Evaluating Astronomical Institutional Productivity Using the Astrophysics Data System Database

Eric Schulman
Institute for Defense Analyses
Alexandria, Virginia

James C. French and Allison L. Powell
University of Virginia Department of Computer Science
Charlottesville, Virginia

Abstract

We used the Astrophysics Data System (ADS) to measure the productivity of the 38 institutions studied by Abt (1993, PASP, 105, 794) during the period 1985 to 1994. The ADS database contains 84,822 astronomical papers published in *Astronomy and Astrophysics*, *The Astronomical Journal*, *The Astrophysical Journal*, *Monthly Notices of the Royal Astronomical Society*, and *The Publications of the Astronomical Society of the Pacific* during this period. For each of these papers we compared the affiliation of each author to a canonical list of ADS affiliations that we had created using automated and manual clustering strategies. We assumed—as did Abt (1993)—that each of the n authors on a paper should result in his or her institution getting credit for n^{-1} papers. We compared our results to those of Abt (1993) and determined that his results were not strongly affected by having neglected the European journals or by having used papers from only one six-month period in the decade 1985-1994.

I. Introduction

Researchers and research institutions are usually evaluated on the basis of how productive they are. In the commercial world, researchers produce products and profits. In academia, they produce papers. However, the evaluation of research institutions in a field such as astronomy can be quite difficult because consistent publication statistics across a number of institutions are hard to come by. Nevertheless, the data do exist in the form of the author affiliations that are listed in refereed journal articles. These data are easily accessible in the bound journals of most science libraries, but they are difficult to use in that form—in order to determine institutional productivities from these bound journals, one would have to read the title page of every article published in a number of journals over a number of years and record the author affiliations for each one. This was done by

Abt (1993) for *The Astronomical Journal* (AJ), *The Astrophysical Journal* (ApJ), and *The Publications of the Astronomical Society of the Pacific* (PASP) for the years 1952, 1962, 1972, 1982, and 1992. However, suggestions were made at the time of this publication that Abt's results could have been skewed by his ignoring major foreign astronomy journals such as *Astronomy and Astrophysics* (A&A) and *Monthly Notices of the Royal Astronomical Society* (MNRAS). In addition, he looked at papers from only one year per decade, which could have underestimated the decadal productivity for some institutions, and overestimated it for others.

During the course of our work on trends in astronomical publication using the Astrophysics Data System database (Schulman *et al*. 1997), we realized that the ADS database could be used to overcome these potential problems with Abt's work. In principle, the ADS database includes information on author affiliations. In practice, however, some papers lacked affiliation information, there were many different standards used in the affiliation field (e.g., "University of Virginia" and "Virginia Univ."), and there were numerous typos (e.g., "Charlottesvill" and "Charlottsville"). The methods we applied to create a canonical list of affiliation strings for the 38 institutions studied by Abt (1993) are chronicled in French *et al*. (1997ab, 1999). The results of using these canonical affiliation strings to study institutional productivity are reported below.

II. Method

We used the 84,822 papers contained in a snapshot of the ADS taken on September 12, 1997, to determine the institutional productivities of the 38 institutions studied by Abt (1993). We assumed—as did Abt—that each of the n authors on a paper should result in his or her institution getting credit for n^{-1} papers. For authors with two or more affiliations, each affiliation shares the credit for that author equally. In addition to the journals that Abt used (AJ, ApJ, and PASP), we used two major European journals: A&A and MNRAS.

One problem with the ADS database is that all papers do not have entries for all fields, especially the affiliation field. When the completeness is relatively high, this problem can be ameliorated by normalizing the productivity results by the fraction of papers with affiliation information for a given year and journal. However, when the completeness is low this can lead to unacceptably large errors. We found the ADS database to be sufficiently complete for affiliations during the period 1985 to 1994, and have therefore restricted our study to this time frame.

III. Results

The astronomical productivity for each of the 38 institutions is shown in Figure 1. The annual productivity ranges from a low of 2.5 papers to a high of more than 125 papers, with most institutions having productivities between 10 and 30 papers per year. It is important to remember that these productivities have been normalized by the number of institutional authors on each paper. The number of authors per paper has been increasing for some time and now has a median value of three (Schulman *et al*. 1997). As a result,

even authors who publish many papers may have relatively low productivities calculated using this method. For example, the lead author on the present paper published 11 papers in ApJ, AJ, and PASP between 1993 and 1997, but his average annual productivity—and therefore his contribution to his institution's productivity—was only 0.65 during this time.

While the raw productivities are interesting, the purpose of this work is to determine whether the potential flaws in the study of Abt (1993) actually manifested themselves. To that end, we divided the 38 institutions into 13 different "productivity ranks" on the basis of their total productivity between 1985 and 1994 in A&A, AJ, ApJ, MNRAS, and PASP. Institutions that do not have significantly different productivities from each other are considered to have the same "productivity rank." We then compared these ranks to the ranks that Abt (1993) determined using the 1992 productivity in AJ, ApJ, and PASP (Figure 2). It is obvious that the ADS ranks and the Abt (1993) ranks are highly correlated. Table 1 shows the data in more detail, and includes the ADS productivity for 1985-1989 and 1990-1994 as well as for 1985-1994. Of the 38 institutions in these studies, for 30 the ADS and Abt rankings differed by no more than one rank. There were three institutions that were underrated by Abt by two ranks: the University of Arizona, the University of California at Berkeley, and the University of Texas. Five institutions were overrated by Abt: the Institute for Advanced Study and the Johns Hopkins University by two ranks, and the University of Virginia, the University of Ohio, and Yale University by three ranks.

IV. Conclusions

We used the Astrophysics Data System database to study the institutional productivities for 38 institutions during the 10-year period from 1985 to 1994, using three American and two European journals. We compared our results to those of Abt (1993) and determined that his results were not strongly affected by having neglected the European journals or by having used papers from only one six-month period in the decade 1985-1994.

References

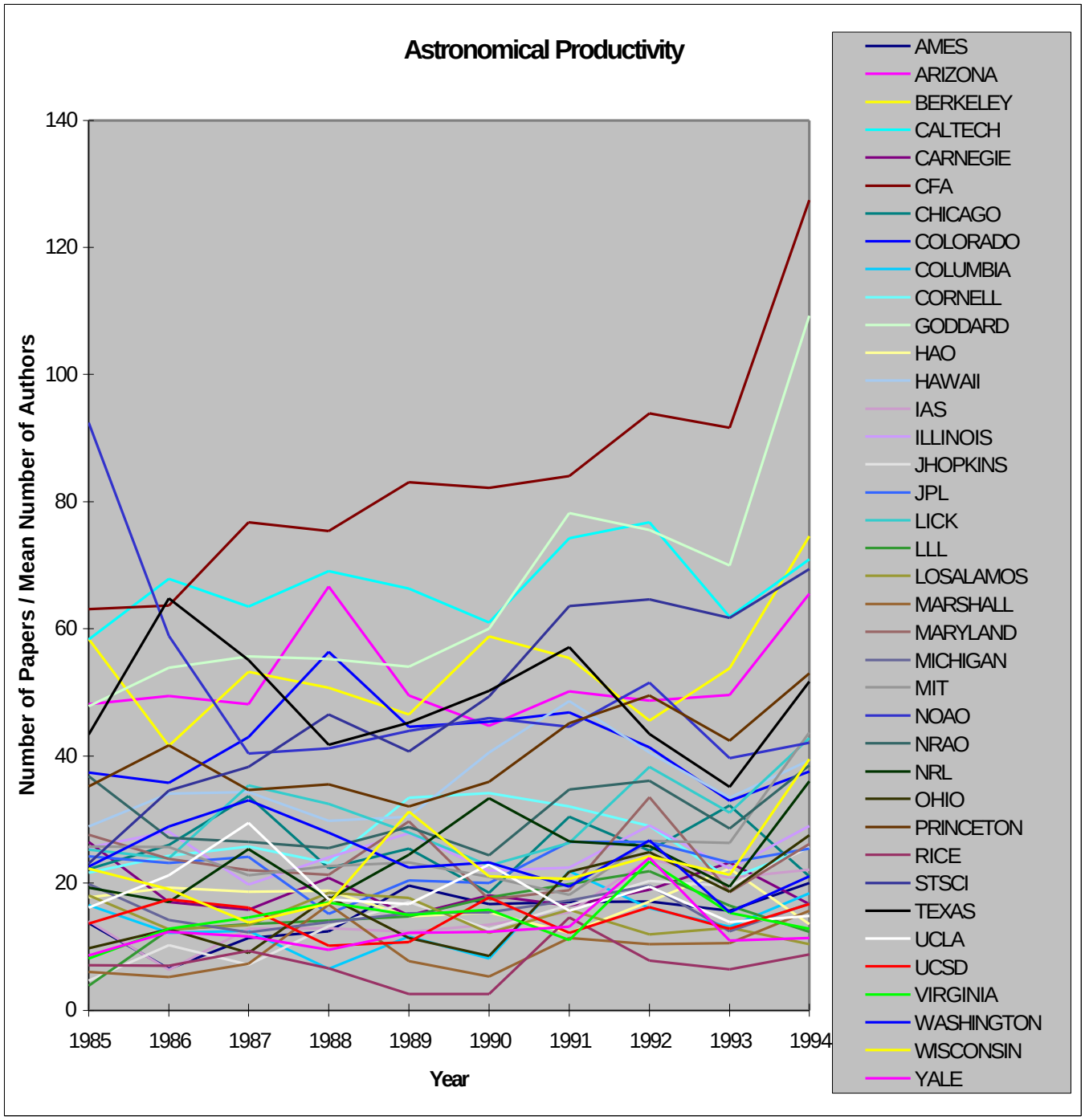
Abt, H. A., "Institutional Productivities," *Publications of the Astronomical Society of the Pacific*, **105**, 794 (1993)

French, J. C., Powell, A. L., and Schulman, E., "Applications of Approximate Word Matching in Information Retrieval," in *Proceedings of the Sixth International Conference on Information and Knowledge Management*, ed. F. Golshani and K. Makki (Association for Computing Machinery, New York), p. 9 (1997a)

French, J. C., Powell, A. L., and Schulman, E., "Clustering Strategies for Creating Authority Files," submitted to the *Journal of the American Society for Information Science* (1999)

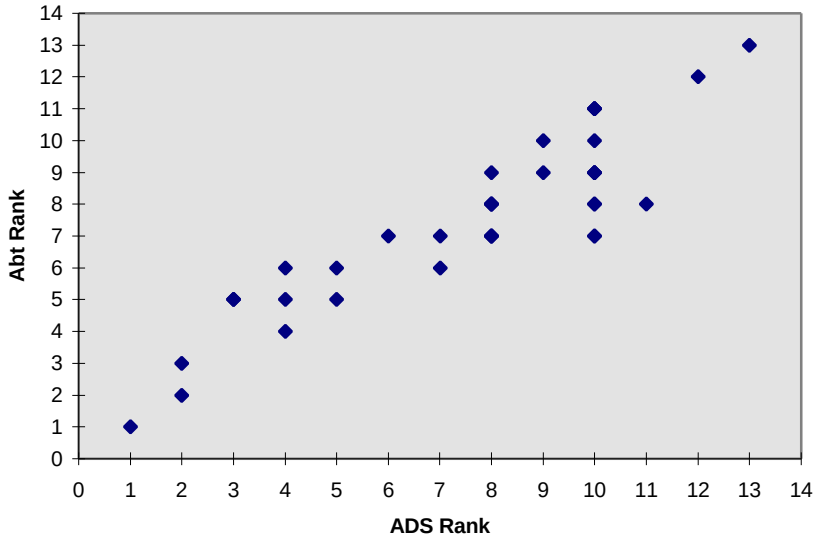
French, J. C., Powell, A. L., Schulman, E., and Pfaltz, J. L., "Automating the Construction of Authority Files in Digital Libraries: A Case Study," *in Research and Advanced Technology for Digital Libraries*, ed. C. Peters and C. Thanos, Springer-Verlag Lecture Notes in Computer Science, **1324**, 55 (1997b)

Schulman, E., French, J. C., Powell, A. L., Eichhorn, G., Kurtz, M. J., and Murray, S. S., "Trends in Astronomical Publication Between 1975 and 1996," *Publications of the Astronomical Society of the Pacific*, **109**, 1278 (1997)


Astronomical Productivity
Number of Papers / Mean Number of Authors
0
20
40
60
80
100
120
140
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
Year
AMES
ARIZONA
BERKELEY
CALTECH
CARNEGIE
CFA
CHICAGO
COLORADO
COLUMBIA
CORNELL
GODDARD
HAO
HAWAII
IAS
ILLINOIS
JHOPKINS
JPL
LICK
LLL
LOSALAMOS
MARSHALL
MARYLAND
MICHIGAN
MIT
NOAO
NRAO
NRL
OHIO
PRINCETON
RICE
STSCI
TEXAS
UCLA
UCSD
VIRGINIA
WASHINGTON
WISCONSIN
YALE

	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994
AMES	13.71	6.47	11.37	12.4	19.57	16.71	17.05	17.09	15.67	20.02
ARIZONA	48.17	49.4	48.12	66.6	49.54	44.78	50.12	48.73	49.63	65.55
BERKELEY	58.35	41.61	53.21	50.73	46.46	58.81	55.44	45.54	53.81	74.59
CALTECH	58.3	67.81	63.49	69.02	66.32	60.98	74.26	76.71	61.91	70.92
CARNEGIE	26.49	17.01	15.77	20.85	14.92	18.12	16.47	18.97	23.27	16.7
CFA	63.12	63.69	76.79	75.37	83.03	82.18	84.02	93.9	91.63	127.41
CHICAGO	22.22	26	33.68	22.38	25.38	18.44	30.45	25.1	32.23	21.06
COLORADO	37.4	35.77	42.94	56.35	44.57	45.35	46.86	41.36	32.96	37.6
COLUMBIA	16.54	12.12	12.49	6.48	11.57	8.08	21.92	16.1	13.09	18.34
CORNELL	21.6	24.27	25.76	23.15	33.42	34.12	32.07	28.97	20.65	25.6
GODDARD	47.86	53.89	55.63	55.26	54	60.01	78.26	75.51	69.99	109.25
HAO	17.77	19.25	18.6	18.74	14.7	15.37	12.17	17.09	22.29	13.48
HAWAII	28.88	34.11	34.27	29.81	30.38	40.52	48.63	40.58	33.43	39.34
IAS	14.07	6.34	12.81	12.91	12.33	13.46	15.44	23.83	20.82	22.09
ILLINOIS	25.57	28.14	19.73	24.07	27.82	21.96	22.47	29.04	22.33	28.97
JHOPKINS	4.65	10.28	7.12	13.19	17.07	12.76	16.53	20.32	19.5	25.63
JPL	24.23	23.17	24.14	15.17	20.37	20.03	26.45	26.5	23.26	25.38
LICK	25.3	23.93	35.42	32.41	27.99	22.76	26.28	38.3	31.12	42.86
LLL	3.83	12.62	13.52	14.12	14.78	17.76	19.84	21.89	16.43	12.29
LOSALAMOS	18.05	12.7	13.4	18.47	17.68	12.21	15.82	11.94	13	10.38
MARSHALL	6.04	5.22	7.27	16.6	7.76	5.28	11.37	10.44	10.56	15.69
MARYLAND	27.64	23.84	21.99	21.27	29.71	17.39	18.88	33.52	18.58	26.17
MICHIGAN	19.9	14.22	12.25	13.87	15.28	15.37	17.31	19.58	12.39	16.76
MIT	25.72	25.63	21.18	22.67	23.26	20.94	18.16	26.65	26.28	43.68
NOAO	92.42	58.88	40.4	41.24	43.92	45.97	44.56	51.56	39.66	42.08
NRAO	36.83	27.15	26.48	25.54	28.76	24.36	34.68	36.09	28.6	38.52
NRL	19.32	17.04	25.32	17.43	24.48	33.37	26.53	25.78	19.47	35.99
OHIO	9.72	12.67	9.05	17.82	11.32	8.57	21.74	24.83	18.59	27.5
PRINCETON	35.19	41.68	34.63	35.5	32.04	35.94	45.14	49.5	42.41	52.96
RICE	7.08	7	9.3	6.57	2.58	2.57	14.58	7.83	6.42	8.82
STSCI	23.12	34.52	38.27	46.54	40.73	49.37	63.56	64.61	61.75	69.41
TEXAS	43.39	64.77	55.12	41.76	45.22	50.24	57.08	43.44	35.13	51.67
UCLA	16.13	21.24	29.5	17.4	16.7	22.92	15.62	19.43	13.87	14.85
UCSD	13.62	17.43	16.13	10.17	10.72	17.74	12.18	16.19	12.81	16.67
VIRGINIA	8.14	12.82	14.56	16.94	15	15.84	11.04	23.4	15.33	12.87
WASHINGTON	22.6	28.95	33.02	27.9	22.45	23.26	19.47	26.76	15.37	21.11
WISCONSIN	22.27	19.03	13.78	16.78	31.23	21.1	20.71	24.23	21.28	39.43
YALE	8.61	12.29	11.58	9.52	12.17	12.33	13.17	23.99	10.96	11.32

Figure 2: ADS vs. Abt Rankings



Rank		Institution	Productivity			
ADS	**Abt**		**85-89**	**90-94**	**85-94**	**Abt92**
1	1	CFA	76.25	117.25	96.75	89.90
2	2	GODDARD	56.16	95.96	76.06	77.20
2	3	CALTECH	68.42	82.52	75.47	64.50
4	4	STSCI	38.56	74.82	56.69	57.40
4	5	NOAO	58.34	53.31	55.83	48.40
5	5	PRINCETON	37.70	54.67	46.19	48.00
3	5	ARIZONA	55.12	63.11	59.11	47.10
3	5	BERKELEY	52.76	70.34	61.55	46.40
4	6	TEXAS	52.72	56.45	54.58	41.90
5	6	COLORADO	45.70	48.30	47.00	40.30
7	6	LICK	30.57	39.45	35.01	35.20
8	7	ILLINOIS	26.38	30.10	28.24	30.50
8	7	NRL	21.84	33.84	27.84	30.10
8	7	MARYLAND	26.21	27.41	26.81	30.10
7	7	NRAO	30.50	39.11	34.81	28.50
8	7	CORNELL	27.00	33.26	30.13	27.00
10	7	VIRGINIA	14.20	18.69	16.45	26.30
10	7	OHIO	12.74	24.65	18.69	26.00
6	7	HAWAII	33.17	48.08	40.63	26.00
8	8	MIT	24.95	33.77	29.36	24.90
8	8	CHICAGO	27.34	30.88	29.11	24.90
8	8	WASHINGTON	28.43	25.04	26.73	24.10
8	8	WISCONSIN	21.70	31.16	26.43	23.70
10	8	IAS	12.32	23.36	17.84	22.10
10	8	JHOPKINS	10.99	23.31	17.15	21.90
11	8	YALE	11.41	16.76	14.09	21.40
8	9	JPL	22.57	29.25	25.91	19.10
10	9	MICHIGAN	15.91	19.30	17.60	19.00
9	9	CARNEGIE	20.02	22.80	21.41	18.80
10	9	LLL	12.39	20.77	16.58	18.60
10	9	UCSD	14.35	18.15	16.25	17.80
9	10	UCLA	21.30	20.46	20.88	16.80
10	10	AMES	13.38	20.90	17.14	16.60
10	11	COLUMBIA	12.49	18.59	15.54	13.50
10	11	HAO	18.76	19.75	19.25	13.00
10	11	LOSALAMOS	16.91	15.12	16.02	12.10
12	12	MARSHALL	9.02	13.15	11.09	10.00
13	13	RICE	6.86	9.54	8.20	6.50