns of individuals, departments, institutions, and, with a bit of effort, individual telescopes. In the current day, when individual telescope projects can run to millions and millions of dollars, those who provide those dollars are increasingly interested in knowing if they are getting their money's worth, whatever that may be. One of the metrics that is increasingly used to try to determine "worth" is documenting the number of times a particular paper has been used by other astronomers. As librarians, we are well aware of the pitfalls of such usage, since citation counting by definition is a quantitative measure and not a qualitative one. One infamous example of the difference this distinction can make is the excitement a few years ago about cold fusion and the physicists who claimed to have produced it. Their paper won a huge number of citations over the next year or so, as other physicists scrambled to prove the initial paper wrong (or right), and the original paper, after an initial high impact, has now faded into the obscurity it so richly deserved. Over a period of more than a couple of years, the "flashes in the pan" disappear and the really useful papers continue to be cited at much higher rates than expected for a merely average paper.

As librarians who have spent more time than we care to consider examining the journal literature and creating bibliographies and paper counts for ourselves, our administrators, and our funding agencies, we have developed certain suspicions about what contributes to "important" papers and what actually has a lasting impact on the science. We had our suspicions but no hard evidence, so we undertook this research to try to gather that evidence.

2. Underlying Assumptions/Suspicions

We admit to ulterior motives in undertaking this investigation. Observatory directors, funding agencies, and the media often seem bent on proclaiming the superiority of their particular facilities compared to all of the others, while ignoring what we consider the fundamental fact that all of modern day astronomy rests solidly on the foundation of work that has gone before. Similarly, the symbiotic relationships that exist among observatories are crucial to the ongoing increase in the body of astronomical knowledge. Looking at specific objects or phenomena at a variety of wavelengths and with coordinated observations often contributes substantially more than a single observation with a single instrument. Comparative charts of how many papers a particular observatory produces are often misleading in that they fail to acknowledge that much of what they are able to elucidate is highly dependent on previous investigations by others or even that although observatory x produced y number of refereed papers last year, 50% of those papers included observations made at other observatories. What we hoped to show by looking at the High Impact Papers was that, to shamefully paraphrase John Donne, "no telescope is an island," that acknowledging the symbiosis and the value of the perhaps less "flashy" observations is as important, and perhaps more so, than the papers that capture the headlines.

3. Methodology

One of ISI's many products is a dataset that provides a "snapshot" of which papers have garnered the most citations in a particular field. The High Impact Papers are the top 200 papers published in a specific year that have the most citations in that and subsequent years. This is a constantly shifting window, of course, as fads, fancies, and interests ebb and flow. The wave of the last few years has been gamma-ray bursts, but something different could take over at any moment. That's the excitement of astronomy for all of us. For this paper, as we noted above, our interest was not so much in following the fads, but rather the larger view. Therefore we looked at the High Impact Papers database in the ISI category "Space Science" that was produced in the spring of 2001. This included papers published 1981–1999 and cited 1981–2000. In order to keep our work to a reasonable level, we then focused on the top fifty papers in each year from 1991–1999, giving us a total of 450 papers (Figure 1). We examined each of these in detail, noting whether it was an observational paper or something else. The

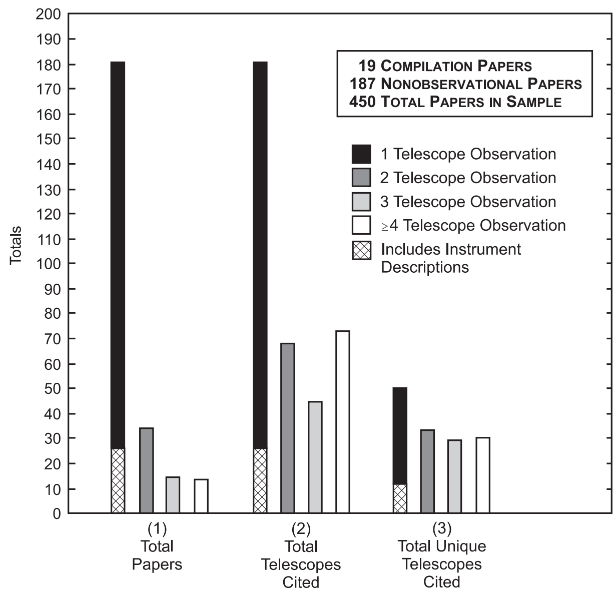

Figure 1.

"something elses" included review papers ("what's our current understanding of funky blue stars?"), techniques ("using *xyz* technique to force the data to come out my way"), laboratory astrophysics ("what can I do to get lots of citations without having to write observing proposals or travel to mountain tops?"), and tons of theory papers (wherein the theorists refuse to look at hard observational data that could conceivably contradict their theories). These papers represented about 40% of the total papers, not surprising given that historically, highly cited papers fall into the "useful" rather than the "fundamental" category.

For those papers containing observational data, we noted the number(s) and name(s) of the telescope(s) used. This turned out to be somewhat less simple than we had anticipated, given that astronomers delight in comparing their results to the results obtained by others investigating the same or similar objects and/or phenomena. To resolve some of the quandaries, we asked of each paper if such comparison observations contributed in a substantive way to the paper's conclusions, but even then, there were some tough calls. For instance, a paper describing *Keck* observations of the Hubble Deep Fields: is that a *Keck* only paper or *Keck* and *HST*? In such cases, we chose *Keck* only, since the *HST* contribution wasn't strictly observational and we tried to be consistent in our decisions, which resulted in looking at specific papers multiple times. Although it is unlikely that everyone performing this investigation would obtain exactly the same results, we believe that our conclusions are internally consistent and relatively bias free.

4. Excursis on Citation Data from ISI versus the ADS

During the investigative stage of this paper, we have learned a lot about the pitfalls and fallacies inherent in this citation counting business. There was the revelation earlier this year that ISI had failed to count most of the citations to one of the breakthrough genome papers, but those sorts of errors are relatively rare. What isn't so rare (and we realize the following is not news to a group of librarians, but we need to say it anyway for the record): astronomers sometimes get their references wrong. From an astronomical perspective, they likely believe it doesn't matter; after all, they're close, right? The cited references to the paper that boasts the largest number of citations in our study appear under several variations on the author's name, over a dozen renderings of the journal title, one citation with the correct page and volume but wrong journal, several dozen permutations of the volume and page numbers, and an interesting selection of publication years. References to the infamous cold fusion paper were rendered in 30+ ways. We point this out to demonstrate the seriously inexact nature of the pursuit of a "real" citation count. Equally mysterious to us is the disparity between citation counts depending on who is doing the counting. On this particular paper, ISI showed 1159 citations as of the end of 2000 and the ADS showed 871 as of late summer of 2001. Because the ADS is dealing with a much smaller "universe" of titles, they are usually better able to reconcile ambiguous references. ISI, dealing with some 5600 titles in a much broader number of disciplines, has to use much more caution in trying to do such reconciliations. The advantage in reconciliation that the ADS has is offset, however, by their not having access to the broader spectrum of citing journals.

Because we noticed that ISI and the ADS seemed to have such differences, we decided to add a systematic investigation of the disparities to our study. Keep in mind for this part of the discussion that our counts from ISI represent a one-time snapshot of citations made available in the spring of 2001, with citations through the end of 2000. We examined the first half of our list in the ADS in the late summer, so in theory the ADS had approximately six additional months to gather citations. For this portion, 1991–1994, in 80 cases the ADS had more citations than did ISI, and in 114 cases, ISI listed more cites. Now basically we have no real concerns in those cases where there are just a few numbers different between the two providers and this was often the case. What is troubling is the wide disparity that occasionally cropped up on both sides. The papers for 1995–1999 were checked in the ADS in late 2001. Because this was almost a year later than the ISI counts, we decided to get updated ISI counts in early 2002 and we found that for this more recent (and therefore more volatile) period, the ADS counts exceeded the ISI ones in only 14 cases. Overall, for the entire period 1991–1999, there were 14% more citations in ISI than in the ADS (ISI totaled 93,324 citations to the ADS' 81,657). Even this number isn't too bad on its face, but when one looks at specific cases, such as a 1999 article in *Science* that had 92 cites in ISI and 0 in the ADS or a 1991 article in *JGR-Space Physics* that had 344 cites in ISI and 3 in the ADS, one starts to get a bit uneasy. Likewise, a 1991 *ApJ* article has 302 cites in the ADS but only 267 in ISI. There are certain areas in which one service does better consistently than the other; for instance, all three of the examples of higher cited ISI papers were in planetary astronomy. The ADS, on the other hand, usually has faster and better citations to observational astronomy and astrophysical papers that are cited in what we know as the primary astronomical journals. And as we are all aware, the ADS is superior from both an ease of use and a cost standpoint. Nonetheless, it is really important to be aware of where each provider fails to deliver when we or others

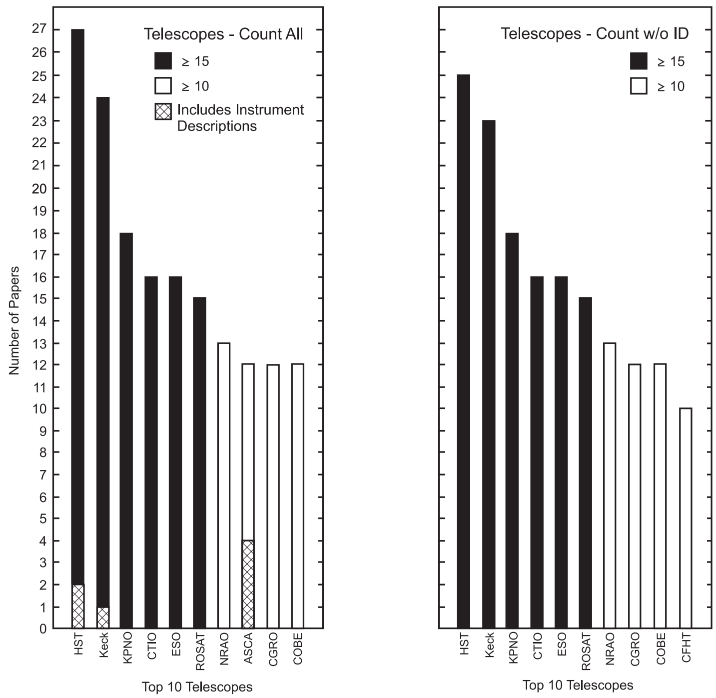

Figure 2.

use these numbers to try to make points about the impact of either telescopes or individuals.

5. A Look at the Numbers from a Variety of Positions

[Note that in the following section, we use the terms "telescope" and "observatory" interchangeably, but that in a few cases (*KPNO*, *MSSSO*, *ESO*, *NRAO*, *CTIO*, and perhaps others), there may be more than one physical telescope involved. For purposes of our discussion, we were concentrating on *observatory* statistics, since that is what most of us need to track. However, in cases where individual telescope statistics could change the rankings significantly, we have made note of it in our discussion.]

After linking specific telescopes to particular papers, we sorted them in several different ways to see what the numbers might be able to tell us and, we believe, they told us quite a lot. Looking first at simply the number of times a specific telescope was used, without regard to whether the use was in conjunction with another telescope or not, gave us a top ten ranking of *HST*, *Keck*, *KPNO*, *CTIO* and *ESO* (tied for fourth), *ROSAT*, *NRAO*, and *ASCA*, *CGRO*, and *COBE* (tied for seventh) (Figure 2, left panel). If we take the instrument description papers out of contention, ASCA drops out and CFHT moves in (Figure 2, right panel). As interesting as these results were, we felt they were not the true picture, since if a telescope only contributed a portion of the data, it ought to only get credit for a portion of the citations. Therefore, we introduced

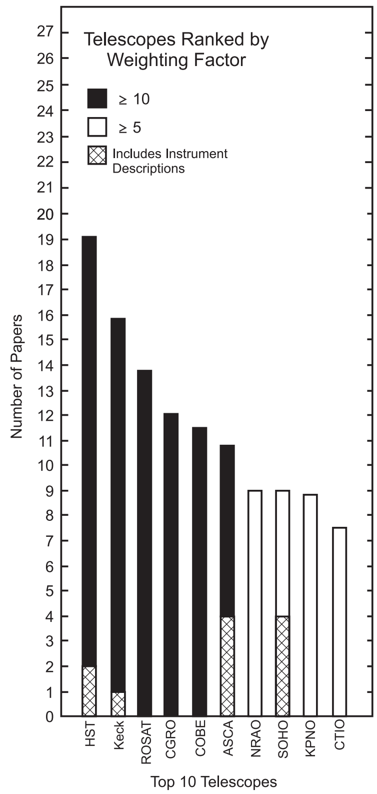

Figure 3.

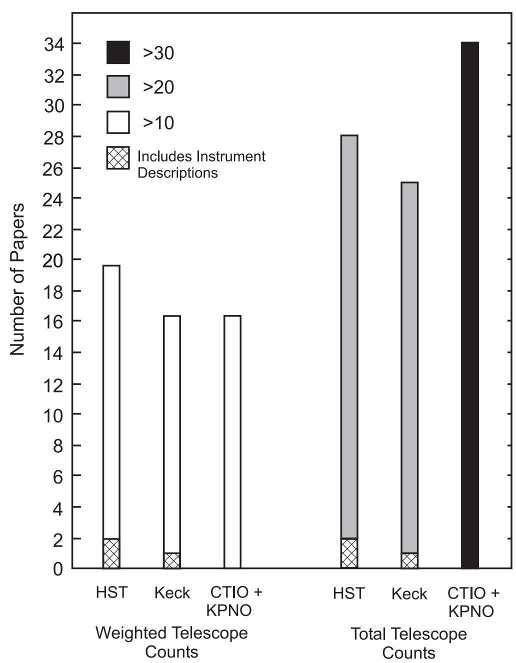

Figure 4.

a weighting system. Single telescope papers were given a 1; two telescopes, .5; three telescopes, .3; and more than three telescopes, .2. Combining the weighted totals for each telescope had an interesting effect on the rankings, but not as much as one might have supposed (Figure 3). *ESO* drops out of the top 10 and *SOHO* moves in. Using these various ways of adjusting the ranking, there were, as you can see, only a few shifts, so what surprises did we find?

We were actually surprised at how well the individual telescopes did in our study. We had anticipated that there would be just a few single observatory papers among the high impact ones but, in fact, single telescopes account for about 41% of the total (183 out of 450 papers) and 40% of the citations during the period, averaging 204 cites per paper. On reflection, however, we realize that these numbers aren't too surprising since the "top performers" all came online during the decade of our study. If one takes away the always highly-cited instrument descriptions, single telescope totals drop to 35% of the papers and only 33% of the citations. The real surprise, given that these exciting new instruments started producing results during the decade, is that the "second tier" did so well. For "best performance in a supporting role," one need look no further than the workhorse *NOAO* telescopes (Figure 4) that, combined, move *Keck* into third place in the weighted total papers calculation. [Note that if we differentiate specific telescopes at the ground-based observatories, two of the top 10 drop into the top 20 and one drops out of the top 20 altogether. However, since astronomers are not always specific about which instrument was used, we feel these dislocations are of only minor importance.]

Indeed, the papers using between two and seven telescopes, 13.5% of the total (61 of 450), averaged 218 citations per paper, suggesting that as outstanding as the "stars" are, even they gather more citations when used in conjunction with other telescopes. Another interesting comparison is to look at the citation records of the compilations. These are the papers that combine data from a variety of observations to produce catalogs and to draw wide-ranging conclusions about particular classes of objects. In this study, we classed those papers that used more than seven telescopes as compilations and these 19 papers, 7.2% of the total observational papers, received only 6.4% of the citations based on observations, averaging only 181 cites per paper. It may simply be that these papers gather citations more slowly and over a longer period of time, but we were surprised that they didn't reflect a larger impact.

Finally, back to our contention that "no telescope is an island": we did a comparison of all 70 specific telescopes cited in any of the High Impact Papers and found, interestingly, that a bit over half of the telescopes appear more often in multiple telescope papers than in single telescope papers. In fact, there are 22 telescopes in our study that *never* appeared on their own and yet when combined with other instruments, make a solid contribution to high impact and, we assume, extension of the science.

6. Conclusions

What conclusions can one draw from all of this? We facetiously entitled this paper "How to succeed in astronomy without having to use a telescope," and indeed one can glean information about that from our work. Of the 450 papers in our study, close to 42% of them didn't involve any interaction with a telescope. Reviews, theories, software descriptions, and laboratory astrophysics kept a significant fraction of the authors either at their desks or lying on a beach somewhere with a laptop computer and an active imagination. Authors taking this route should try to insure that they're also brilliant, however, since there are an awful lot of such papers that never get cited by anyone at any time.

The second way to succeed appears to be tied to being the author of an instrument description for some eagerly awaited telescope project. These papers may fade into obscurity over the long term, but for those younger astronomers during the critical tenure-track decade, such papers can be remarkably useful in boosting one's citation count.

Third, if you have a result you can report concisely, do seriously consider making it a *Letters* paper. Contrary to popular belief, *Letters* articles really do have staying power. Given that they make up such a small percentage of the literature, it's quite interesting that they make up 20% of the High Impact Papers.

Besides these somewhat tongue-in-cheek conclusions, there are a couple of serious ones. One is to not sell the ground-based "supporting" telescopes short. Some of the older and/or smaller telescopes are obviously continuing to make a strong contribution to the science. Of the 70 telescopes mentioned in these papers, 22 of them never stood alone and yet they obviously have made a strong contribution to the 48 others in order to end up being a part of these high impact papers. The other extremely important conclusion that we can draw is that studying the impact of a specific instrument in isolation is not actually appropriate, given all of the considerations we mention above, both with regard to potential pitfalls in the citation data and to the frequency with which important results are the product of cooperation and symbiosis, rather than rugged individualism.

Finally, be aware when your director or funding agent or users' committee asks for a "few simple numbers based on citation counts," that there are no simple numbers and that citation analysis is not for the faint of heart!

Acknowledgments. The authors thank Sharon Toolan (ST ScI) for preparing the manuscript and Pat Smiley (NRAO) for the figures. We also thank Karen Morrison of ISI for providing the citation information for the Fleischmann and Pons cold fusion paper. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.