CITATION ANALYSIS AND THE NASA ASTROPHYSICS DATA SYSTEM (ADS)

P.C. van der Kruit
Kapteyn Astronomical Institute
University of Groningen
www.astro.rug.nl/∼vdkruit

Part 1. General issues and the use of ADS for bibliometric studies (written in August 2003, with additions in August 2004 and 2005)

1.1 Introduction

There is considerable interest in citations to scientific publications as a means to provide statistics that can be used for widely varying purposes such as those related to science policies at national levels, funding within universities, employment issues and promotions of scientists, etc. I have been interested in this for a long time and have published two papers in refereed journals on these subjects[1]. I review and discuss some issues related to bibliometric studies and investigate to what extent the NASA Astronomy Data System can be used for citation analyses.

1.2 Previous work

In my first paper[2] I collected various data for 15 member countries of the Organisation for Economic Co-operation and Development (OECD) that have to do with the practising of astronomy: (1) using the report of the Astronomy expert meeting of the Megascience Forum of the OECD, the level of astronomy funding, size of the research communities, relative commitment to ground-based versus space-based astronomy, etc.; (2) from other sources the size of the population, Gross National Product and size of the total research community; (3) from the paper of Schubert *et al.*[3] data on publication and citation scores in the *Science Citation Index* of these countries in astronomy and the total research effort (excluding social and economic sciences).

In the study I used the following parameters: The number of researchers as a fraction of the total population, the budget per head of the population, the budget available per researcher, the astronomy budget as fraction of the GNP, the citation rate (citations per

[1] The fact that these are published in refereed journals does not necessarily make this scientific research and I prefer myself to refer to this –as this note itself– as my "recreational research".

[2] P.C. van der Kruit: *A comparison of astronomy in fifteen member countries of the Organisation for Economic Co-operation and Development*, Scientometrics **31**, 155-172 (1994). The full paper is available on my Web-site.

[3] A. Schubert, W. Glänzel and T. Braun: *Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields in 1981-1985*, Scientometrics **16**, 3-478 (1989).

Table 1: Summary of various rankings of OECD countries.

	Astronomy Funding	Astronomy Output	Astronomy Budget-Output	All Science Output	Astronomy-All Science
1	Finland	Netherlands	Canada	Switzerland	Netherlands
2	France	USA	Netherlands	Sweden	Australia
3	Switzerland	Germany	Australia	USA	Italy
4	Germany	Australia	USA	Denmark	Germany
5	UK	Switzerland	UK	UK	USA
6	Italy	UK	Denmark	Netherlands	France
7	USA	Canada	Sweden	Canada	UK
8	Netherlands	Italy	Belgium	Belgium	Switzerland
9	Australia	France	Germany	Germany	Canada
10	Belgium	Belgium	Switzerland	Finland	Belgium
11	Spain	Denmark	Japan	Australia	Spain
12	Denmark	Sweden	Italy	France	Japan
13	Japan	Japan	France	Japan	Finland
14	Sweden	Finland	Spain	Italy	Denmark
15	Canada	Spain	Finland	Spain	Sweden

paper), the number of citations per researcher, the number of publications per researcher, and three parameters calculated by Schubert *et al.* (*op. cit.*) from which I also took the relevant data. These are (1) *the relative citation rate*: first an expected citation rate is calculated for each country which is based on the number of publications and the citation rates for the journals in which these papers have appeared; the relative citation rate then is the ratio between the observed and expected citation rates; (2) *the activity index*: for each country this is the ratio of the world share it has in the number of publications in astronomy and that share for all sciences; (3) *the attractivity index*: for each country this is the ratio of the world share it has in the number of citations in astronomy and that share for all sciences.

Using these data the 15 countries have been ranked on the following aspects (see table 1): (1) the relative level of astronomy funding; (2) the relative level of performance in astronomy; (3) the correspondence between funding and performance in astronomy; (4) the relative level of performance of the total science effort; and (5) the performance in astronomy relative to that in all sciences. Remember that all of this refers to the early nineties.

Other interesting results that can be inferred from the data collected in this paper are: (1) one out of every 75,000 inhabitants of these OECD countries is an astronomical researcher; (2) each citizen of these countries spends on average 2.5$ per year on astronomical research (either from the ground or in space); (3) the average budget per researcher amounts to roughly 200,000$ per annum; (4) the average budget for astronomy amounts to 0.016% of the Gross National Product and of order 1% of the total budget for civilian R & D; (5) an astronomical researcher from these countries produces on average 1.7 papers each year and these papers receive on average ten citations in the first five years; (6) researchers in science (excluding economic and social sciences) make up 0.08% of the population in these countries and one in about 65 of these researchers works in astronomy or astrophysics; (7) most countries spend about one-third of their astronomy budget on salaries, one-sixth on basic support and half on observing facilities (in a ratio one to two for ground-based versus

Table 2: Comparison of mean citation rates and sample sizes to Trimble's studies.

(Sub-)group	AAS		RAS		NL	
	cits/yr	#	cits/yr	#	cits/yr	#
All	28	393	22	188	36	103
Universities	40	162	28	78	39	74
U+O/I/L	33	303	26	148	36	103
		%		%		%
Theory	28	34	30	35	26	17
Observations	31	55	20	49	41	71
Instruments	19	12	14	17	9	12
Sun	27	9	12	6	17	8
Solar system	24	12	17	5	22	8
Dynamics	26	4	17	10	–	–
Stars	22	27	24	25	38	31
MW+ISM	29	13	26	14	35	21
Galaxies	51	13	36	13	50	19
HEAP	36	17	27	18	31	8
Cosmology	73	5	40	10	26	3
X+γ	30	9	27	21	25	15
UV+IR	24	8	11	13	49	15
Optical	29	47	25	41	43	31
Radio	31	26	13	26	36	39

space).

In a second paper[4] I looked into citation scores of individual researchers in the country. For this a list has been compiled for essentially all scientific staff at Netherlands astronomical institutes with their first author citations in the period 1986 to 1992, using the Science Citation Index. In addition for each person for the year 1992 the number of cited publications has been compiled, their earliest, most frequently and the most recent cited paper. Various conclusions can be drawn from this: (1) the earliest cited paper usually is from near the year of the Ph.D.; (2) the most recent cited (first author) paper is usually less than 5 years old; (3) a personal high citation rate is generally a result of a reasonable number of citations to most of ones papers rather than due to a single, highly cited paper; (4) Netherlands astronomy is most visible in observational studies, especially concerning stars, interstellar medium and our own and other galaxies and using in particular UV, IR and optical facilities. Comparison in mean citation rates determined in other studies are also made, resulting in an excellent score for the Netherlands astronomical community.

This study was triggered by a citation analysis that was performed by the CWTS (see below) commissioned by the Netherlands Foundation for Research in Astronomy (ASTRON, of which was at that time chairman of the Board) and two studies that Virginia Trimble[5] had published in the same journal. She had taken samples of astronomers at random from the membership directories of the professional societies for the USA (the American Astronomical Society, AAS) and the UK (the Royal Astronomical Society, RAS). She and I

[4]P.C. van der Kruit: *The astronomical community in the Netherlands*, Quart. J. Roy. Astron. Soc. **35**, 409-423 (1994). This paper is also available on my Web-site.

[5]V. Trimble, Quart. J. Roy. Astron. Soc. **34**, 235 and 301 (1993).

compared for researchers the average values of their number of citations per year in the Science Citation Index in various categories. For institutions we have Universities (U) and research institutes (observatories, institutes and laboratories or O/I/L). Further categorisations were theory, observational or instrumentational, area's of astronomy (MW+ISM means the Milky Way Galaxy and interstellar medium; HEAP is high-energy astrophysics) and wavelength coverage of instruments used. The result is displayed in table 2.

1.3 Henk Spruit's Curve of Growth

Henk Spruit[6] found that the curve for each astronomer in the total annual number of citations to first author papers in the Science Citation Index as a function of years after his/her Ph.D. has a fairly uniform shape. He determined that curve using citation data of 72 astronomers of the Netherlands nationality that were active (held research positions anywhere in the world) at the time of writing. Each individual can then be fitted to this *"curve of growth"* (a term from astronomical spectroscopy) with a single amplitude factor α. The average astronomer has an α equal to unity. Henk Spruit has updated this most recently in 2000, using a sample of 176 astronomers of Dutch nationality or working at institutes in the Netherlands. My α then came out as 3.07 ± 0.19 (the sixth from the top of these 176 astronomers)[7].

1.4 The CWTS method and the NOWT

The use of citations per paper is an obvious and widely used parameter. Yet it is impossible to compare even subfields in a certain disciplines, let alone comparisons between fields of science. Various solutions to this can be found such as for examples the use of the activity and attractivity indices of Schubert *et al.* described above. The *Center for Science and Technology Studies* (Centrum voor Wetenschaps - en Technologie Studies, CWTS) under the direction of A.F.J. (Ton) van Raan has developed a method where the number of citation of each publication (where possible) is being normalised to average citation scores of similar publications. These normalising publications are in principle those in the same years in either the same journal or in a mix of journals together covering the relevant field of science. The CWTS method is widely used in the Netherlands.

The CWTS defines the following parameters:

– *P*: the number publications that can be used for the method (usually only papers in international refereed journals).

– *C*: the number of citations to those publications. Self-citations are excluded (a self-citation is defined in the strict sense that it involves the same (co-)author in both the citing and cited publication). The percentage of self-citations usually is of the order of 25%, at least in astronomy.

– *CPP*: the number of citations per publication (citation rate). Again self-citations are excluded.

– *JCSm*: The average citation rate of all papers published in the same journals in the same years (also corrected for self-citations).

[6] H.C. Spruit: *A 'curve of growth' for astronomers on the Citation Index*, Quart. J. Roy. Astron. Soc. **37**, 1-9 (1996). This paper is available on the ADS (see below for the URL).

[7] Spinoza winner Ewine van Dishoeck came out of course ahead of me at 4.04 ± 0.20; the other Spinoza winner Ed van den Heuvel had 2.44 ± 0.15. From the Kapteyn Institute Xander Tielens had 3.59 ± 0.27 and Peter Barthel 2.02 ± 0.18. *Note added in 2004*: This year's Spinoza winner Michiel van der Klis had 2.80 ± 0.35.

– *FCSm*: The average citation rate of all papers published in journals in the relevant (sub)field of science in the same years (corrected for self-citations).
– *CPP/JCSm* and *CPP/FCSm*: The impact of the papers with respect to either the citation rate in the same set of journals or the average world citation rate in the same field.
– *JCSm/FCSm*: The relative impact of the journals in which the papers are being published.

Impact ratios $CPP/JCSm$ and $CPP/FCSm$ larger than unity indicate performance above world average. The CWTS assigns a significance level to this, but I have heard critisism of the validity of this. It has been pointed out that the CWTS does not properly allow for the fact that the distributions in this sort of analysis are very skew, even to the extent that the intrincic distributions may have higher order moments that are infinite. Important to note also is that the CWTS uses the same time windows for publication of the cited and citing papers and that is often is 5 years or less. That means that a basic assumption is that the citation profiles of all papers are the same, an assumption that is particularly important for papers published recently in the time window.

A first study of the Faculty of Mathematics and Natural Sciences at the University of Groningen was commissioned in 1996 (when I was Dean) and completed in 1997. My reason for initiating that had to do with the fact that I wanted to use measures of quality for allocation of research funding. In addition to the results of national peer review visiting committees (organized at the time per discipline by the association of Dutch universities VSNU) I wanted to have a look at the citation scores of the leading researchers in the research units. This approach was later called "back-to-the-future" by the CWTS. In that approach all publications of the permanent staff members were considered independetn of their affiliation at the time. It involved the time window 1980 to 1995, which was as far back as the CWTS could do the study. At that time the analysis was not sophisticated enough to correct the citation data for self-citations.

My own statistics came out as $P = 31$, $C = 1223$, $CPP = 39.45$ and the impact ratios were $CPP/JCSm = 3.28$ and $CPP/FCSm = 2.89$. My self-citation rate was only 11.1%.

Recently (2003) an update study was performed, again in a study commissioned by the Faculty. This time it concerned the period 1992-2001. My statistics now were $P = 11$, $C = 136$ (this time corrected for self-citations), $CPP = 12.36$ and the impact ratios were $CPP/JCSm = 2.64$ and $CPP/FCSm = 1.80$. My self-citation rate was 15.0%. My work with the most impact occured in the first half of the eighties, while my administrative work strongly increased during the last part of the eighties up till now.

My impact ratios compared to journals is higher than compared to the field. That is because I had not much choice than to publish in the European journal *Astronomy & Astrophysics*[8], which has a lower citation rate then US journals.

I have two problems with the CWTS method. The first concerns the way that papers published in *Nature* or *Science* are treated. Astronomical publications in *Science* are rare (sometimes one finds astronomical review papers, but an astronomical research paper is exceptional), while those in *Nature* are often submitted to that journal as an equivalent to a "Letter to the Editor" in astronomical journals. Ed van den Heuvel has pointed out that –in contrast to astronomical journals in general–, only 10 to 15% of astronomy articles submitted to *Nature* are accepted. Many breakthoroughs in astronomy and astrophysics have indeed appeared in *Nature*. Therefore there is no question that it is prestigeous to

[8] *Astronomy & Astrophysics* is sponsored by most West-European countries except the UK. Each country pays a contribution in ratio to its Gross National Product. In the Netherlands the institutes contribute to the national contribution roughly in proportion to the number of papers published. In return astronomers from these European countries don't have to pay page charges. As a result, most of us are forced to publish in that journal in spite of the somewhat lesser standing that it has and the fact that it is read less in the USA. However, with modern preprint servers and ADS (see below) this latter effect has mostly disappeared.

publish in *Nature*, but in my opinion not to the extent that I notice in other disciplines. In any case, the number of citations that papers in molecular biology, biochemistry and other fields of research recieve for a publication in *Nature* is much higher than in astronomy. The citation rate that the CWTS uses for the normalisation is over the whole journal and this is certainly not representative for astronomy[9]. Especially the Amsterdam group (working on compact object, black holes in compact binary systems and gamma-ray bursts) are being disadvantaged by this. I have asked the CWTS on a few of occassions to see if it is not possible to find the expected number of citations in *Nature* by selecting them according to discipline or at least conduct a study to estimate the extent of the problem, but I have not seen any such study.

A second problem arises as follows. The journal *Astrophysics and Space Science* (A&SS; a Kluwer journal) does not have a very good reputation in astronomy. It is not cited much and papers in it are often rejected first by the major journals. The editors have over the last ten years or so changed the policy of the journal. Often nowadays proceedings of conferences are published in it (as single issues or volumes). The journal is listed as a refereed journal (which it is for the research papers), so gets included in CWTS analyses. But often it contains one- or two-page abstracts of conference poster papers (refereed only by the conference organising committee) and invited reviews. Abstracts get very few, if any, citations. In practice people tend to refer to research papers and not the reviews at conferences. The citation rate of the journal is not high, and the inclusion of these contributions is not correct[10]. Of course, the CWTS in its verification round with the researchers (the CWTS sends around a list of the papers they find of each researcher and lets the researchers check these) does not allow deletion of records found, only additions.

Note added in 2005: The Faculty of Mathematics and Natural Sciences has again commisioned a study of its staff by the CWTS, now concerning the ten-year period 1994 to 2003. This time I was able to convince them that reviews and conference contributions in *Astrophysics & Space Science* should not be included. The CWTS determined my $CPP/JCSm$ as 2.67 and $CPP/FCSm$ as 2.04.

Although I support use of citation analyses I stress that proper use involves much care. Too easily wrong conclusions are being drawn. The next paragraphs are from notes that I prepared as part of a reaction by the Board for the Exact Sciences (Gebiedsbestuur Exacte

[9]In 1987 I was co-author on a paper in *Nature*. Roelof Bottema (a student of mine) was first author and it involved the observation of an edge-on spiral galaxy in Westerbork, displaying a highy unusual "warp" in the neutral hydogen gas layer. The paper was called: *"The progious warp of NGC4013"* and the issue had a picture of it on the cover. According to ADS (see below) it has received 25 citations. Other astronomical papers in *Nature* of that year (there were 25 in ADS) had a citation rate of 10.0 citations per paper. Roelof published –delayed by a serious illness– a much more involved and detailed analysis of these data in 1995 and 1996. These papers have now 18 and 16 citations respectively, while the average for the journal volumes were 14.62 and 11.26. So, the *Nature* paper attracted relatively much attention, and the work involved indeed was of good quality. But 25 references in 16 years (or 10 for the average astronomical *Nature* paper) is not a very high rate by any means.

[10]In 1994 I organised together with my colleague Gerry Gilmore from Cambridge a symposium in the series of the International Astronomical Union. Kluwer published all IAU proceedings and thus did so with ours on "Stellar Populations". It was a highly succesfull volume and Kluwer actually ran out of stock in 1998, which had never happened before. I agreed that the volume would be reprinted, but only the invited reviews and I gave these authors the chance to add a maximum two-page update, which half of them did. Then Kluwer decided that the reprint would not be profitable and instead publised the reprint as a volume of A&SS in 1999. On the *ISI Web of Science* I find eight papers with my name on it since 1999. Five of these are in A&SS. Two are the preface to the reprinted proceedings and the original preface. Another is my own review with a short update and a fourth one is an extended summary of a panel discussion during the symposium, which was not changed at all. For such publications you don't expect to get many citations, but the inclusion of these in a citation analysis of research articles in refereed journals is clearly wrong.

Table 3: Origin of Dutch papers and citation impacts for natural sciences.

Discipline	Fractional papers Total	Univ. (%)	OZI's (%)	Indust. (%)	Citation impact Total	Univ.
Physics	2250	78	14	8	1.36	1.09
Chemistry	2231	79	8	13	1.38	1.17
Biology	1200	67	32	2	1.11	1.04
Earth sciences	575	59	38	4	1.07	1.08
Computer science	522	78	13	9	1.05	1.23
Materials science	470	69	17	14	1.26	1.16
Mathematics	466	89	9	2	1.06	1.03
Astronomy and astrophysics	391	62	36	2	1.19	1.30
Chemical engineering	254	78	5	17	1.37	1.31
Statistical methods	182	87	10	3	0.97	0.92

Wetenschappen) of NWO (Netherlands Organisation for Scientific Research) on the 2000-version of the "Nederlands Observatorium voor Wetenschap en Technologie":

It is already some time ago that the *Netherlands Observatory of Science and Technology – Science and Technology Indicators 2000* was published. I will use the acronym from the Dutch title (NOWT) in what follows. The relevant parts of the report are contained in (their) Tables 3.5 and 3.8 and in Figure 3.6 (Table 3.5 appears as Table 10 in the English summary). The numerical information, on which these tables are based, is available in the statistical appendix and therefore only known to those who have taken the trouble to download this from the CWTS Website.

The approach shows the usual problem of a compromise between a study over a long timescale (as would be necessary for a thorough study) and the wish of administrators to have a measurement of the scientific quality now, or if not possible at least last year. The window of 5 years is small; the procedure is correct however in that it normalises in an appropriate way, but assumes that every paper has the same citation profile.

The NOWT notes a dominance of research at the universities, as judged from the origin of papers. Of all (fractional) papers in 1997 and 1998, 75% is authored by researchers at universities, 20% at publicly funded research institutes outside the universities (OZI's) and 5% at industrial or privately owned laboratories. Table 3 gives the numbers for the natural sciences together with the citation impacts[11] (total and universities only), taken from the statistical appendix. Examination of this information leads us to the opinion that the methods used in the NOWT do not give a clear picture of the performance of Dutch research, especially in astronomy and computer science.

For astronomy and astrophysics table 3 suggests that somewhat less than two-thirds of astronomical and astrophysical papers come from the universities, while more than one-third would originate from publicly funded institutes outside the universities. The NOWT notes (page 54, left column) that *"the substantial contribution of non-university publicly funded research establishments in the disciplines Astronomy* [...] *can for the larger part be attributed to ESTEC"* (the rest must be publications in astronomical journals from the ASTRON and SRON institutes). ESTEC is manifestly not part of Dutch astronomy. In fact, Dutch astronomy is the Netherlands Research School for Astronomy (NOVA), which constitutes the four astronomical institutes at the universities of Amsterdam, Groningen, Leiden and

[11] These are in effect the parameter $CPP/FCSm$ in usual CWTS studies.

Table 4: Specialisation and citation impact for broad categories of disciplines.

Discipline	Share %	Specialisation	Citation Impact Total
Social sciences	7.2	1.48	0.98
Life sciences	20.4	1.14	1.08
Medical sciences	35.5	1.11	1.09
Engineering	6.8	1.07	1.20
Arts and Humanities	1.5	0.92	1.13
Physical sciences	27.6	0.78	1.28

Utrecht[12], and which is recognized as the number one top research school in the Bonus Incentives Scheme ("dieptestrategie"). From the table we furthermore see that the citation impact of the universities (almost all of Dutch astronomy and by definition all of NOVA) is 1.30[13]. In the NOWT, however, the impact of Dutch astronomy is quoted as 1.19, which includes the ESA publications. Here another bias is present, namely that astronomical papers in *Nature* and *Science* are collected in a separate category "multidisciplinary" and are not contributing to the astronomy and astrophysics impact.

Referring back to table 3, it is interesting to note that except for astronomy and computer science (and in a very minor way for earth sciences) papers from the universities have, on average, a lower citation impact than those from the OZI's and industry combined. The very high citation impacts of physics, chemistry and materials science do not result from academic research; the FOM institutes and industrial laboratories must produce very highly cited papers. This is important information that is not addressed in the analysis of the NOWT.

With the information in the statistical appendix it is possible to estimate the numerical information for broad groups of disciplines (table 4). Specialisation is the ratio of the percentage of Dutch papers in a discipline relative to that percentage averaged over all sciences. This table clearly confirms the generally held perception that the Netherlands invests (much) more than average in social sciences. Restricting ourselves to the physical sciences, we see that in general the Netherlands appears to be investing below average. In spite of this the citation impacts are very high. The physical sciences are responsible for a substantial fraction of the scientific output of the Netherlands and the conclusion that in citation impact the Netherlands ranks third in the world is for a major part a result of performance in physical sciences. Table 4 is organised in order of decreasing specialisation. The general trend in the final column suggests that more support does not necessarily lead (on average at least) to better research; in fact, there is a definite anti-correlation.

1.5 The Burstein paper

David Burstein[14] has produced an enormous dataset (the *Astronomy Citation Database* or ACD) on citation data of astronomers worldwide. The main list (available through

[12] Nijmegen has recently joined NOVA as the fifth institute.

[13] For the Kapteyn Astronomical Institute the value is 1.37. Leiden has 1.39, Amsterdam 1.19 and Utrecht 1.05. *Note added in 2004*: The values for 1998-2001 are respectively 1.45, 1.35, 1.22 and 1.23, and for the whole of the universities 1.29.

[14] D. Burstein: *Astronomers and the Science Citation Index, 1981- 1997*, Bull. Am. Astron. Soc. **32**, 917-936 (2000).

anonymous ftp upon request from the author) encompasses the citation statistics for about 6500 astronomers, who received more than 100 citations over the period 1981.0 to 1997.5 in the Science Citation Index. Of these about 6500 astronomers, 4617 are "unique-1" (names *and* initials are uniquely identified). In a further 357 cases the name and initials were not unambiguously associated and 1484 cases remained confused with others with similar names. Although this shows a limitation, it remains true that the main unique-1 sample contains the vast majority and can therefore be used for analyses with sufficient confidence.

My critisism to this work is that in spite of the fact that the master list has for each astronomers his number of citations per paper, he presents lists of astronomers with simply the largest total number of citations. That is not the proper way to use these data, even though I must admit that many of these people also have high numbers of citations per paper. It is true that Dave himself is in the list, but he also has a very high citation rate[15]. The database is still of use, since it was attempted to clean up most of the name confusion in the Science Citaton Index.

After the paper became available on astro-ph[16], I had an email exchange with Dave Burstein (we work in closely related research areas, so we know each other well). I tried to make the point that one should not be preoccupied with total citation numbers and how I felt he had to provide means to normalise these. We did not get to any agreement and in the end he noted: *"I never made the claim in my paper that the kind of analysis I did was the be-all and end-all. Rather, it is the kind of analysis one can do with a very large database that many others were likely to do. In doing the analysis I found a number of systematic problems that affect the statistics at the few percent level, thus alerting both the ISI and others who might do this kind of analysis."*

Burstein also presented the details of the distributions. In particular that for the citations per paper averaged over astronomers it is interesting. It makes it possible to answer a question that I always asked myself, namely what does it mean to have a certain impact ratio in terms of where one ranks among researchers? I only look at the unique-1 sample. The median astronomer has 18.3 citations per paper. The modal value is around 15 and the mean is 26.90. Indeed one has to be very careful because the distribution is very skew. The "average" astronomer with 26.9 citations per paper in this dataset actually has a better score than 78.1% of his/her colleagues. Twice average already means a higher score than 92.3%. What we would still like to know is how this compares between disciplines.

Where do I stand myself in the ACD? I am being listed in the category unique-1 with 31 papers and 1462 citations, giving 47.16 citations per paper. Of the 4617 astronomers in that list I rank about number 450 from the highest citation rate, so that 9.7% of these astronomers have a higher citation rate than I do. In the absolute numbers I do less well.The median number of papers cited is 24 and about 38% of the astronomers have published more papers than I have. Not surprising since most of my papers are by myself or with a single collaborator. The median number of citations is 381 and 8.5% of the astronomers have more citations than I do.

Finally I compared these citation numbers with the ADS (see below). I do find 29 publications for this time interval, after leaving out two errata, one book review and a half-page abstract of an invited talk I had given at a meeting of the Royal Astronomical

[15]Henk Struik ended his paper cited above with: *"Finally, a selection effect is apparent in studies like the present one: they tend to be done for populations that score well compared with the average, and by individuals who themselves do rather well on the citation index."* I commented on the draft of this paper that the word "done" be better replaced with "published", but Henk ignored this comment.

[16]Astro-ph is the commonly used preprint server in astronomy, maintained at Cornell University, New York, since April 1992. It also provides preprint servers in physics, mathematics and computer science. The URL is www.lanl.gov/archive/astro-ph.

Society. ADS lists 1446 citations over the same period or 49.86 per paper.

1.6 The NASA Astrophysics Data System

The Astrophysics Data System (ADS) is a NASA-funded project which maintains four bibliographic databases containing at this time (August 2003) more than 3.3 million records: Astronomy and Astrophysics, Instrumentation, Physics and Geophysics, and preprints in Astronomy. The main body of data in the ADS consists of bibliographic records, which are searchable through abstract service query forms and full-text scans of much of the astronomical literature which can be browsed through a browse interface. In addition to its databases, the ADS provides access and pointers to a wealth of external resources, including electronic articles, data catalogs and archives. Currently there are links to over 1.3 million records. The ADS is maintained at the Harvard-Smithsonian Center for Astrophysics in Cambridge, USA with many mirror sites, in particular at the European Southern Observatory headquarters in Garching bei München, Germany[17]. Apart from the provision of information on papers in journals, conference proceedings and observatory publications the availability of scans of the actual publications that can be downloaded in postscript of pdf format is a most useful tool. Almost all the astronomical literature (including the major journals from their beginning, sometimes back into the nineteenth century) can be seen on-line on the screen. Astronomers these days read papers in journals almost exclusively from their computer screens or using their printers. Of course visits to libraries remain necessary for consulting textbooks and conference proceedings.

In addition to this one can select lists of papers by author name, journal, words in the title or abstract, etc. and obtain lists of papers that cite each article.[18] This makes it easy to perform bibliographical analyses at the level of individual researchers. The question is whether it is possible to use ADS for an analysis like that of the CWTS.

The ADS notes that its citation data are incomplete. However, one may think that it is at least internally consistent so that a comparative study would still be possible. Since ADS is so easily available this would be very nice.

I have performed a test of the ADS by repeating the CWTS studies of my publications using the ADS. The easiest manner is to request for each of one's paper a list of all papers in that particular volume of the journal, using the journal query form, sorted by citation count. The number of papers appears at the top of the listing. That gives the total number of citations for each paper and the total for all papers. There is also the option of restricting the range of publication dates of the citing papers. From this I find the average number of citations of papers in that journal volume. In the list I then also locate my own paper and note its citation count. What I prefer myself is to subtract the citations to my paper from the total and calculate the average citation count of the remaining papers. This then can be used to calculate the impact ratio.

I find 65 papers with me as (co-)author in refereed journals with a total of 3450 citations, or 53.08 citations per paper. For the articles in the same journal volumes the mean citation rate is 20.88, so my impact score would be 2.54 over the whole period since my first paper in 1970. These numbers are for citations included in ADS up to August 2003. In the process I find the impact ratios of individual articles. Two are larger than 10, nine between 5 and 10, twentyfour between 2 and 5 and ten between 1 and 2. That means that 69% of my papers had a more than average impact. There are no papers with no citations. The paper

[17] The URLs are adswww.harvard.edu at Harvard and esoads.eso.org at ESO in Europe.

[18] *Note added in 2005*: It is now even possible to perform word-searches of the full text of all scanned articles in the ADS.

Table 5: Annual citation statistics for major astronomical papers.

Year	A&A	MN	ApJ	AJ	ApJS	A&AS	Total
1970	351	153	644	141	21	44	1363
	8035	3822	21410	3449	1525	836	39180
	22.89	24.98	33.25	45.74	72.62	19.00	28.75
1980	793	390	1243	226	64	172	3286
	14788	9052	44401	4092	6403	4026	85133
	18.65	23.21	35.72	18.11	100.05	23.41	25.91
1990	992	536	1606	357	121	150	4127
	18118	13129	46274	10426	5171	3506	100976
	18.26	24.49	28.81	29.20	142.74	23.37	24.47
1995	1233	752	2121	504	138	270	5018
	18183	14305	50335	10657	6058	3419	102957
	14.75	19.02	23.73	21.14	43.90	12.66	20.52
2000	1571	902	2418	526	182	268	5866
	11951	9106	28515	6290	1567	1971	59400
	7.61	10.10	11.79	11.96	8.61	7.35	10.13

in footnote 4 above has been cited only once, namely by Spruit in the paper in footnote 6. However, the 75 other papers in the same volume have been cited together only 28 times.

For the period 1980 to 1995 (the first CWTS study) I find only 28 papers (CWTS had 31). I don't really know how that arises, since I don't have the list the CWTS used and therefore cannot compare the input in detail. These have an average citation rate of 76.77 citations per paper. For the same volumes of the journals I find a citation rate of 28.00, giving an impact ratio of 2.74. Here CWTS had 3.28.

For the period 1991 to 2001 (the second CWTS study) I find 11 papers (CWTS also had 11; since I still have the CWTS list from the verification round I can be sure we have the same papers). These now have an average citation rate of 22.27 citations per paper. For the same volumes of the journals I find a citation rate of 10.38, giving an impact ratio of 2.15. CWTS found 2.64.

The number of citations in ADS is higher then what CWTS has. This may not be surprising, since I did use citations up to August 2003, while the CWTS studies had citation windows similar to the publication windows. When I look at the 11 papers in the period 1991 to 2001, I find in ADS 149 citations over the same time interval. CWTS has 160 (including self-citations, as in my number). The whole list of 149 ADS references contains no more then 4 papers that are not in the four main astronomical journals (ApJ, AJ, MNRAS and A&A). For completeness I note that a close look at the citing papers produces a number of self-citations of about 30, while CWTS had 26. The ADS and CWTS are not very different. I do not understand the differences, but they are small.

Taking the same window 1991-2001 in the ADS I find for the journals involved in the 11 papers mentioned above that the expected number of citations is 64.40. The CWTS must have had an expected 51.5 citations (but excluding self-citations) in order to get an impact ratio of 2.64 with a number of citations of 136. The difference of 20% is probably the result of self-citations in the journal papers.

It is of interest to compare the ADS citation counts to those on the *ISI Web of Science*. On the WoS I find 7 papers under "van der Kruit". There are no papers before 1999. When I spell my name as "vanderkruit" I do get my publications before that (back to 1988). In

Table 6: Citations per paper for *Monthly Notices of the Royal Astronomical Society.*

1960	20.75	1970	24.38	1980	22.59	1990	23.75	2000	8.05
1961	11.91	1971	29.74	1981	28.58	1991	22.34	2001	4.76
1962	22.96	1972	25.77	1982	28.84	1992	23.85		
1963	24.27	1973	22.86	1983	25.15	1993	22.41		
1964	25.47	1974	22.95	1984	24.61	1994	20.32		
1965	22.81	1975	23.26	1985	27.15	1995	17.88		
1966	21.29	1976	24.93	1986	28.35	1996	17.75		
1967	23.82	1977	23.87	1987	25.23	1997	15.30		
1968	27.63	1978	27.23	1988	26.92	1998	13.10		
1969	30.65	1979	23.10	1989	24.82	1999	11.45		

total there are 14 papers in refereed journals. With ADS I find 346 citations to these and in WoS 355. So, ADS *is* very complete compared to the Science Citation Index. Restricting to the period 1991 to 2001 for the publication dates, the WoS gives 245 citations up to now, while ADS gives 240. It would seem that it is possible to use the ADS for bibliographic studies.
Note added August 2005. The new (2005) study of the CWTS was for my citation scores based on 8 publications with 189 citations, corrected for self-citations to 163 citations. One of the publications was an erratum! Excluding that one, I find in ADS 205 citations over the same period, of which 16 are citations by articles on which I was a (co-)author. The CWTS self-citations include those where my co-authors are self-citing. ADS thus gives again *more* citations than the CWTS extracts from the *Citation Index.*

Up till now I only compared the citation of the papers to those of others in the same journal volume. That is maybe a small sample to use for normalisation. Indeed it turns out that variations in the average citations per paper between volumes can be as large as 20% and that contributes then to the "noise" in the impact ratio. From year to year the variation is more like 10%, so that it is preferable to use annual averages.

What about averages over the field? In astronomy (at least my kind of astronomy; for planetary science of solar physics it may be different) this can be very well done by restricting oneselves to the four major journals: *Astrophysical Journal* (ApJ) en *Astronomical Journal* (AJ) in the USA, *Monthly Notices of the Royal Astronomical Society* (MN) in the UK and *Astronomy & Astrophysics* (A&A) in the rest of Europe. ApJ and A&A also have an important Supplement Series (A&A until recently). Table 5 gives some statistics for a few selected years. For each year the top row is number of papers, middle row number of citation up to August 2003 and the lower row citations per paper.

These are more cumbersome to get from ADS because of the large numbers involved. Note that there is a significant difference between the journals. The European journal *Astronomy & Asthophysics* is generally considered of lesser quality than the three others (see above). The citation rate indeed is lower than the others.

Important to note is that the citations are of the level of 25 per paper up till the early nineties. These are citations up to August 2003. Also note that the value for the MNRAS is more or less similar to the worldwide number. It is of interest to find this number over all years. To be more definite I have only included citations up to the end of 2002.

On this basis it seems a reasonable approximation to find for each paper the citations up to the end of 2002 and determine the impact factor with respect to the field by assuming an expected citation rate of 25 per paper for papers published up to about 1990, going to down to 20 up to 1994 and then down to 11 or so in 1999. Of course this cannot be done this way for very long, but is reasonable for the moment.

For myself I then get for 59 papers published up to 1998 with a total of 3208 actual citations and 1408 expected citations, yielding an impact ratio of 2.28.

My conclusion is that the citation data in the ADS is very similar to the Science Citation Index and therefore to the input of CWTS studies. However, I do not really understand why the impact ratios found by CWTS are systematically higher than the ones I find with the ADS. Because it can more easily handle large numbers, CWTS has an advantage in that it can calculate $JCSm$ and $FCSm$ using all publications of the relevant year. CWTS certainly can do what a simple application of the ADS cannot do, namely exclude self-citations in determining $JCSm$ and $FCSm$. Still for many purposes ADS is an easy tool to find reasonably accurate estimates of an individual's impact ratios.

Part 2. Application of ADS to Dutch astronomy (written in August 2004, with additions in August 2005)

2.1 Introduction

I have now also done (in the summer of 2004) an excercise where I looked at the permanent staff members at the Dutch institutes. The aim was to study which parameters could be used for bibliometric studies in connection with what the ADS offers in terms of possibilities and options. So I made up a list of astronomers that at the time were active researchers at institutes in the Netherlands and I will first describe how I made up that list.

2.2 Astronomers in the Netherlands and data extraction from the ADS

I used as a starting point the listing of the staff at the institutes of NOVA from the NOVA Web-site at the Leiden homepage (http://www.strw.leidenuniv.nl/nova/people.html). This is not entirely up-to-date and I next added persons on the basis of the telephone directories of the institutes and information on recent hires that I had from personal information. The institutes also list retired staff members that are still active. I looked up in ADS what their most recent publication was and checked that this was indeed a paper in a refereed journal or an invited review at an international symposium. Generally speaking I felt that this most recent paper should not be earlier then 2002 in order to qualify as "active". I will first describe what I did in the case of retired persons.

For Amsterdam Pik Sin Thé is listed, but his most recent publication is a paper in 2001[19]. As far as I am aware he is no longer working regularly at the institute and this may very well be his last involvement in research. Joop Hovenier is still active.

For Groningen Adriaan Blaauw, Tjeerd van Albada, Stuart Pottasch and Hugo van Woerden are active researchers. However, the NOVA list has also Ulrich Schwarz as retired staff member, whose most recent publication is from 2000[20]. I excluded Ulrich.

[19]It is with D. de Winter as first author and has Pik Sin as number 4 of 8 authors. It concerns a photometric catalogue of southern emission-line stars. I note that also retired staff member Tjin A. Djie is a co-author; he has no more recent publications in ADS. Tjin A. Djie is not listed as retired staff member in the NOVA tables.

[20]There is a Udo Schwarz, a former Ph.D. student from Amsterdam, that is not to be confused with Ulrich in ADS.

In Leiden there is a long list. Adriaan Blaauw is included here as well as in Groningen and correctly so. Luc Braes is listed, but his most recent publication according to ADS is 1976! Butler Burton does not live in the Netherlands or work in Leiden anymore; his most recent publications are in 2002 and concern searches for high-velocity clouds in the Dwingeloo Survey and his name has both Leiden and NRAO as his affiliations. I decided to exclude Butler from the list, since he is not really an active researcher *in the Netherlands*. Arnout van Genderen is still active. Cees van Houten has deceased and his wife Ingrid van Houten-Groeneveld has recently reported observations of asteroids and eclipsing binaries, but has not published research papers in recent years. King Kwee's most recent (co-authored) paper in a refereed journal was 1992. Alex Ollongren is not active as a researcher; his most recent publication stems from 1992. Cees van Schooneveld has not appeared on a paper since 1990, at least in ADS. Jaap Tinbergen had invited reviews in 2003, but no research papers since 1993. I decided not to include Jaap.

For Utrecht, Tony Hearn is listed in the NOVA tables, but his most recent publication was in 1998. The same holds for Max Kuperus (also most recent publication in 1998). Harry van der Laan was co-author of a paper in 2001. I left these persons out as active researchers.

The institutes of ASTRON, JIVE and SRON are no formal part of NOVA, so their staffs are not listed in the NOVA documentation. I took the lists from telephone directories and applied similar criteria. I don't know of any active retired staff from ASTRON, but for SRON I included Johan Bleeker (is still publishing) and Rolf Mewe (although he died during the execution of this excercise).

I routinely checked the years of the most recent publication for all persons on the list when I consulted ADS. The only person for whom I found that the most recently published paper was older then from 2003 was Harvey Butcher. His most recent involvement in a research paper stems from 1999. In spite of this I retained Harvey, since he clearly is important in setting the scientific scene in the Netherlands. The NOVA list still had Rudy Wijnands on it, but as far as I know he now works at M.I.T.

In listing these researchers I did so under the heading of their first employer in case they held multiple affiliations. In the case of Michael Perryman and Conny Aerts this was of course not possible.

In extracting information from the ADS it is important not to confuse names. I encountered two difficulties even after using the first initial. These were Rob Rutten and Richard Strom. With Rob it is of course René Rutten (currently director of the Isaac Newton Group of Telescopes at La Palma); I required that the word "solar" was in the abstract. In the case of Richard it is a Robert Strom, working at the Lunar and Planetary Laboratory in Flagstaff, Arizone and who is involved in planetary research. Here we have to be careful, because Richard has been (and is) involved in radio studies of Jupiter with Imke de Pater. I required the word "radio" to be in the abstract. For Leo van den Horn and Stephan Schlemmer it was necessary to also include physics papers to get a good indication.[21]

In the end I had a list of 98 persons, 10 of which are retired, but still active researchers. For all of these I checked the ADS and extracted the relevant information up to early August 2004. It is listed in three parts in table's 7, 8 and 9 in the appendix to this note. For these tables there are the following general points of explanation to make:

- Name. I have only used the last name, even though in ADS I did use the first initial in case there were any persons with similar names.

[21] It is also possible to use the option "middle initial name search". This provides a list of all combinations that occur with a particular last name and first initial (or last name only) and you can select from this list. It is somewhat more cumbersome, but works well. Also it can be used as a check.

- The category "all publications" simply takes all that ADS has in its database regardless of where the publication appears.

- The category "refereed journals" has only publications in refereed journals. It should be noted, however, even though this sounds good, a simple check shows that at least in my own case a few book reviews in *Astrophysics and Space Science* are included, just as conference contributions when these go to that journal. Also errata are included as separate papers. All these points apply to every individual.

- The extraction of the information was done early August 2004 without any limitation on periods of time. So it always concerns all publications of each individual from the earliest one to August 2004.

For table 7, which concerns simply numbers of publications and citations, the columns are the following:

- n: The total number of publications or papers in the ADS database.

- $cits$: The total number of citations in the ADS database to these publications.

- cpp: The number of citations per paper, calculated as $cits/n$.

- $yr1$: The year of the oldest paper of which the person has been (co-)author.

- ppy: The number of papers per year since the first paper appeared. This is calculated as $n/(2004 - yr1)$.

- cpy: The number of citations per year since the first paper appeared. This is calculated as $cits/(2004 - yr1)$.

In table 8 I use what ADS calls "normalised" papers and citations. ADS itself says the following for normalised counts: *"This option sorts the results list according to the score, normalized to the number of authors in the article. This weighs articles with fewer authors higher."* For normalised citation counts it says: *"This option sorts the results list according to the number of citations that each article has, normalized to the number of authors for the article. It weighs the citation count such that articles with fewer authors are deemed more important."* I can see the possible use of correcting for the number of authors, because people publishing many multi-author papers tend to accrue larger number of papers and citations during their career. The information in table 8 is fairly similar to that in table 7:

- nn: The total normalised number of publications or papers. The ADS gives here integer numbers. If less than 10 I added up the individual values and used the outcome of that to calculate derived properties, even though I have not listed these fractional values.

- $ncits$: The total normalised number of citations in the ADS database.

- $ncpnp$: The number of normalised citations per normalised paper, calculated as $ncits/nn$.

- na: This is the average number of authors on a paper from the person involved, calculated from comparing the number of papers to the normalised total or n/nn.

- $nppy$: The normalised number of papers per year since the first paper appeared. This is calculated as $nn/(2004 - yr1)$.

Table 10: Number of expected citations per paper as a function of the year of the first paper.

1960	24.51	1970	24.36	1980	23.79	1990	19.69	2000	8.45
1961	24.53	1971	24.34	1981	23.74	1991	19.03	2001	6.73
1962	24.81	1972	24.24	1982	23.41	1992	18.93	2002	4.73
1963	24.83	1973	24.26	1983	23.02	1993	17.46	2003	2.85
1964	24.81	1974	24.28	1984	22.76	1994	16.45	2004	1.00
1965	24.74	1975	24.27	1985	22.49	1995	15.42		
1966	24.71	1976	24.20	1986	22.04	1996	14.44		
1967	24.76	1977	24.11	1987	21.46	1997	13.09		
1968	24.74	1978	24.05	1988	20.98	1998	11.70		
1969	24.60	1979	23.85	1989	20.32	1999	10.21		

- *ncpy*: The normalised number of citations per year since the first paper appeared. This is calculated as $ncits/(2004 - yr1)$.

Table 9 then collects the following, all concerning only papers in refereed journals:

- *cpp*: This is the number of citations per paper that was derived in table 7.
- *exp*: This is the expected number of citations per paper. It is based on the method introduced at the CWTS and discussed at length in part I of this document. As shown there we may take the citation scores of papers in the *Monthly Notices of the Royal Astronomical Society* as collecting citation numbers that are typical for papers in all journals in the field. In order to estimate the expected citation rates I first redid table 6 for the summer of 2004. Next I calculated the expected average number of citations under the assumption that a person would publish every year the same number of papers. The result is in table 10. One can see that a young astronomer who published his/her first paper in 1999 can expect to have a mean citation rate of about 10 citations per paper if he/she performs at the average world level. This expected number of citations per paper increases to about 25 with time since the first paper.
- *ir*: This is an impact ratio similar to the one used by the CWTS, but only as a first (but presumaby reasonable) approximation. I will discuss this in more detail below. The *ir* is calculated as *cpp*/*exp*. I have used in the actual calculation more decimals than listed in the previous columns.

I then look at the most cited paper in refereed journals for each person, either as any author on the paper or as first author[22]. The columns contain:

- *cits*: The total number of citations.
- *yr*: The year of publication of the paper.
- *auth*: The position of the person among the author. So x/y means that the person concerned has position x among y co-authors.

[22] If you type the symbol ˆ in front of a person's name you get only those papers for which he/she is first author.

Table 11: Publications and citations in ADS for the permanent staff at the institutes in the Netherlands: means, medians, etc.

Part I

Name	all publications			refereed journals					
	n	*cits*	*cpp*	*n*	*cits*	*cpp*	*yr*1	*ppy*	*cpy*
mean	150	2351	14.8	79	2176	26.2	1978	3.5	91.6
dispersion	108	2173	7.2	58	2040	12.7	14	2.4	70.7
ten-percentile	37	472	7.0	19	423	11.8	1966	1.0	20.6
first quartile	66	1095	9.6	38	1267	17.3	1972	1.7	39.7
median	130	1684	13.8	60	2051	24.7	1980	2.8	76.0
third quartile	213	3253	18.7	102	3203	33.1	1989	4.8	126.4
ninety-percentile	313	5646	24.0	175	5195	43.8	1995	6.8	178.9

Part II

Name	all publications			refereed journals					
	nn	*ncits*	*ncpnp*	*nn*	*ncits*	*ncpnp*	*na*	*nppy*	*ncpy*
mean	54	770	12.9	26	684	24.9	3.4	1.5	27.3
dispersion	40	757	7.3	21	675	14.3	1.3	1.1	22.3
ten-percentile	13	96	5.1	6	87	9.2	2.0	0.4	5.8
first quartile	24	215	7.2	11	180	15.3	2.4	0.6	11.4
median	46	508	11.6	22	420	22.0	3.3	1.0	19.7
third quartile	71	1549	17.0	33	1025	30.8	4.1	1.4	36.0
ninety-percentile	114	1932	25.2	52	1636	49.1	5.3	2.0	56.9

Part III

Name	refereed journals			most often cited					
	cpp	*exp*	*ir*	*cits*	*cits*	*cpy*	*ir*	*yrc*	*auth*
mean	26.2	21.3	1.20	214	145	10.9	5.7	8.3	5.0
dispersion	12.7	3.2	0.57	152	133	10.6	4.9	7.1	7.9
ten-percentile	11.8	15.4	0.58	59	34	2.7	1.6	1	1
first quartile	17.3	20.3	0.86	77	55	4.2	2.8	3	2
median	24.7	23.9	1.11	186	101	6.2	4.4	7	3
third quartile	33.1	24.5	1.45	352	199	12.1	7.4	10	5
ninety-percentile	43.8	24.6	2.08	426	386	22.2	14.3	20	9

2.3 The distribution of the parameters

In table 11 I show some characteristics of the distributions of the various properties that I listed. The lay-out is generally the same as in table 7, 8 and 9, except for the third part (derived from table 9). I show for the most cited paper for the case that the person involved is co-author at any position, only the distribution of the total number of citations *cits*. For the case that this person is first author I show in addition to *cits* the distribution of

- *cpy*: This is the number of citations per year that that particular paper has received.
- *ir*: This is the actual impact ratio of that paper, which is the total number of citations compared to that of the average paper in *Monthly Notices of the Royal Astronomical Society* in the same year. These latter numbers were those which I used to construct table 10.
- *yrc*: This is the number of years of publication of the most cited paper since that of the first paper (*yr*1).
- *auth*: This is the total number of co-authors on the paper concerned.

Before looking at these distributions in more detail I first note that the numbers of "all publications" compared to "refereed papers" indicate that although the first category as expected has significantly more publications, the number of citations does not increase very much. In fact, the ratio of the number of all publications to the number of refereed papers is 1.95 ± 0.39, while that ratio for the corresponding citations is 1.09 ± 0.11. So, on average astronomers have almost the same number of miscellaneous publications as papers in refereed journals; yet this only increases the number of citations one has by about 10%. This seems to make the use of the category "all publications" not very helpful, while of course it also does not improve statistics by much. I will not use this category further.

The most important thing to note is that the r.m.s. values are in most cases comparable to the mean value itself, which indicates that the distributions are intrinsically rather skew. This is well known and the usual situation in case of bibliometric parameters. I therefore have indicated the medians and some percentiles. There almost always are very extended tails to high values. I will treat that separately below when I take the thirty highest ranks for a large number of the parameters. Here I look at the "usual" values, for which I will take the range of values for the persons between the two quartiles.

First look at the number of refereed papers each person is involved in. The number of such papers people have co-authored is usually in the range between 40 and 100, of course depending on the person's age. Exceptionally it is less than 10 or more than 175. The number of citations to these papers is usually between 1000 and 3000, giving a average number of citations of around 20 to 30 per paper. These numbers are however not providing much insight, since they depend very much on the age of the persons and the times when they wrote their most cited papers. Also it is potentially heavily influenced by whether or not one usually works in larger groups or as individual or collaborations between two or three people. So, the next properties were designed to see how these infuences can be corrected for.

I have taken the year of the first publication of a refereed paper as the starting year of each person's career. The median is 1980 and the quartiles 1972 and 1989. However, the full distribution is interesting for the following reason. If collected in periods of 5 years, the usual number of persons is (since 1965) about 10 or 2 per year. This is about the mean rate at which people will go into retirement. With a staff of active astronomers of order 100 and a general length of a career of 40 years or so we expect such a number. However, there are in the distribution two periods of 5 years, namely 1970-1974 and 1985-1990, where the number is more like 20. Individual years that have 3 or more entries are 1967 (4), 1972 (9), 1987 (5), 1988 (6) and 1995 (6). It is not clear to me why this occurs, but it seems a bit more than expected from random fluctuations.

Next we see that usually each persons is involved in 2 to 5 refereed papers per year and "earns" between 40 and 125 citations for each year of the career. So, these numbers take to some extent into account the lengths of a person's career, although an older person is able to "earn" much more citations due to the average length of time that the papers can have been cited.

Next I turn to the matter of the "normalised" counts. This is introduced by the ADS and I was interested in seeing to what extent this is a useful way of treating publications and persons. It simply attributes each paper and its citations in proportion to the total number of co-authors. The usual range of normalised number of papers is now between roughly 10 and 30; this is still a range of about the same factor three as for absolute data, but of course this may reflect the range in years over which persons have been publishing. The normalised number of citations is also rather similar to that of not-normalised citations, namely a factor of about 2 between 15 and 30. The range in normalised citations per normalised paper is similar to unnormalised values, namely roughly between 15 and 30. This is not surprising since these are very similar parameters that differ only in the way each contributing paper is weighted in the calculation of the result. The ratio $cpp/ncpnp$ is 1.14 ± 0.33. The r.m.s. scatter shows that in individual cases the different weighing can make an important difference. In particular it makes a large difference for people that are on one or a few papers that mostly have a large number of co-authors and also are highly cited.

From comparison of absolute and normalised numbers of papers we can estimate that on average a person publishes with has between 2.5 and 4 (co-)authors. Exceptionally it is more than 5. Less than 2 on average is also rare; only 8 persons come in this category (Adriaan Blaauw, Vincent Icke, Piet van der Kruit, Arnoud van Genderen, Bram Achterberg, Bob Sanders, Cees de Jager and Jan Kuijpers).

In terms of normalised counts the usual number of papers per year is between 0.5 and 1.5 and the number of earned citations between 10 and 35 per year.

The "impact ratio" ir is designed to be similar (but less rigorous) to the fundamental parameter the CWTS uses. It is calculated here with two important assumptions in the calculation of the expected number of citations per paper. The first is that the citation rates in the *Monthly Notices of the Royal Astronomical Society* are typical for astronomy and astrophysics. The second is that it is assumed that each person has during his/her career published each year the same number of papers. This assumption is probably only roughly true, but should give a reasonable approximation. For example, my value of 2.16 is in good agreement with the value of 2.28 that I found in Part 1.

The median value is 1.11, which means that the citation rate of the median Dutch astronomer is 11% higher than the world average. This is an excellent performance for Dutch astronomy. Half of Dutch astronomers have an impact ratio between 0.85 and 1.45 and about two-thirds have $ir > 1.0$ or are performing above world average. Only 10 % is lower than 0.6 and a full one-quarter is above 1.45. Nine persons even have impact ratio's above 2.0.

Finally, the most cited paper on which one is first author the number of citations is between 55 and 200 or between 4 and 12 citations per year since the time of publication. The individual impact ratio of these papers (calculated as the total number of citations divided by the average number of citations a paper in Mon. Not. R.A.S. of the same publication year has) ranges roughly between 3 and 7 with a median of 4.4. The papers are published between 3 and 10 years after the first publication with a median of 7 years and have a total of 2 to 5 co-authors (including the first author). In fact, of the persons in the list 16 are sole author of their most cited paper and another 25 have a second author in addition to themselves as first author. Together that is more than a third of the sample.

2.4 Highly cited papers

It is of interest to look at the papers with the highest individual impact ratio (in the column ir in part III of table 11). These would give a nice overview of the most important contributions of currently active astronomers in the Netherlands to the progress in the field. I list the ones in ADS (except articles in the *Annual Review of Astronomy & Astrophysics* series, which are listed as refereed papers in ADS, but really are sollicited reviews) for which the impact ratio exceeds 10.0 and include papers with an active Dutch astronomer as a co-author. These are:

1. Impact ratio **59.3**:
K.K. Kwee and H. van Woerden, *A method for computing accurately the epoch of minimum of an eclipsing variable*, Bull. Astron. Inst. Neth. 12, 327-330 (1956).

2. Impact ratio **24.2**:
A. Blaauw, *On the origin of the O- and B-type stars with high velocities (the "run-away" stars), and some related problems*, Bull. Astron. Inst. Neth. 15, 265-290 (1961).

3. Impact ratio **20.2**:
P.D. Barthel, *Is every quasar beamed?*, Ap. J. 336, 606-611 (1989).

4. Impact ratio **19.1**:
R. Mewe, E.H.B.M. Gronenschild and G.H.J. van den Oord, *Calculated X-radiation from optically thin plasmas. V – Improved atomic data for* $n = 1 - 2$ *lines plus satellites from H- and He-like ions,* $2l - 2l'$ $(n \geq 2)$ *and* $3l - 3l'$ $(n \geq 3)$ *transitions from Fe and Ni lines, and lines from Li to Ne isoelectronic sequences of elements C to Ca*, Astron. Astrophys. Suppl. Ser. 62, 197-254 (1985).

5. Impact ratio **19.1**:
P.A. Shaver, R.X. McGee, L.M. Newton, A.C. Danks and R.S. Pottasch, *The Galactic abundance gradient*, Mon. Not. R.A.S. 204, 53-112 (1983).

6. Impact ratio **18.7**:
G. Boella, R.C. Butler, G.C. Perola, L. Piro, L. Scarsi and J.A.M. Bleeker, *BeppoSAX, the wide band mission for X-ray astronomy*, Astron. Astrophys. Suppl. Ser. 122, 299-307 (1997).

7. Impact ratio **18.3**:
A.G.G.M. Tielens and D. Hollenbach, *Photodissociation regions. I – Basic model*, Ap. J. 291, 722-754 (1985).

8. Impact ratio **17.9**:
T.M. Heckman, L. Armus and G.K. Miley, *On the nature and implications of starburst-driven galactic superwinds*, Ap. J. Suppl. Ser. 74, 833-868 (1990).

9. Impact ratio **16.2**:
T.J. Galama, P.M. Vreeswijk, J. van Paradijs, C. Kouveliotou, T. Augusteijn, H. Bohnhardt, J.P. Brewer, V. Doublier, J.-F. Gonzalez, B. Leibundgut, B. and 39 further co-authors (among which P.J. Groot, R. S. Strom, R.A.M.J. Wijers, J. in 't Zand and J. Heise), *An unusual supernova in the error box of the gamma-ray burst of 25 April 1998*, Nature, 395, 670-672 (1998).

10. Impact ratio **15.9**:
J. van Paradijs, P.J. Groot, T. Galama, C. Kouveliotou, R.G. Strom, J. Telting, R.G.M. Rutten, G.J. Fishman, C.A. Meegan, M. Pettini and 21 further co-authors (among which J. Heise and J. in 't Zand), *Transient optical emission from the error box of the gamma-ray burst of 28 February 1997*, Nature 386, 686-689 (1997).

11. Impact ratio **15.4**:
E. Costa, F. Frontera, J. Heise, M. Feroci, J. in 't Zand, F. Fiore, M.N. Cinti, D. dal Fiume, L. Nicastro, M. Orlandini and 16 further co-authors, *Discovery of an X-ray afterglow associated with the gamma-ray burst of 28 February 1997*, Nature, 387, 783-785 (1997).

12. Impact ratio **15.3**:
K.A. van der Hucht, P.S. Conti, I. Lundstrom and B. Stenholm, *The sixth catalogue of Galactic Wolf-Rayet stars, their past and present*, Space Sci. Rev., vol. 28, no. 3, 1981, p. 227-306 (1981).

13. Impact ratio **14.3**:
T. de Graauw, L.N. Haser, D.A. Beintema, P.R. Roelfsema, H. van Agthoven, L. Barl, O.H. Bauer, H.E.G. Bekenkamp, A.-J.Boonstra, D.R. Boxhoorn and 52 further co-authors (among which K.A. van der Hucht, J.M. van der Hulst, H.G.J.L.M. Lamers, E.A. Valentijn, L.B.F.M. Waters and P.R. Wesselius), *Observing with the ISO Short-Wavelength Spectrometer*, Astron. Astrophys. 315, L49-L54 (1996).

14. Impact ratio **14.2**:
E.F. van Dishoeck and J.H. Black, *The photodissociation and chemistry of interstellar CO*, Ap. J. 334, 771-802 (1988).

15. Impact ratio **13.4**:
L.J. Allamandola, A.G.G.M. Tielens and J.R. Barker, *Interstellar polycyclic aromatic hydrocarbons – The infrared emission bands, the excitation/emission mechanism, and the astrophysical implications*, Ap. J. Suppl. Ser. 71, 733-775 (1989).

16. Impact ratio **13.3**:
H. Butcher and A. Oemler Jr., *The evolution of galaxies in clusters. I – ISIT photometry of Cl 0024+1654 and 3C 295*, Ap. J. 219, 18-30 (1978).

17. Impact ratio **13.0**:
H.R. Butcher and A. Oemler, Jr., *The evolution of galaxies in clusters. V – A study of populations since z approximately equal to 0.5*, Ap. J. 285, 426-438 (1984).

18. Impact ratio **13.0**:
G.B. Field, D.W. Goldsmith and H.J. Habing, *Cosmic-ray heating of the interstellar gas*, Ap. J. 155, L149-L154 (1969).

19. Impact ratio **12.7**:
D. Bhattacharya and E.P.J. van den Heuvel, *Formation and evolution of binary and millisecond radio pulsars*, Phys. Rep., 203, 1-124 (1991).

20. Impact ratio **12.5**:
G.S. Vaiana, G. Fabbiano, R. Giacconi, L. Golub, P. Gorenstein, F.R. Harnden Jr., J.P. Cassinelli, B.M. Haisch, H.M. Johnson, J.L. Linsky and 6 further co-authors (among which R. Mewe and C. Zwaan), *Results from an extensive Einstein stellar survey*, Ap. J. 245, 163-182 (1981).

21. Impact ratio **12.4**:
P.T. de Zeeuw, R. Hoogerwerf, J.H.J. de Bruijne, A.G.A. Brown, and A. Blaauw, *A HIPPARCOS census of the nearby OB associations*, Astron. J. 117, 354-399 (1999).

22. Impact ratio **12.4**:
L.J. Allamandola, A.G.G.M. Tielens and J.R. Barker, *Polycyclic aromatic hydrocarbons and the unidentified infrared emission bands – Auto exhaust along the Milky Way*, Ap. J. 290, L25-L28 (1985).

23. Impact ratio **12.2**:
J.H. Black and E.F. van Dishoeck, *Fluorescent excitation of interstellar* H_2, Ap. J. 322, 412-449 (1987).

24. Impact ratio **12.0**:
A. Dressler, A. Oemler, Augustus Jr., W.J. Couch, I. Smail, R.S. Ellis, A. Barger, H.R. Butcher, B.M. Poggianti and R.M. Sharples, *Evolution since* $z = 0.5$ *of the morphology-density relation for clusters of galaxies*, Ap. J. 490, 577-591 (1997).

25. Impact ratio **11.8**:
H.J. Habing, *The interstellar radiation density between 912 Å and 2400 Å*, Bull. Astron. Inst. Neth. 19, 421-431 (1968).

26. Impact ratio **11.6**:
H.H. Aumann, C.A. Beichman, F.C. Gillett, T. de Jong, J.R. Houck, F.J. Low, G. Neugebauer, R.G. Walker and P.R. Wesselius, *Discovery of a shell around Alpha Lyrae*, Ap. J. 278, L23-L27 (1984).

27. Impact ratio **11.3**:
P.C. van der Kruit and L. Searle, *Surface photometry of edge-on spiral galaxies. I. A model for the three-dimensional distribution of light in galactic disks*, Astron. Astrophys. 95, 105-126 (1981).

28. Impact ratio **11.3**:
R.F. Peletier, R.L. Davies, G.D. Illingworth, L.E. Davis, and M. Cawson, *CCD surface photometry of galaxies with dynamical data. II - UBR photometry of 39 elliptical galaxies*, Astron. J. 100, 1091-1142 (1990).

29. Impact ratio **11.2**:
J.B.G.M. Bloemen, A.W. Strong, H.A. Mayer-Hasselwander, L. Blitz, R.S. Cohen, T.M. Dame, D.A. Grabelsky, P. Thaddeus, W. Hermsen and F. Lebrun, *The radial distribution of Galactic gamma rays. III - The distribution of cosmic rays in the Galaxy and the* CO-H_2 *calibration*, Astron. Astrophys. 154, 25-41 (1986).

30. Impact ratio **11.1**:
G. Neugebauer, H.J. Habing, R. van Duinen, H.H. Aumann, B. Baud, C.A. Beichman, D.A. Beintema, N. Boggess, P.E. Clegg, T. de Jong and 18 further co-authors (among which G.K. Miley and S.R. Pottasch, *The Infrared Astronomical Satellite (IRAS) mission*, Ap. J. 78, L1-L6 (1984).

31. Impact ratio **10.9**:
R.C. Vermeulen and M.H. Cohen, *Superluminal motion statistics and cosmology*, Ap. J. 430, 467-494 (1994).

32. Impact ratio **10.9**:
K.C. Chambers, G.K. Miley and W. van Breugel, W., *Alignment of radio and optical orientations in high-redshift radio galaxies*, Nature 329, 604-606 (1987).

33. Impact ratio **10.7**:
G. Hasinger and M. van der Klis, *Two patterns of correlated X-ray timing and spectral behaviour in low-mass X-ray binaries*, Astron. Astrophys. 225, 79-96 (1989).

34. Impact ratio **10.7**:
R. Mewe, J.N. Lemen and G.H.J. van den Oord, *Calculated X-radiation from optically thin plasmas. VI - Improved calculations for continuum emission and approximation formulae for nonrelativistic average Gaunt factors*, Astron. Astrophys. Suppl. 65, 511-536 (1986).

35. Impact ratio **10.6**:
W. Rice, C.J. Lonsdale, B.T. Soifer, G. Neugebauer, E.L. Koplan, L.A. Lloyd, T. de Jong and H.J. Habing, *A catalog of IRAS observations of large optical galaxies*, Ap. J. Suppl. Ser. 68, 91-127 (1988).

36. Impact ratio **10.2**:
R.A.M.J. Wijers , M.J. Rees and P. Meszaros, *Shocked by GRB 970228: the afterglow of a cosmological fireball*, Mon. Not. R.A.S. 288, L51-L56 (1997).

37. Impact ratio **10.1**:
S. Rappaport, F. Verbunt and P.C. Joss, *A new technique for calculations of binary stellar evolution, with application to magnetic braking*, Ap. J. 275, 713-731 (1983).

38. Impact ratio **10.0**:
W.E.C.J. van der Veen and H.J. Habing, *The IRAS two-colour diagram as a tool for studying late stages of stellar evolution*, Astron. Astrophys. 194, 125-134 (1988).

In compiling this list I have ignored papers that cannot really be seen as part of a specific Dutch contribution to astronomy. E.g. there is a paper by C. Stoughton et al. with almost 200 co-authors, among which Amina Helmi, entitled *"Sloan Digital Sky Survey: Early data release"* (2002), which has the very high impact ratio 34.2. This survey is not a Dutch undertaking; furthermore Amina's affiliation in the paper is the Max Planck Institut in Munich. A paper with impact ratio 19.1 by M.A.C. Perryman et al. (among which Rudolf le Poole of about 20 co-authors),: *"The HIPPARCOS Catalogue"* (1997) is also ignored. It concerns a HIPPARCOS satellite data product; not only the satellite itself, but also the production of this cataloque was an ESA project and not one of Dutch astronomy. It is true that the Netherlands played a major part in this satellite project (many others, as Adriaan Blaauw and Tim de Zeeuw were involved in addition to Rudolf, but are not on the paper). Finally, the paper by C.A. Christian et al. *"Video camera/CCD standard stars (KPNO video camera/CCD standards consortium)"* (1985) has Harvey Butcher as a co-author. It has an impact ratio of 10.1, but was written well before Harvey came to the Netherlands; in itself that is OK (as for numbers 16 and 17), but this one concerned an instrument on Kitt Peak Observatory in the USA.

There are some surprises in this list. First the highest ranked paper on the computation of properties of an eclipsing binary by King Kwee and Hugo van Woerden is (at least to me) a real surprise. It has no less than 393 citations, of which still 34 in 2003. It is a (short) paper that remains a source of citation up to the present. For 1956, *Monthly Notices* has on average 6.64 citations per paper (930 citations to 140 papers) and the years before and after give similar numbers. The famous Kwee, Muller and Westerhout paper in the *Bulletin of the Astronomical Institutes in the Netherlands* of 1954 on *"The rotation of the inner parts of the Galactic System"*, has only 45 citations. The other 31 papers in the B.A.N. of 1956 have a total of 515 citations for an average of 16.61 per paper. Interestingly this is much higher than *Monthly Notices.* This note by Kwee and van Woerden really is an all-time champion.

In the second place I note that a number of these highly cited papers describe a space mission, an instrument on a satellite or on a ground-based facilitity, a survey or a catalogue. There is of course nothing wrong with that and it shows that being involved with such a project does not automatically remove the chance of obtaining a good citation score.

The rest of this section has been added in 2005:
In the mean time an interesting article has appeared by Frazer Pearce[23], in which he also studies the use of the ADS for bibliometric studies. Among others he determines distributions of citations of astronomy and astrophysics papers and numbers of citations in a recent 5-year period for individual researchers. In addition he lists the 1000 most cited refereed papers[24] and the 500 most cited non-refereed publications[25] in the ADS. This is interesting to compare with the list just presented.

In Pearce's list of 1000 most cited papers there are quite a few that are theoretical cosmology (and published in physics journals such as the *Physical Review*) and solar system and planetary science publications. E.g. the highest cited paper (with 3127 citations up to December 2004) is a paper by J.P. Perdew and A. Zunger (respectively from New Orleans and Boulder in the USA) entitled *"Self-interaction correction to density-functional approximations for many-electron systems"* and published in *Physical Review B (Condensed Matter)* in 1981. This is not astronomy or astrophsics in my view, but rather related (and undoubtedly relevant) physics. But the second and third, respectively by Shakura and Sunyaev on *"Black holes in binary systems. Observational appearance"* in 1973 (2684 citations) and by Kurucz on *"Model atmospheres for G, F, A, B, and O stars"* from 1979 (2670 citations) are indeed famous papers in astronomy and astrophysics.

Looking through Pearce's list of 1000 most cited refereed papers I find back all papers in my list of papers above with the highest impact ratios. There are five additional ones:

- Rank **359**:
  A. Blaauw, *The O Associations in the Solar Neighborhood*, Ann. Rev. Astron. Astrophys. 2, 213-246 (1964).

- Rank **531**:
  G. Miley, *The structure of extended extragalactic radio sources*, Ann. Rev. Astron. Astrophys 18, 165-218 (1980).

- Rank **887**:
  E.F. van Dishoeck and J.H. Black, *Comprehensive models of diffuse interstellar clouds - Physical conditions and molecular abundances*, Ap. J. Suppl. 62, 109-145 (1986).

- Rank **897**:
  F. Verbunt and C. Zwaan, *Magnetic braking in low-mass X-ray binaries*, Astron. Astrophys. 100, L7-L9 (1981).

- Rank **947**:
  T.S. van Albada, *Dissipationless galaxy formation and the R to the 1/4-power law*, Mon. Not. R.A.S. 201, 939-955 (1982).

The papers in the *Annual Review of Astronomy and Astrophysics* by George Miley and Adriaan Blaauw had not been included in my list with high impact ratios, since these are sollicited reviews rather then refereed papers. The review by Adriaan Blaauw has an impact ratio of 17.8, but compared to the other papers in the *Annual Review* of 1964 it is 9.2. It is by far the most cited contribution of that year. For the paper by George Miley the impact ratio is 17.1, but compared to the *Annual Review* of 1980 it is 2.2. There actually is a contribution by Popper on *"Stellar masses"* in the same volume (1980) that has more than twice as many citations as George's.

[23] F.R. Pearce, *Citation measures and impact within astronomy*, Astronomy & Geophysics **45**, 2.15-2.17 (2005).

[24] See http://www.nottingham.ac.uk/ ppzfrp/top1000.html.

[25] See http://www.nottingham.ac.uk/ ppzfrp/nonref.html.

The two lists together should comprise a major fraction of the most important contributions of Dutch astronomy. In order of total number of citations we have:

First author	NL author(s)	year	*ir* 2004		Pearce		
			rank	*ir*	rank	*cits*	NL
Mewe	Mewe	1985	4	19.1	174	620	1
Barthel	Barthel	1989	3	20.2	181	605	2
Tielens	Tielens	1985	7	18.3	192	589	3
Shaver	Pottasch	1983	5	19.1	224	549	4
Heckman	Miley	1990	8	17.9	253	523	5
van der Hucht	van der Hucht	1981	12	15.3	308	478	6
van Dishoeck	van Dishoeck	1988	14	14.2	343	457	7
Blaauw	Blaauw	1964		17.8	359	446	8
Field	Habing	1969	18	13.0	383	433	9
Boela	Bleeker	1997	6	18.7	415	418	10
Allamandola	Tielens	1989	15	13.4	441	408	11
Kwee	van Woerden	1956	1	59.3	447	405	12
Butcher	Butcher	1978	16	13.3	467	399	13
Vaiana	Mewe	1981	20	12.5	493	392	14
Allamandola	Tielens	1985	22	12.4	507	388	15
Bhattacharya	van den Heuvel	1991	19	12.7	511	385	16
Butcher	Butcher	1984	17	13.0	528	380	17
Miley	Miley	1980		17.1	531	379	18
van der Kruit	van der Kruit	1981	27	11.3	571	365	19
de Graauw	de Graauw et al.	1996	13	14.3	573	364	20
Habing	Habing	1968	25	11.8	577	363	21
Black	van Dishoeck	1987	23	12.2	580	362	22
van Paradijs	Groot et al.	1997	10	15.9	590	359	23
Galama	Groot et al.	1998	9	16.2	599	357	24
Costa	Heise/in 't Zand	1997	11	15.4	638	346	25
Rice	de Jong/Habing	1988	35	10.6	667	338	26
Aumann	de Jong/Wesselius	1984	26	11.6	710	330	27
Hasinger	van der Klis	1989	33	10.7	747	322	28
Bloemen	Bloemen/Hermsen	1986	29	11.2	773	317	29
Mewe	Mewe	1986	34	10.7	781	316	30
van der Veen	Habing	1988	38	10.0	795	313	31
Neugebauer	Habing et al.	1984	30	11.1	801	312	32
Blaauw	Blaauw	1961	2	24.2	812	311	33
van Dishoeck	van Dishoeck	1986		9.9	887	295	34
Verbunt	Verbunt	1981		9.9	897	294	35
Rappaport	Verbunt	1983	37	10.1	933	289	36
van Albada	van Albada	1982		9.8	947	286	37
Peletier	Peletier	1990	28	11.3	956	284	38
Chambers	Miley	1987	32	10.9	968	283	39
Dressler	Butcher	1997	24	12.0	963	283	40
de Zeeuw	de Zeeuw/Blaauw	1999	21	12.4		257	
Wijers	Wijers	1997	36	10.2		253	
Vermeulen	Vermeulen	1994	31	10.9		209	

Pearce's list of most cited *non-refereed* publications is based on citations up to April 2005. It is led by C.W. Allen's famous book *"Astrophysical Quantities"*, which has had 4353 citations up to April 2005 in the ADS. However with a different code it appears again at rank 29 with 1202 citations. The former is the 1973 edition and the latter one the edition of 1963. This indicates (see also below) that one has to be carefull using the ADS for such compilations, when for some publications various editions and reprints are available.

In second place is *"Numerical Recipes in FORTRAN. The Art of Scientific Computing"* by William H. Press et al. with 3865 citations. The other places down to rank ten has books that probably most astronomers know well and own, in particular workers in the Galactic and extragalactic field. It includes two editions of the *"Reference Catalogue of Bright Galaxies"*, and books as *"Galactic Dynamics"* by James Binney and Scott Tremaine, *"The Large-scale Structure of the Universe"* by Jim Peebles, *"Gravitation"* by Charles Misner, Kip Thorn and John Wheeeler, *"Physical Processes in the Interstellar Medium"* by Lyman Spitzer and *"Astrophysics of Gaseous Nebulae and Active Galactic Nuclei"* by Donald Osterbrock. All these have more than 2000 citations.

At rank 10 we find *"Table of Integrals, Series and Products,* by I.S. Gradhsteyn and I.W. Ryzhik with 1991 citations. It appears again at rank 14 with 1732 citations; and also at rank 81 with 544 citations. Adding these brings it to a total of 4267 which is good for second place. These are different editions, respectively from 1980, 1965 and 1994. The *"Handbook of Mathematical Functions with Formulas, Graphs and Tables"* by Milton Abramowitz and Irene A. Stegun is at rank 13 with 1846 citations, but again it appears at rank 25 with 1335 additional citations, at rank 42 with 899 citations and rank 224 with 241 citations. These are editions from 1965, 1972, 1970 and 1968 respectively.

Allan Sandage's *"The Hubble Atlas of Galaxies"* appears at rank 41 with 961 citations and Chip Arp's *"Atlas of Peculiar Galaxies"* at rank 108 with 460 citations.

The "Dutch" contributions to the list of non-refereed papers (now including all persons that worked extensively in the Netherlands) are:

- Rank **54** with 729 citations:
  C.E. Moore, M.G.J. Minnaert and J. Houtgast, *The Solar Spectrum from 2935Å to 8770Å* (1966).

- Rank **68** with 615 citations and at rank 491 with 117 citations (total 732 citations, which would have been rank 53):
  H.C. van de Hulst, *Light Scattering by Small Particles* (1957). The first one is the 1957 edition and the second the 1981 Dover reprint.

- Rank **93** with 503 citations:
  S.R. Pottasch, *Planetary Nebulae - A Study of Late Stages of Stellar Evolution* (1984).

- Rank **166** with 310 citations:
  A. Lauberts and E.A. Valentijn, *The Surface Photometry Catalogue of the ESO-Uppsala Galaxies* (1989).

- Rank **284** with 192 citations:
  K.G. Begeman, *HI Rotation Curves of Spiral Galaxies*, Ph.D. thesis at the University of Groningen (1987).

- Rank **296** with 188 citations:
  A.H. Broeils, *Dark and Visible Matter in Spiral Galaxies*, Ph.D. thesis at the University of Groningen (1992).

- Rank **348** with 160 citations:
  M. Minnaert, J. Houtgast and G.F.W. Mulders, *Photometric Atlas of the Solar Spectrum from λ3612 to λ8771 with an Appendix from λ3332 to λ3637* (1940).

- Rank **358** with 155 citations:
  W.H.G. Lewin, J. van Paradijs en E.P.J. van den Heuvel (editors), *X-ray Binaries*, Cambridge Astrophysics Series (1995).

- Rank **456** with 125 citations:
  H.J.G.L.M. Lamers and J.P. Cassinelli, *Introduction to Stellar Winds* (1999).

- Rank **466** with 123 citations:
  H.C. van de Hulst, *Multiple Light Scattering. Tables, Formulas and Applications* (1980).

- Rank **469** with 122 citations:
  A. Blaauw, *OB Associations and the Fossil Record of Star Formation*, contribution to a NATO Advanced Science Institute on "The Physics of Star Formation and Early Stellar Evolution" (1991).

- Rank **494** with 116 citations:
  W.B. Burton, *The Structure of our Galaxy derived from Observations of Neutral Hydrogen*, review in: Galactic and extragalactic radio astronomy (2nd edition), 295-358 (1988).

Obviously we have some very fundamental contributions to research here, either in the form of books, other separate publications or contributions to symposia and summer schools, etc. Contributions to *Annual Reviews* should have been included here rather then under the refereed papers. The two Groningen theses were unusually frequently cited, probably because they were much later and only partly published in refereed journals.

Finally, I comment on the fact that in Pearce's two lists not a single publication appears by Jan Oort. ADS shows his most cited publication as 275 (as of August 2005) for his 1977 paper in the *Annual Review of Astronomy and Astrophysics* on *"the Galactic Center"*. His review from 1983 on *"Superclusters"* has 189 citations. His most cited research paper in a refereed journal is *"The formation of galaxies and the origin of the high-velocity hydrogen"* in Astron. Astrophys. 7, 381-404 (1970). His famous paper on the "Oort Cloud" (*"The structure of the cloud of comets surrounding the Solar System and a hypothesis concerning its origin"*) from 1950 and published in the *Bulletin of the Astronomical Institutes of the Netherlands* (B.A.N.) has 198 citations and that constitutes an impact ratio of 39.9[26].

2.5 The use of the various parameters

The main reason for executing this excercise was to see which parameters are useful as instruments to judge the impact of scientific work.

In table 12 (see the Appendix) I have ordered the persons according to various parameters and list the thirty highest scores. Tables 12.A and 12.B list straightforwardly the number of papers each person has been involved in and the number of citations. This depends of course very much on the length of one's career and one likely scores higher here if one usually works in a large group and publishes with more co-authors. On the other

[26] An interesting observation is that the 139 papers in the 1950 *Monthly Notices of the R.A.S.* have on average only 5.96 citations in August of 2005. But a full 97 of these papers have no citations in ADS at all!

hand, being involved with many papers and collecting many citations to these constitutes genuinely a high-profile contribution to the progress of science. It is certainly an accomplishment, especially for those that are leaders of a productive group. Making these lists, I would say, is a useful excercise if one wants to evaluate where the strong contributions are, but it cannot be the full answer.

Table 12.C lists the highest scores in citations per paper and that is a first indication of what we may approximate with attributes such as quality, usefulness or visibility. It is easily seen that a good score here involves a long career; except for two cases the year of the first paper is earlier than 1990 and in 20 of the 30 cases earlier than 1980.

Table 12.D normalises the number of papers to the average over the years since the first publication. Here it would be expected that persons usually involved in publishing multi-author papers would have an advantage. Indeed the first fifteen have the parameter na from part II on average 4.5 ± 1.3 and a median of 4.1, while table 11 shows that the mean for the full sample is 3.4 ± 1.3 and the median 3.3.

Table 12.E has the number of citations normalised to those per year since the first year of publication. Here the effect of the number of authors is less pronounced. The average is now 4.2 ± 1.4 and the median 3.9 for the first 15 persons in the table.

I now turn to the normalised papers and citations. This concept should help to minimise the effects of multi-authored papers on a person's scores. It has been used before, of course, sometimes with the feature of giving the first author some extra weight, either half of the full attribution of the paper or citations or twice that which the other co-authors individually get[27]. The ADS decided to simply give all authors the same weight, independent of their position in the order of co-authors. It would allow a "correction", possibly removing some of the biases for persons with many multi-authored papers.[28] I am not convinced that the full correction is justified.

In table 12.F we see the highest scores with normalised number of papers. There is some difference, but note for example that of the first 9 persons in tables 12.A and 12.F (absolute and normalised number of papers) 5 are in both tables. The same happens with the scores for the normalised number of citations (table 12.G); for that property the comparison is to table 12.B and for the first 10 in those two tables no less than 7 are the same. In a way this is comforting, since although it seems that the normalisation is a fair way of allowing for the effect of multi-authorship, it does on the other hand not change things in a substantial way.

The normalised number of citations per normalised paper is the next parameter. The thirty highest scores are in table 12.H. I already noted that on average this property does not differ much from its absolute value (the average ratio was 1.14 ± 0.33). However the r.m.s. value shows that the scatter is larger than the difference from 1.0 and therefore in individual cases the effect might be important. Indeed we see when we compare tables 12.C and 12.H that generally the same names occur, but some have changed quite significantly in the order. For example Jan Willem Pel is now well on top with a value of 77.9, while he

[27] A thing not known much in astronomy is a practise that I encountered in my function as dean of the Faculty or when serving on review committees within the Faculty, is the importance of being *last* author. Particularly in chemistry and biology multi-authored papers usually have the person who did most of the actual work –often a graduate student or a postdoc– as first author and the leader of the group –both in the sense of director of the research and first applicant of the grant– as last author. In any case in reviewing persons in these disciplines the number of times one has been *last* author is an important measure of one's standing. This is, as far as I am aware, completely unknown in astronomy and astrophysics.

[28] In my own case, where the average number of authors is na=1.5 (the median is 3.3), it turns my relatively low production of ppy=2.2 papers per year compared to a median of 2.8 into a relatively high production of $nppy$=1.5 normalised papers per year compared to a median of 1.0.

Table 13: Persons at least 8 times in table 12.

Name	total	1-15	16-30	Name	total	1-15	16-30
Tielens	13	13	0	Franx	10	7	3
Miley	13	12	1	Pottasch	10	7	3
van Dishoeck	12	10	2	van der Hulst	10	3	7
Habing	11	11	0	Waters	9	7	2
van der Kruit	11	7	4	Butcher	9	5	4
Lamers	11	7	4	Mewe	9	4	5
de Jong	11	6	5	van den Heuvel	8	5	3
de Zeeuw	11	2	9	Sanders	8	2	6
van der Klis	10	8	2	Kuijken	8	1	7
Verbunt	10	9	1	Langer	8	1	7

was number 15 in the absolute value (39.7). Also Peter Barthel has made a difference; he now is at number 14 with 39.1 citations per paper, while in table 12.C he did not appear (his value of 27.4 citations per paper is lower than the 30.1 for the cut-off in table 12.C).

Table 12.I gives the thirty highest scores in normalised number of papers per year since the first publication. The mean vale of na for the first 15 persons is 3.1 ± 0.8 and the median 3.2. This is much closer to those numbers for the full sample (3.4 ± 1.3 and 3.3). Indeed the largest part of the bias towards persons often on multi-authored papers is removed. It remains of course true that values of relatively high na still apply to the first 5 persons in the table.

The same thing is happening in table 12.J, which lists the number of normalised citations per year since the first year of publication. The mean na of the first 15 now is 3.1 ± 0.9 and the median 2.9. But again na is on the higher side for the first 5 persons on the list. On the other hand, Xander Tielens, George Miley, Marijn Franx, Michiel van der Klis and Ewine van Dishoeck would be on everybody's list, I suspect, of top astronomers in the Netherlands.

In table 12.K I list the highest scores in impact ratio. Although I repeat that this is not done as thoroughly and rigorously as the CWTS, this is similar to the impact ratio's the CWTS calculates with respect to the field. Note that the high scores of Frank Helmich, Rudolf le Poole and Peter Roelfsema result from high citation rates for papers with many authors, on which they are not the first author. This is visible in table 12.K by comparing the most often cited papers on which they appear with that on which they are first author. As Scott Trager and Joop Schaye show, one does not need to have a long career to score high in this list. Also a relatively small number of papers per year, as Jan Willem Pel and Jan Lub have, is not a disadvantage; what they published deserved clearly to be cited often.

Tables 12.L and 12.M list the highest number of citations of the person's most frequently cited papers.

2.6 Concluding remarks

Taking everything together I conclude that the ADS is useful for bibliometric studies and in particular can be used to define various parameters describing the citation scores of individuals and groups of individuals. It does in any case take multidisciplinary journals much better into account than the CWTS. As this second part has shown, it is possible to take various approaches to this and it is not entirely obvious what the best parameter is.

Table 15: Persons at least 4 times in table 12, parts C, D, E, H, I, J and K.

Name	total	1-15	16-30	Name	total	1-15	16-30
Tielens	7	7	0	Lamers	5	3	2
Miley	7	6	1	van der Hulst	5	3	2
Franx	7	5	2	Koopmans	5	2	3
van Dishoeck	6	4	2	Schaye	5	2	3
van der Kruit	6	4	2	van der Klis	4	4	0
Kuijken	6	2	4	Waters	4	4	0
de Zeeuw	6	2	4	Fender	4	3	1
Langer	6	1	5	Falcke	4	2	2
Peletier	6	0	6	Pottasch	4	2	2
Habing	5	5	0	Jaffe	4	1	3
de Jong	5	4	1	Sanders	4	0	4
Verbunt	5	4	1	Wijers	4	0	4
Butcher	5	3	2				

There certainly is merit in correcting for the length of the career and for multiple authorship on papers. But it is not always fair to fully correct for these effects.

To conclude this report I take the point of view that any of the 13 parameters in table 12 have some merit for listing persons and then look which persons occur most often in the various individual parts of table 12. Table 13 shows the result for all persons that occur at least in 8 of the 13 parts of table 12. This also includes all persons, as it happens, that occur at least 5 times in the top half (the highest 15) in those tables. The full list of all persons in the sample appears as table 14 in the appendix.

It is maybe not remarkable that almost all these persons all have a relatively long history of publications. The fact that a number of the tables disadvantage younger persons is of course the reason for this. When we look in the table in detail we see that we have here persons that come from all of the major academic institutes and the two NWO foundations. Many of then have had or still have major administrative duties and have devoted much of the time that they could otherwise have productively used for their research instead for the benefit of their own organisations, astronomy in the Netherlands or for international bodies. Xander Tielens is the obvious exception, although a few others are not known for a heavy involvement in administration. For those with extensive involvement with administration, to some extent of course their scientific status has been a factor in them being asked for these functions. On the other hand, they are still active researchers in the sense that they regularly publish research papers.

I have been surprised by Ed van den Heuvel's rank in these orderings. He did not come in among the first thirty in the values for citations per paper (table 12.C), number of papers and citations per year (tables 12.D and 12.E), number of normalised citations per normalised paper (table 12.H) and impact ratio (table 12.K). It is true that his value for na is somewhat low (2.3), certainly lower than his colleagues in Amsterdam. In any case the "corrections" involving normalising of papers and citations does help restore him to a large extent to the prominent position he deserves. He still does not make it in citations per paper, but in normalised papers per year and normalised citations per year he now comes in as number 16 and 19. Ed is an astronomer with a very high standing, so this shows we have to be carefull with citation statistics to measure standing and quality.

It might be of interest to find out in table 12 for which parameters persons in table 13 have not come in among the thirty highest. In particular I look at the citations rates in tables 12.C (citations per paper), 12.H (normalised citations per normalised paper) and 12.K (impact ratio). Frank Verbunt and Henny Lamers miss tables 12.C and 12.K, but they do get in for table 12.H at 23 and 27. Interestingly Michiel van der Klis, Stuart Pottasch, Rens Waters and Ed van den Heuvel miss all three of these listings. None of the other persons in table 12 miss tabels 12.C and 12.H, and 12.K is only missed by Rolf Mewe.

This brings up again the point that I have repeatedly made throughout the years and that is the disadvantage astronomers might have as a result of the fact that their papers in *Nature* are not taken into account by the CWTS studies. The question then is whether or not this is the reason for the somewhat lower score on citation rates by Amsterdam astronomers. In ADS of course papers in *Nature*, but also in *Science*, are treated in a way fully equivalent to papers in dedicated astronomical journals. Indeed in the list above of papers with a high impact ratio some Amsterdam *Nature* papers feature highly. If we look at the citation lists for Michiel van der Klis, Ed van den Heuvel, Ralph Wijers, Jan van Paradijs and Titus Galama, we find that among their 10 most highly cited papers respectively 3, 3, 3, 6 and 5 are in *Nature*. This seems to be consistent with the notion that the citation numbers in ADS are as complete for papers in these journals as they are for other journals. I find it difficult to accept that in the case of the use of ADS publishing in multi-disciplinary journals would lower one's citation scores. The point does apply for CWTS studies, since there the papers in *Nature* are either not counted in astronomy and astrophysics, or are normalised by a wrong expected number of citations. But it does not hold here, since we normalise to citations in typical *astronomical* journals.

With these qualifications I conclude that apparently, the differences in the Netherlands are small between the institutions and a major workload in administration and/or management does not seem to have a large adverse effect on ones standing when it is measured in terms of publications and citations.

I remarked that the derivation of table 13 disadvantages younger persons. I therefore also did that exercise using only those parts of table 12, where this disadvantage is not present or at least not strong. I used those parts that involve (normalised) citations per (normalised) paper (parts C. and H.), (normalised) papers or citations per year (parts D., E., I. and J.) and impact ratio (part K.). The result for persons with at least 4 entries in these parts of table 12 are listed in table 15. The full list is in the appendix as table 16. Indeed some younger staff members now appear in this listing, such as in particular Leon Koopmans, Joop Schaye, Rob Fender and Heino Falcke.

Finally, I make a comparison in terms of impact ratios between the institutions. The Netherlands Observatory for Science and Technology (NOWT; see above and note that in the mean time a 2003 version has appeared) has produced these with the CWTS method. There are of course the fundamental differences that the NOWT uses a window for publications and citations of only a few recent years and that the data collected here concern the full publishing career of the currently active staff members. Also the NOWT concerns all publications from the institutions and not only those with permanent staff members on it.

I show the result in table 17. For each institute I have summed up the actual total number of citations (*cits* in table 7) of all staff members together and then the expected number (*exp* from in Table 9 multiplied by the number of papers n in table 7). This then gives the impact ratio ir. However, this is weighted by the total number of papers each staff member has, so we may also –as an alternative– simply average the impact ratios of the staff members and this gives $\overline{ir}$. The final two columns give the results from the two recent

Table 17: Impact ratios per institute compared to the Netherlands Observatory for Science and Technology.

	ADS				NOWT	
	cits	*exp*	*ir*	$\overline{ir}$	2000	2003
Amsterdam	33279	31268	1.06	0.96	1.19	1.22
ASTRON + JIVE	19400	19789	0.98	0.99		1.66
Groningen	43939	29705	1.48	1.31	1.37	1.43
Leiden	54666	40805	1.34	1.32	1.39	1.35
Nijmegen	4100	4509	0.91	1.06		
SRON	31794	31908	1.14	1.31		1.04
Utrecht	20493	19415	1.06	1.00	1.05	1.23
Netherlands	207671	173399	1.20	1.20	1.30	1.29

versions of the NOWT. For SRON and ASTRON technical research published in technical papers is likely to be part of the papers used in the NOWT and the result is probably not correct. Indeed, in these two cases the difference is considerable. In general, for the university institutes the results are comparable. Comparing the NOWT values between the institutes it appears also that the effect of multi-disciplinary journals on the Amsterdam impact ratios is not as important as I suspected Fbefore.

My overall conclusion is that the ADS can be used for bibliometric studies. I conclude that the numbers of citations collected by articles in *Nature* do not appear significantly different from those in other journals, so that the average citation impact of a group is probably not systematically lowered by not taking those into account. I also note that the use of "normalised" publication and citation counts is a useful way of allowing for the effects of different publication habits in terms of number of co-authors. However, I am not convinced of any clear and unambiguous measure of an individual's stature in the field. The various measures that I used in table 13 might be the best compromise, although the length of one's career is an important factor in some of the measures. I stress that about two-thirds of tenured astronomers at Dutch institutes have citation rates (impact ratio's) above the world average.

A Appendix

The tables in this appendix show first the publication and citation scores of the permanent staff in the institutes in the Netherlands (Tables 9, 10 and 11). Then these data are sorted for the thirty highest scores in a number of ways (Table 12.A-M), and finally there are listings (Tables 14 and 16) of the frequence of occurence for all persons in the sample in Table 12, corresponding to the listings in tables 13 and 15 in the main text.

Note added October 2005: **I will distribute this document in principle without the appendix, since the information contained in it is really out of date by now. Anyone who is interested in his/her position in any listing can go to ADS and find his/her scores (remember I did this in August 2004!) and compare to the distributions that I list in Table 11.**