Report on the Citation Analysis of the Netherlands School for Astronomy (NOVA)

dr. Hans J.T. Buist and prof. dr. Pieter C. van der Kruit

August 29, 2016

Abstract

This report aims to give a concise overview of the performance of Dutch astronomy performed at the 4 NOVA institutes at the Universities of Amsterdam (Anton Pannekoek Institute), Groningen (Kapteyn Astronomical Institute), Leiden (Sterrewacht Leiden), and Nijmegen (Department of Astrophysics, Radboud University). The citation statistics are obtained from NASA's ADS publication record service. The study consists of two parts: we investigate the refereed articles from the institutes from 2010–2015, and we investigate all refereed articles from staff members active in the 2010–2015 period and compare the citation metrics with those from staff members of top institutes in the United States and Europe.

1 Introduction

In this study we aim to provide numbers that quantify the impact of refereed articles produced by the NOVA Federation. This study was initiated by NOVA as a continuation of an earlier study, Kamphuis & van der Kruit (2010).

We followed two separate procedures, one where we looked at all refereed articles from the NOVA institutes in the years 2010–2015, and one where we tracked publications from the full career of staff members of the NOVA institutes that were active in the 2010–2015 period. For both studies we only used refereed articles, and no other publication forms.

In the first study, the goal is to show how well publications from NOVA institutes perform. To this end a comparison was made with publications in refereed astronomy journals, but also to lists of revelant and impactful journals (these lists will be specified later on). Several different statistics for these publications are given, including the degree to which the NOVA institutes collaborate on articles.

In the second study, we investigate how well the tenured NOVA staff performs. In that case we compare them to staff members from top institutes in the United States and Europe. These top institutes are listed in Appendix A, and are based mostly on bibliographic measures from a study by Kinney (2008), but also for a part on their reputation in the field.

We use NASA's Astrophysics Data System (ADS) as the base system for gathering publication records and citation statistics. ADS is the main database for astronomical articles, from both peer-reviewed journals and non-reviewed sources

such as arXiv.org. We therefore consider ADS to be highly appropriate to investigate the impact of astronomy articles. Articles on ADS are uniquely identified by their bibcode. ADS keeps track of any possible changes of this code, such that articles can still be found by their old bibcode. This may for example occur when the publication year of an article was changed, or because it moved from arXiv.org to a refereed journal. The latest development for ADS is called 'Bumblebee', which provides an application programming interface (API) that allows scripts to execute queries on the database. We used this interface to search on ADS using a Python module by Vladimir Sudilovsky and Andy Casey[1].

2 Refereed NOVA articles from 2010–2015

For the first part of this study, we perform a citation analysis on all refereed articles from the NOVA institutes in the years 2010–2015, and compare this with articles in the field of astronomy.

2.1 Method

For this part we rely on the institutes' efforts to gather all publications in the years 2010–2015 using their own bibliograpy management systems. The alternative of performing a blind search by affiliation is impractical because there are still too many publications where the affiliation information is missing or incorrect. The premise is that publications were only included if they could be attributed to a person working at the institute at the time of publication. In practise there is some leeway, for example from recently joined staff members that might have added publications that were not written at the institute yet, and from PhDs and postdocs that do not correctly attribute articles that were actually mostly prepared at a NOVA institute. However, we found through manual inspections of publication list that this represents only a small amount compared to the total number of publications. Furthermore, these effects are opposite, and we expect that they tend to cancel each other.

2.1.1 Finding articles on ADS

We used the properties and number of citations of articles as provided by ADS. To this end it was important to first find all articles on ADS. This was relatively straightforward for the NOVA institutes in Nijmegen and Groningen, as they use bibliography systems that keep the bibcodes. More effort was needed for the institutes in Leiden and Amsterdam, as their systems generally do not store bibcodes, but provide DOIs for some publications. In the absence of bibcodes or DOIs, we used (combinations of) title, authors, publication year, and journal name to find the articles on ADS. The final lists of articles on ADS were manually verified by the institutes to see if important articles were missing. We found that almost all refereed articles from the NOVA institutes are available on ADS. Indicative for this is that for Leiden only 3 of their about 2000 refereed articles did have a DOI but were not found on ADS, and in total 9 publications with a DOI were not found on ADS. These publications were not published in astronomical journals.

[1]The Python ADS module can be found at `https://github.com/andycasey/ads`

2.1.2 Selection process

Using the properties of articles as provided by ADS, we only kept publications that were refereed and classified as an article according to ADS. More specifically, in ADS query language we required `doctype:article` and `property:refereed`. In astronomy, refereed articles typically have a much higher quality standard than conference proceedings and non-refereed articles, although the latter also includes manuscripts in the preprint stage. Our selection ensures that preprints are only counted when they get a final publication date within the time window of this study, 2010–2015. The article properties were finalised in June 2016, so we expect the ADS listings up to 2015 to be complete, and that remaining preprints from 2015 may receive publication dates in 2016 or beyond. A similar situation can happen to preprints from 2009 that are published in 2010, and these two effects will compensate each other. As a final step, we also excluded errata, and journals that contained the word 'proceedings' or 'conference'. This was done because the bookkeeping of article properties on ADS is not completely homogeneous, and it can sometimes occur that proceedings are flagged as refereed articles. In total, 4129 articles were kept in the analysis.

2.1.3 Citation statistics

The main raw statistic is the number of citations of an article. We found that typically the (long-term) fraction of citations from non-refereed articles is small (~ 0.1–0.15), as shown in Table 1. Clearly the fraction goes down with time, as recently citing articles are more likely to be in preprint status. We therefore decided to keep all citations, refereed and non-refereed.

Table 1: The ratio of citations from refereed publications to the total number of citations

Institute	Total	2010	2011	2012	2013	2014	2015
Amsterdam	0.82	0.85	0.81	0.82	0.82	0.82	0.80
Groningen	0.85	0.87	0.85	0.87	0.84	0.83	0.74
Leiden	0.86	0.88	0.88	0.87	0.85	0.82	0.73
Radboud	0.80	0.82	0.81	0.81	0.82	0.80	0.74
NOVA	0.84	0.86	0.85	0.85	0.83	0.81	0.74

Another important statistic is the number of citations normalised w.r.t. the number of authors of the cited article. This was adopted to remove the unwanted weight towards large collaborations that publish articles with many authors. On the other hand, it does not take into account that often the first author contributes most of the work, nor does it give due credit when more authors are from the same institute. When we compare to the general field there is no affiliation information used for the author, so using normalisations that benefit the first author would not have been fair.

Finally, we also looked at the number of first-author publications by the institutes, and the performance thereof. First-author publications are typically more directly related to science performed at the institute of interest and therefore give important information on the institutes' leading position in research.

2.1.4 Comparison with the field of astronomy

We compared the citations of refereed articles in our sample to the general field, using refereed articles that were labelled as 'astronomy' in the ADS database. We also looked for the number of citations to articles in 4 top journals: Monthly Notices of the Royal Astronomical Society (MNRAS), Astronomy and Astrophysics (A&A), Astronomical Journal (AJ) and the Astrophysical Journal (ApJ). Additionally, we also did this for a 'top 6' of journals, which is the former list extended with Nature and Science.

We define the 'impact' following Kamphuis & van der Kruit (2010) as the ratio of the number of citations per paper of an institute, over the number of citations per paper in the field:

$$\text{impact} = \frac{N_{\text{cit, inst}}}{N_{\text{pub, inst}}} \Big/ \frac{N_{\text{cit, field}}}{N_{\text{pub, field}}}. \tag{1}$$

Essentially, this is a ratio of the mean numbers of citations per paper compared to a field, and indicates if the group stands out from the field. In all cases we only compare refereed astronomy articles.

Other advanced statistics such as the h-index are not relevant for this part of the study because we focus on a selected publication time-window for an institute, and do not look at the full careers of astronomers.

2.1.5 First-authorship detection

First-authorship in astronomy is often given to the person that contributed the most to a scientific article. The number of first-author articles can therefore give an impression of the contributions of NOVA to astronomy. We determined which articles had a NOVA-affiliated first author by checking the affiliation and/or the first author's name. To this end a list was provided by Wilfried Boland with tenured staff members at the four NOVA institutes, and optionally their secondary affiliations. Examples of secondary affiliations are the NWO funded Netherlands Institute for Radio Astronomy (ASTRON) and the Netherlands Institute for Space Research (SRON).

For the Netherlands it is a tradition that a large fraction of publications are lead by PhDs, postdocs and non-tenured staff. For this reason we did not limit the matching of the first author to staff members only. We decided that a first-author match was already achieved when the first author had an affiliation of one of the NOVA institutes. But any secondary affiliation could only be matched when the name of the author also matched. In case no affiliation information was present we could only rely on matching the author names with our institute author list. This level of complexity was necessary to avoid accidentally marking authors from other institutes as a NOVA first author, which would be more likely to happen if they have a commonly occuring family name such as 'Smith'.

To further improve the completeness of the first-author publications we augmented the staff list with all active non-tenured staff, and a selection of postdocs and PhDs. This helped to label first-authorship in cases where no affiliation information was available on the article record, or when the author used their secondary affiliation. This exercise was not meant to be complete, as otherwise all PhDs and postdocs that were active between 2010 and 2015 would need to be added.

Name variants were expanded in every reasonably possible form, and we used regular expressions to ignore variations in the occurrence of dots and white space. As an example, the author *Pieter Corijnus van der Kruit* might appear on articles as *P. C. van der Kruit*; *Piet van der Kruit*; *Pieter C. van der Kruit*; *Kruit, Piet van der*; and several more variations. To this end we first performed a search by family name on ADS to retrieve existing initials and spellings, and then added these to our list of name variations.

In a similar manner, we also took into account variants of the institute affiliations. Taking Groningen as an example, we found that 'kapteyn' or the wrongly typed 'kepteyn' captures most cases, as the official affiliation is often 'Kapteyn Astronomical Institute, ...'. As another example, Nijmegen resulted in variations like 'radboud', or 'nijmegen', as the affiliation is often 'Department of Astronomy, Radboud University, ... Nijmegen'.

2.2 Results

In Table 2 we show the outcome of the analysis of the articles that were provided by the institutes for the period 2010–2015. The most important results are the impacts, which show that typically the NOVA institutes are performing well above the general field of astronomy (roughly twice as many citations per paper), and also above the average article in the top 4 astronomy journals. We also find that the additional impact of Science and Nature within the field of astronomy is not significant, but is most likely larger beyond the field of astronomy.

If we look at the number of refereed publications, we see that these grow significantly with time for some institutes, which may be related to the fact that former Utrecht staff members moved to other institutes[2]. We speculate that year-to-year fluctations may be caused by the availability of new data from instruments and large surveys (e.g. Herschel, LOFAR, SDSS, Planck, LIGO), and the NOVA-funding cycle, that periodically results in a higher or lower influx of PhD students and postdocs.

The normalised number of publications takes into account the number of authors per publication, and shows a mostly similar picture for the NOVA institutes. We found that Radboud initially had many articles within large collaborations, while towards 2015 they are much more similar to the other NOVA institutes in this regard. Not surprisingly, Leiden, being the biggest institute with the largest number of faculty, postdocs and PhD students also publishes the most.

The number of citations per paper shows very clearly the increase of citations as time passes. A surprising increase of the number of citations was found for Radboud in 2014, which does not show up in the normalised citations per paper. This suggests strongly that Radboud authors contributed to a large and impactful collaboration which resulted in this spike, and indeed we found this to be the Planck collaboration. This is also reflected in the high impact number for Radboud in 2014. At the same time we find that in the more recent years the normalised number of citations becomes more similar for the NOVA institutes, with Leiden keeping its leading position.

[2]The Utrecht department of astronomy was dissolved in 2011.

Table 2: Citation statistics for the articles provided by the NOVA institutes for the period 2010–2015. The total gives the result over the full period, while the year columns indicate the number of publications during that year.

Number of refereed publications

Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	965	138	148	173	174	166	166
Groningen	1048	191	167	139	184	176	191
Leiden	2004	288	303	345	376	316	376
Radboud	836	128	104	130	149	188	137
NOVA	4129	583	626	698	743	738	741
All astronomy	146758	22776	24893	24982	24349	24598	25160
Big 4	45616	6779	7216	7926	7721	7682	8292
Big 4 + 2	47738	7072	7681	8341	8032	7997	8615

Number of normalised refereed publications

Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	146.6	19.9	24.4	29.0	25.8	24.6	22.8
Groningen	171.9	29.5	34.9	23.4	30.4	26.8	26.8
Leiden	301.9	41.9	50.8	53.3	52.6	47.7	55.7
Radboud	108.7	13.3	11.1	20.9	23.1	21.7	18.5
NOVA	666.6	95.2	108.4	119.0	119.4	110.7	113.9

Citations, per paper

Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	32.0	47.4	46.0	35.8	37.7	20.7	8.1
Groningen	27.5	41.7	38.9	33.0	27.2	16.8	9.4
Leiden	34.1	55.0	50.0	41.3	30.5	21.7	13.0
Radboud	37.3	39.5	36.9	22.7	27.2	70.0	15.6
NOVA	32.8	49.9	42.5	35.2	30.0	32.8	11.8

Citations, normalised, per paper

Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	3.4	6.2	4.0	4.5	3.4	1.9	0.9
Groningen	3.1	5.0	4.5	3.4	2.7	1.9	1.0
Leiden	4.0	6.7	6.2	4.9	3.4	2.4	1.3
Radboud	2.2	2.2	2.9	2.7	2.7	2.0	0.9
NOVA	3.6	6.3	5.0	4.4	3.3	2.3	1.2

Impact

Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	2.2	2.1	2.4	2.2	2.9	2.2	1.7
Groningen	1.9	1.8	2.0	2.0	2.1	1.8	2.0
Leiden	2.4	2.4	2.6	2.5	2.3	2.3	2.7
Radboud	2.6	1.7	1.9	1.4	2.1	7.4	3.3
NOVA	2.3	2.2	2.2	2.1	2.3	3.4	2.5
All astronomy	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Big 4	1.5	1.5	1.5	1.5	1.4	1.5	1.5
Big 4 + 2	1.5	1.5	1.5	1.5	1.5	1.5	1.5

Table 3: First-author publications of the NOVA institutes.

First-author publications

Institute	Total	2010	2011	2012	2013	2014	2015
Amsterdam	249	36	43	42	47	39	42
Groningen	327	53	56	43	64	53	58
Leiden	604	76	98	84	105	118	123
Radboud	177	18	17	31	34	36	41
NOVA	1225	163	189	186	222	225	240

In Table 3 we show the number of refereed first-author articles. Remarkable here is the strong growth of Radboud and Leiden. The number of first-author articles by a specific author is not very likely to change, but the number of authors at an institute can change a lot. Most likely the increase is caused by an increase in the number of PhDs and postdocs, and possibly also from hiring former staff from the Utrecht astronomy department.

We conclude that the publications from the NOVA institutes have a much higher impact than typical refereed astronomy articles, even when only high-quality journals are considered. Leiden, being the largest institute within NOVA, stands out extra in its impact and quantity of research. This was also confirmed by a Centre for Science and Technology Studies (CWTS) study peformed for the Faculty of Science of the University of Leiden (unfortunately not publicly released). In that study, the Mean Normalised Citation Score (MNCS), which may be compared to the 'impact' in this report, was found to be 2.15 for Leiden Observatory, indicating that they stand out in the field of astronomy. Although no such study was performed for the other NOVA-institutes, we may extrapolate that in a CWTS study, the NOVA institutes would perform at a similarly high level in astronomy.

2.2.1 Individual publications

In Figs. 1 and 2 we present the citation histories of NOVA institute publications with more than 50 citations. These diagrams show the cumulative number of citations with time, counted per month and with the zero-point at the final publication date. Some articles already show many citations before this moment because they were well cited in their preprint stage. The curves typically show a steadily increasing number of citations, with sometimes a small 'hook'-shape towards the end of the curve, which we suspect to be caused by preprint articles that cite the article in question. These preprints will eventually move to a later citation date when the official publication date is inserted.

Noticable in these figures is also a group of articles at Radboud where some lines quickly move off diagram. These are the Planck paper series, which are extremely well cited, and the most cited article goes up to 4000 citations. However, these articles are clearly exceptions from the others.

Another interesting quantity to look at is the number of highly cited articles. Well cited articles will eventually reach a certain publication level, for example a treshold of 100 citations. If we use this as a criterion for highly-cited articles, we can expect to find a bias towards older articles since they had more time to

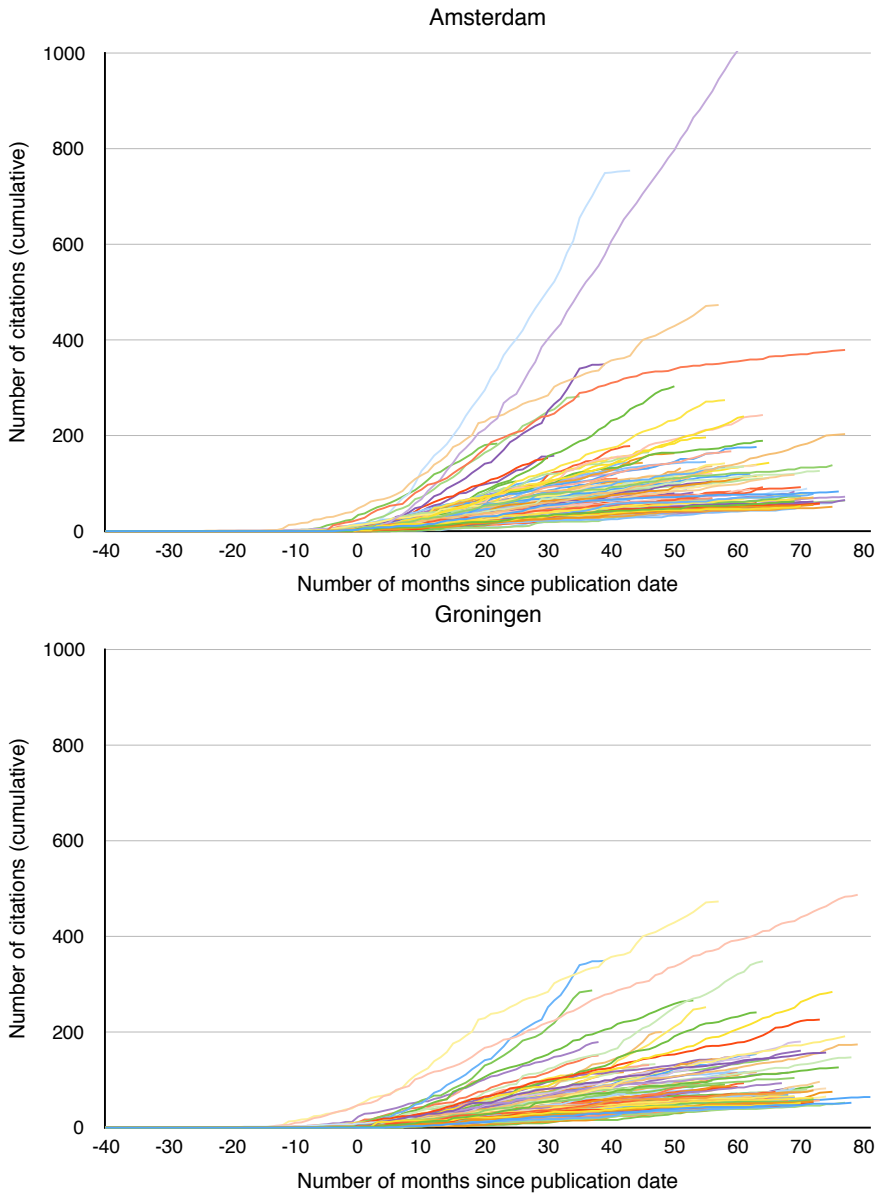

Figure 1: Citation histories of individual articles with more than 50 citations. The zero-point is the final publication date. The top panel shows Amsterdam, and the bottom panel Groningen. Lines typically start at negative time due to the preprint stage of articles. All lines end at July 2016.

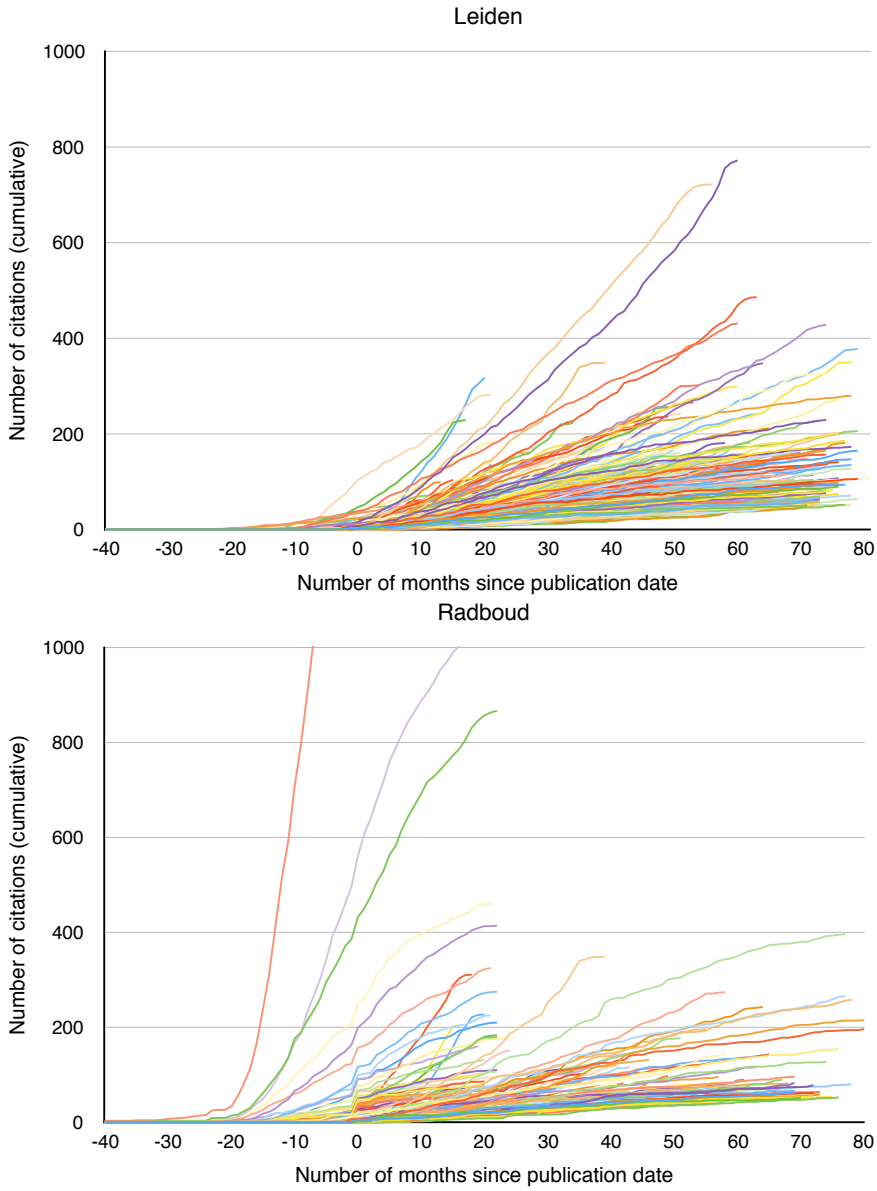

Figure 2: As Fig. 1. The top panel shows Leiden, and the bottom panel Radboud. The highly cited articles that go off-diagram for Radboud are the Planck paper series.

get cited. Instead, a more reasonable approach seems to compute the citation rate. We define the citation rate as the ratio of the total number of citations over the time since the final publication date. This assumes the citation rate is roughly constant with time, a reasonable assumption as can be seen in Figs. 1 and 2.

Table 4: High and low performing publications

Number of articles with more than 100 citations

Institute	Total	2010	2011	2012	2013	2014	2015
Amsterdam	54	6	18	13	12	5	0
Groningen	37	13	11	7	6	0	0
Leiden	131	37	35	29	20	8	2
Radboud	44	7	6	4	6	18	3
NOVA	228	57	53	46	38	29	5

Number of articles with more than 2 citations/month on average

Institute	Total	2010	2011	2012	2013	2014	2015
Amsterdam	70	3	13	13	19	14	8
Groningen	55	8	8	7	9	9	14
Leiden	188	22	28	28	34	29	47
Radboud	75	6	5	4	10	39	11
NOVA	334	36	40	45	64	84	65

Number of articles without citations

Institute	Total	2010	2011	2012	2013	2014	2015
Amsterdam	16	1	2	3	0	2	8
Groningen	27	0	6	1	5	3	12
Leiden	44	3	0	4	3	9	25
Radboud	24	2	2	7	5	3	5
NOVA	107	6	10	15	13	16	47

In Table 4 we show the number of articles that would be considered top-cited, by selecting those that meet the 100 citations treshold. As mentioned before, this is biased towards older articles. We therefore also show the number of papers that have a mean citation rate per month at 2 or more. We chose 2 because it gives a similar total number of top-cited papers to when selecting by at least 100 citations. The citation rate yields a lot less dependence on the publication year. In the bottom part of Table 4 we show the few papers that are (still) without any citations. About half of them is from 2015, suggesting that they may get cited later. Some of these articles could be interdisciplinary and published in journals where ADS performs less good on the tracking of citations, such as those in chemical physics and computational physics. And a small portion consists of book chapters and conference proceedings that were incorrectly labeled as articles on ADS, and therefore more often have low citation rates.

2.2.2 Publications divided by NOVA network

So far we have looked at the results per institute and for NOVA as a whole. For this section we investigated the internal division of NOVA by scientific topic, the so-called NOVA networks. Network 1 focuses on the formation and evolution of galaxies, Network 2 on the formation and evolution of stars and planetary systems, and Network 3 focuses on the astrophysics of black holes, neutron stars and white dwarfs. Note that a further specification of these themes and the corresponding NOVA staff members can be found on the NOVA website.

Our division of NOVA networks is based on a staff list provided by Wilfried Boland. This list includes several people that are not tenured, but long-term present at an institute. The network determination in this study is therefore mostly complete, but it leaves out publications from a PhD student or postdoc without a staff member being co-author.

Table 5: Articles in NOVA networks

Number of articles

Network	Total	2010	2011	2012	2013	2014	2015
Network 1	1368	181	209	243	264	227	244
Network 2	989	193	142	143	170	170	171
Network 3	828	99	96	137	154	171	171

Number of articles with more than 2 citations/month on average

Network	Total	2010	2011	2012	2013	2014	2015
Network 1	154	20	19	18	29	29	39
Network 2	32	3	6	5	7	5	6
Network 3	42	8	3	4	9	11	7

In Table 5 we show the networks and their number of publications. Clearly Network 1 is the largest, and produces many more high-citation rate articles. This suggests that the citation behaviour in Networks 2 and 3 may be different, possibly also related to the size of the field the networks are focussed on.

2.2.3 Collaborations between NOVA-institutes

In this section we show the collaborations between the NOVA-institutes. Collaborations are counted when the same article appears in more than one institute publication list. In Table 6 we show the articles that were shared between institutes. Typically, about 1/4 of the publications are shared with another NOVA institute, while if we look at the total output of the NOVA institutes, they make up only about 1/8 of the publications. This is because shared articles are most often shared between the same two NOVA institutes.

In Table 7 we show in more detail how the collaborations between specific NOVA institutes are. Leiden and Groningen have many shared articles, and this is also because they both do a lot of research on galaxies in NOVA Network 1: “Formation and evolution of galaxies: from high redshift to the present”. And they also collaborated a lot in NOVA Network 2: “Formation and evolution of stars and planetary systems”, because both were involved in the Herschil-HIFI instrument. At the same time, Radboud and Amsterdam collaborate a lot,

Table 6: Shared articles by the NOVA institutes. Collaborations are shown between any institute, and between 2, 3, and 4 NOVA institutes. For comparison we list the total number of publications.

Articles shared between institutes

Institute	Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	Any institute	281	63	37	37	53	46	45
	2 Institutes	172	24	24	27	35	34	28
	3 Institutes	58	21	11	5	7	7	7
	4 Institutes	51	18	2	5	11	5	10
	All publications	965	138	148	173	174	166	166
Groningen	Any institute	346	78	52	44	66	46	60
	2 Institutes	256	44	43	36	53	37	43
	3 Institutes	39	16	7	3	2	4	7
	4 Institutes	51	18	2	5	11	5	10
	All publications	1048	191	167	139	184	176	191
Leiden	Any institute	412	74	57	51	80	67	83
	2 Institutes	312	47	45	42	63	51	64
	3 Institutes	49	9	10	4	6	11	9
	4 Institutes	51	18	2	5	11	5	10
	All publications	2004	288	303	345	376	316	376
Radboud	Any institute	238	51	29	30	52	36	40
	2 Institutes	126	13	16	19	35	20	23
	3 Institutes	61	20	11	6	6	11	7
	4 Institutes	51	18	2	5	11	5	10
	All publications	836	128	104	130	149	188	137
NOVA	Any institute	553	104	79	73	111	87	99
	2 Institutes	433	64	64	62	93	71	79
	3 Institutes	69	22	13	6	7	11	10
	4 Institutes	51	18	2	5	11	5	10
	All publications	4129	583	626	698	743	738	741

mostly through Network 3: “The astrophysics of black holes, neutron stars and white dwarfs”.

Table 7: Who collaborates with whom?

Articles shared between institutes

Institute	Description	Total	2010	2011	2012	2013	2014	2015
Amsterdam	Groningen	104	37	13	11	18	10	15
	Leiden	144	33	16	15	25	26	29
	Radboud	193	50	23	26	39	27	28
Groningen	Amsterdam	104	37	13	11	18	10	15
	Leiden	297	61	42	38	57	41	58
	Radboud	86	32	8	8	15	9	14
Leiden	Amsterdam	144	33	16	15	25	26	29
	Groningen	297	61	42	38	57	41	58
	Radboud	122	25	13	12	26	21	25
Radboud	Amsterdam	193	50	23	26	39	27	28
	Groningen	86	32	8	8	15	9	14
	Leiden	122	25	13	12	26	21	25

3 The performance of publications by NOVA staff active in 2010–2015

For the second part of this study, we performed a citation analysis on all refereed articles from the NOVA staff members that were active in the period 2010–2015. We compared this to a set of reference institutes where we performed the same analysis. The main difference with the first part is that here we focus on the careers of astronomers working at the NOVA institutes.

3.1 Method

In this part we used ADS to search on (variants of) the name of the staff members. We started by searching the websites of the institutes and staff members to find the names of all tenured and retired staff members, and we used ADS to find alternative name variants and missing initials. In the process, we found that some institutes add physics professors to their institute lists, which we accepted on our list if they really published on astronomy topics. Retired staff members were also preliminary included. Some institutes list staff as ‘research staff’ or ‘lecturers’, and we excluded them as they were found not to be actively publishing. The final condition for inclusion in our analysis was to have at least 2 articles per year in the years 2010–2015. To not accidentally ‘retire’ ordinary staff members, a condition was added that they should have worked at least 20 years before being excluded for not being active.

We also tried to find the year of staff members’ PhD thesis to help resolve name conflicts with authors with similar names that published far before the author in question. In case the year of the PhD could not be found in the CV

or biography we used ADS to find the PhD thesis. In those cases where no PhD thesis could be found we instead used their first publication as start time. We then assumed a wide margin before the start time, of 6 years (based on the length of the Dutch master and PhD trajectory together), and assumed that no publications should be found from the author before this margin. Given that a PhD officially takes 3 to 4 years, this margin should at least include the most relevant publications near the time of finishing the PhD thesis.

An important difference with the first part of this study is that first-authorship was only given to the staff member where we performed the search for, as the goal was to find the performance of the staff member, and then aggregate this by institute. The importance of first-authorship in this part of the study is lower, given that many staff members are co-authors on publications by PhD students and postdocs from the groups they lead.

3.1.1 Finding the right articles

When searching refereed articles on ADS by author name there is a lot of by-catch from other authors, often in a very different field but with the same (family) name and initials. We used keyword matching to find if articles were likely to belong to the author. To start, we generated a library of articles that matched name and affiliation, and used this library to match all articles by the author. Typically we found that about 20-30 publications were needed (refereed, non-refereed) to find most, if not all articles. If the initial library of articles by an author was not big enough (for example for a new staff member), we added affiliations from earlier held positions.

The keywords were generated from the abstract, title, journal and the official keywords on ADS. We compared this with a corpus to filter out very common words such as 'and' and 'often'. The corpus was completed with a list of often occuring words in scientific publications, that were found by manually checking keywords. The aim was to have at least 5 keywords, but we made sure that words that are equally likely to occur are all included, or all excluded, to avoid randomly selecting keywords. We also used a simple word reduction algorithm that ensured that 'X-rays' and 'X-ray' were seen as the same keyword. A match on keywords occurs when at least 35% of the keywords match (with a minimum of 2 keywords matched).

Unfortunately this still led to a lot of accidentally matching keywords, for example 'X-rays', which is a topic both in physics and astronomy. We therefore also excluded certain keywords if they were used in publications by authors with the same family name and similar initials. As an example, we take *Hensen, Annie W.*, who might publish as *Hensen, A. W.* or *Hensen, A.*. But these might also be variants for *Hensen, Arnold* and *Hensen, Antoine Wilmeny*. The goal is to find publications of the latter authors and exclude the keywords they use when trying to match articles. The names on this 'not'-author list were manually inspected to ensure that no typographic error in the author name was missed, as this could contaminate the excluded keywords list.

This procedure resulted in a list of known publications, a list of very likely candidates and a list of unlikely candidates. Additionally a 'maybe'-list for further inspection was created with publications where the full first name was used. In the rare cases where the 'not'-library was in fact empty, all publications were included, even if no match on keywords could be found. In those cases the

name is unique and there is no need for a strict selection, an example being 'van Dishoeck, Ewine F.'. A manual inspection was done for authors with relatively high numbers of unmatched articles or 'maybe'-articles (compared to the initial library and the 'not'-library). If needed we added extra affiliations to improve the search.

3.2 Results

Table 8: Statistics of refereed articles belonging to authors in the NOVA, Top US and Top European institutes. Unless indicated otherwise, these are medians over the group of authors. Here we use pp for 'per paper', and pnp for 'per normalised paper', and are averages per author. The 'per year' quantities were computed using the time since the final publication date.

(Median) number of...	NOVA	Top US	Top EUR
Authors in sample	76	321	74
Papers	147.0	142.0	123.5
Normalised papers	26.3	33.2	25.8
First-author papers	17.0	24.0	13.0
Normalised first author papers	7.0	10.2	5.7
Citations	6840.5	8086.0	7137.0
Normalised citations	1120.8	1664.1	1368.4
First-author citations	946.0	1597.0	894.5
Normalised first author citations	391.7	584.6	283.7
Citations, pp	44.2	57.9	58.8
Normalised citations, pp	8.7	11.5	9.5
First-author citations, pp	7.4	11.4	6.5
Normalised first author citations, pp	2.8	4.3	2.0
Citations/yr, pp	6.35	7.57	9.3
Normalised citations/yr, pp	0.81	1.05	1.05
First author citations/yr, pp	3.94	5.08	5.18
Normalised first author citations/yr, pp	1.35	1.68	1.64
Number of normalised citations, pnp	45.1	52.3	48.2
h-index	42.5	48.0	45.5
h-index normalised	15.0	20.0	16.5
h-index first author	12.0	16.0	9.5
h-index first author normalised	9.0	11.0	7.5
Papers/yr	6.2	5.1	5.6
Normalised papers/yr	1.0	1.1	1.2
Citations/yr	293.9	295.9	330.8
Normalised citations/yr	48.2	61.1	56.2
First-author papers/yr	0.8	0.9	0.6
First-author citations/yr	45.3	57.3	33.7
Publishing years	23.0	29.0	24.5

In Table 8 we show the medians of the publication statistics distributions for authors in the NOVA, Top US and Top European institutes. And in Fig. 3 and Fig. 4 we show the distributions of several of the quantities in Table 8. The frequencies were normalised to the total number of authors in the group, which

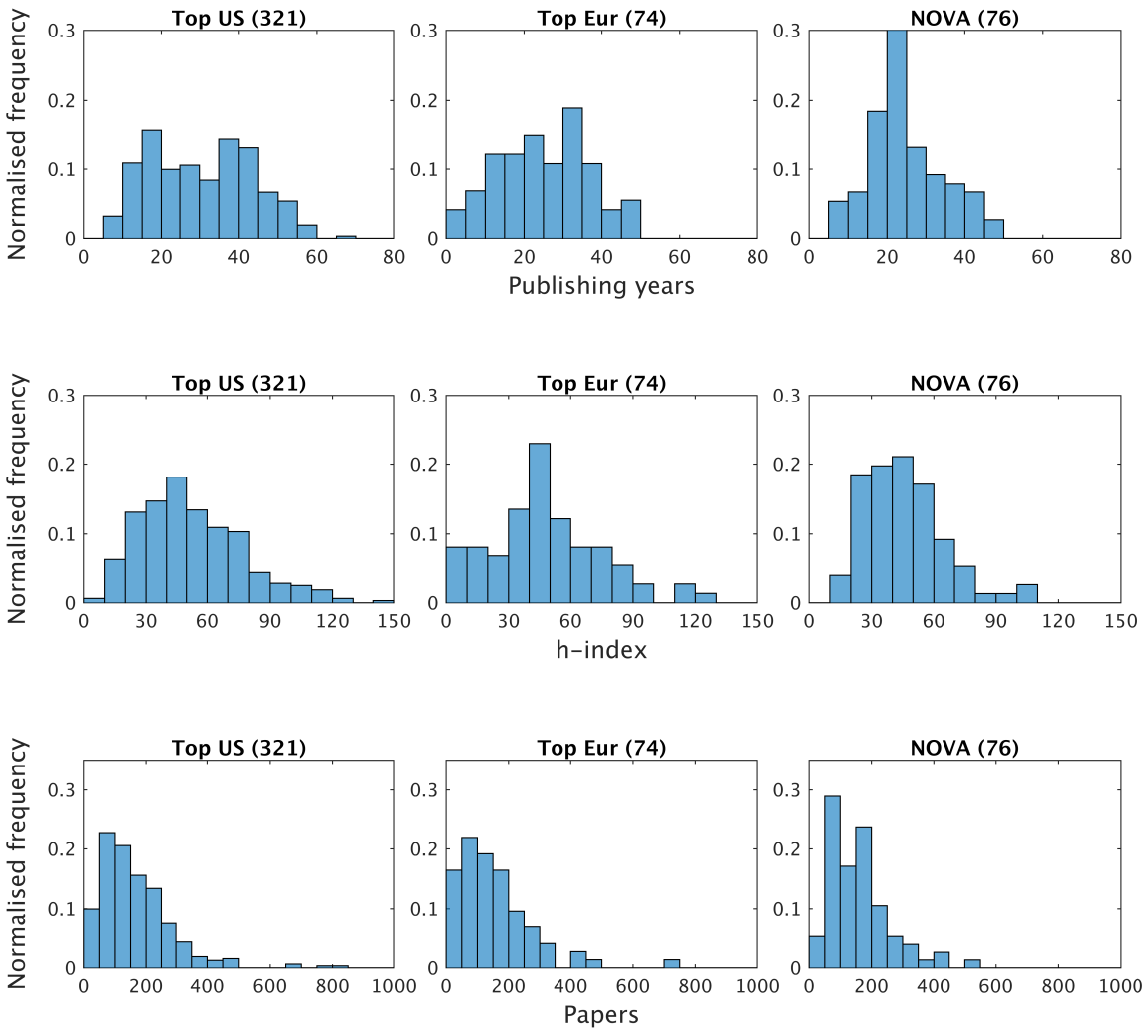

Figure 3: Histograms (frequency normalised to the total number of authors) for the distributions of the quantities as also listed in Table 8.

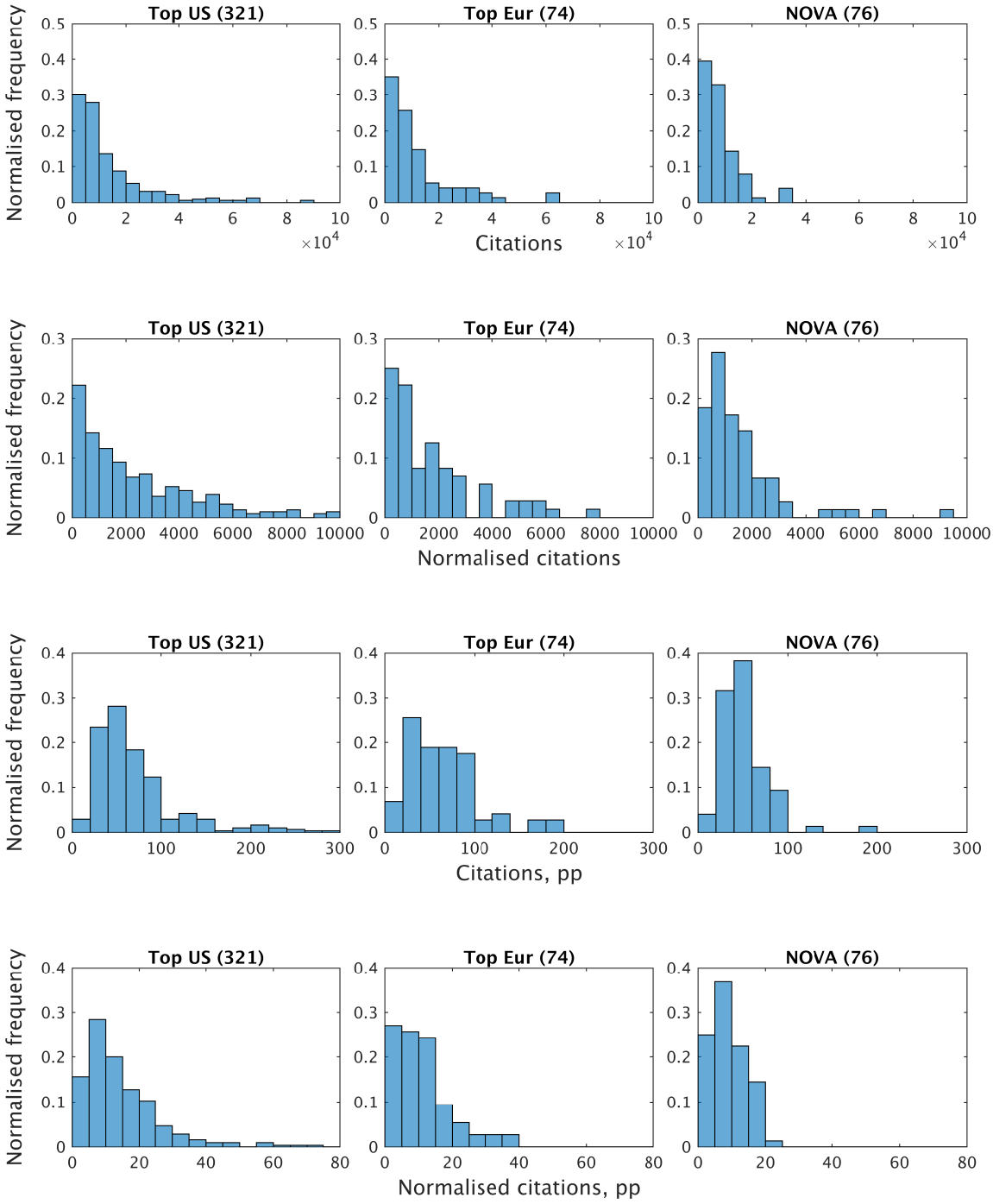

Figure 4: Continued from 3. Note that these are per author quantities, and not referring to individual publications.

enables a direct comparison of the distributions. The numbers listed as 'pp' or per paper indicate that an average was taken over all publications by the author.

From Fig. 3 and Table 8 it is clear that the NOVA institutes have a relatively young staff, and the distribution peaks near the median of 23 publishing years. The Top US and Top European institutes have a secondary peak at 35-40 publishing years, indicating a strong contribution from more senior staff members. For the European Top institutes this is compensated by many younger staff members that have much fewer publishing years. Overall we conclude that the NOVA institutes consist of a predominantly younger population than the Top US and Top European institutes. This can also explain the ranking of the total number of citations since articles have had more time to get cited, and the h-indices, for which the median is the highest for the US, then the European institutes and then NOVA.

If we look at the number of papers, then clearly the NOVA staff members are more active than their US and European counterparts, while their citations are only slightly lower (Fig. 4). Instead, in normalised citations, the distribution for NOVA peaks at a slightly higher number, even though the median is lower. In the citations and normalised citations per paper we see similar distributions, but with generally smaller tails for NOVA. This simply means that there are fewer exceptionally impactful articles where NOVA staff members contributed to. One explanation for these tails could be that the Top European and Top US institutes are involed in some of the biggest and highest impact collaborations such as SDSS and WMAP, while NOVA is not directly involved.

Overall we conclude that the performance of authors from NOVA is comparable to and not far from the performance of authors from the Top US and Top European institutes. We expect that most of the differences are due to the different length of the careers of the staff members between these groups. In the next section we will compare with the results from the previous study, Kamphuis & van der Kruit (2010). They found slightly different numbers, but also concluded that the numbers for the Top US and NOVA groups were far above those of the average astronomer.

3.2.1 Comparison with the previous study

In Kamphuis & van der Kruit (2010), a similar study was done to gather the citation statistics of NOVA staff, compared with a sample of American Astronomical Society (AAS) members, a sample of International Astronomical Union (IAU) members, and with the same list of Top US institutes. Their intention was to find how the 'typical' astronomer performs. We inserted this table in Table 9. Note that the NOVA-key column is not in the other tables of this report, because NOVA abandoned the key-researchers scheme.

As can be seen, some of the statistics are quite different, which we attribute partly to the differences in counting (we were rather strict on only including refereed articles), to the increased publication pressure in the field, and to improvements in ADS that allow it to better track the citations. Even though in an absolute sense, the numbers of this study and of the previous study are not always comparable, it needs to be stressed that in a relative sense, the performance of NOVA and the Top US institutes is much higher than of the average AAS and IAU member, and the differences between NOVA and the Top US and European institutes is relatively small.

Table 9: Table 1 from Kamphuis & van der Kruit (2010) with labels updated to be consistent with Table 8. This table shows very clearly that NOVA performs much better than the 'average astronomer', here taken to be a member of the American Astronomical Society (AAS) or International Astronomical Union (IAU).

(Median) number of...	NOVA	Top US	AAS	IAU	NOVA-key
Authors in sample	79	177	172	193	26
Papers	90	94	26	58	123
Normalised papers	26	25	7	20	33
First-author papers	21	22	8	19	23
Normalised first-author papers	11	10	3	9	13
Citations	3325	4175	544	1042	4558
Normalised citations	704	929	86	271	1213
First-author citations	795	971	112	256	1166
Citations, pp	36	43	21	18	38
First-author citations, pp	39	39	13	15	45
Normalised citations pnp	33	35	17	14	37
h-index	31	34	12	17	39
h-index first-author	13	14	4	8	14
h-index normalised	14	16	5	9	16
h-index first-author normalised	9	10	3	6	10
Papers/yr	3.9	3.3	1.6	1.9	5.6
Normalised papers/yr	1.1	0.9	0.4	0.7	1.5
Citations/yr	131	141	34	34	229
Normalised citations/yr	34	33	6	10	57
First-author papers/yr	0.9	0.9	0.5	0.6	1.2
First-author citations/yr	40	33	8	9	59
Publishing years	25	30	18	30	22

3.2.2 Comparing the NOVA institutes

Table 10: Statistics of refereed articles belonging to authors in the NOVA institutes. Unless indicated otherwise, these are medians over the group. Here we use pp for 'per paper', and pnp for 'per normalised paper'. The citation rates are in units of papers per year, and were computed using the time since the first publication. The other 'per year' indicators were computed over the timespan of the staff member's career.

(Median) number of...	NOVA	Amsterdam	Groningen	Leiden	Radboud
Authors in sample	76	16	22	27	11
Papers	147.0	129.5	113.0	160.0	176.0
Norm. papers	26.3	24.9	25.2	26.0	27.5
First-author papers	17.0	15.5	17.5	16.0	24.0
Norm. first-author papers	7.0	6.4	7.3	6.5	10.9
Citations	6840.5	7023.0	5787.5	9749.0	6833.0
Norm. citations	1120.8	1087.3	1120.8	1438.6	832.4
First-author citations	946.0	807.5	812.5	1255.0	1083.0
Norm. first-author citations	391.7	322.8	404.0	395.1	381.7
Citations, pp	44.2	47.1	42.3	46.9	38.8
Norm. citations, pp	8.7	8.2	9.9	8.7	5.9
First-author citations, pp	7.4	5.8	8.0	7.4	8.3
Norm. first-author citations, pp	2.8	2.2	2.8	3.5	2.7
Citation rate, pp	6.35	7.24	6.65	6.19	5.85
Norm. citation rate, pp	0.81	1.01	0.76	0.82	0.77
First author citation rate, pp	3.94	4.72	3.54	4.44	3.04
Norm. firs-author citation rate, pp	1.35	1.6	1.22	1.30	1.17
Number of norm. citations, pnp	45.1	48.2	44.7	44.4	34.0
h-index	42.5	41.5	41.5	52.0	42.0
h-index normalised	15.0	15.5	14.0	19.0	13.0
h-index first author	12.0	11.5	12.5	12.0	17.0
h-index first author normalised	9.0	9.0	9.0	8.0	9.0
Papers/yr	6.2	6.7	4.6	6.3	8.8
Normalised papers/yr	1.0	1.0	1.0	1.1	1.5
Citations/yr	293.9	341.1	213.4	382.4	259.0
Norm. citations/yr	48.2	50.4	43.0	48.4	46.4
First-author papers/yr	0.8	0.7	0.7	0.7	1.3
First-author citations/yr	45.3	46.4	30.3	48.9	53.6
Publishing years	23.0	21.0	24.0	24.0	19.0

In Table 10 we show the same statistics as in Table 8 but instead compare NOVA and the different institutes. Because the number of authors in the sample is small we expect a larger statistical error. As to be expected, the relatively new department in Nijmegen has the lowest number of publishing years for its staff members, which also affects the h-indices. At the same time, the more senior department in Leiden has by far the most citations and the highest h-indices.

Groningen staff members have published fewer papers per year, but they stand out in the number of normalised citations per paper. The normalised number of papers that a staff member publishes flattens out to about 1 normalised paper/year for all NOVA institutes, with Radboud performing above

the rest with 1.5 normalised papers per year. This pattern seems to repeat also in the normalised number of first-author articles, with Leiden, Amsterdam and Groningen having 0.7 normalised first-author papers per year and Radboud 1.3 per year. In the (normalised) citation rates per year, computed since the time of publication, we find that Amsterdam stands out. Regarding the normalised citations/year over the full career of staff members we see that the institutes are roughly in the same range.

Most of these differences can be attributed to the different age composition of the staff members at the NOVA institutes, but also to differences in publication practises. Overall, we may conclude that Leiden stands out in absolute numbers because of its size and more senior staff, but that the other NOVA institutes come out at a very similar or sometimes higher level when looking at the detailed numbers. Combined with the excellent outcome of the CWTS study on Leiden Observatory, we repeat here that NOVA as a whole is performing on a very high level.

4 Conclusions

In this study we investigated the performance of NOVA-related publications and NOVA staff members. In the first part of this study we looked at NOVA-related publications, and found that they stand out from the average astronomical publication, but also from the average publication in the top journals in astronomy. In the second part of this study we looked at the bibliographic performance of NOVA staff members, and found that they publish at a very comparable level to that of the top institutes in the US and Europe. There are very likely intrinsic differences due to staff age, scientific field, and the participation in large collaborations that affect the statistics. However, using a table from the previous study, we stress that NOVA and the Top-US institutes stand significantly from the 'average astronomer'. Given that NOVA performs at a level very similar to that of the Top-US and Top-European institutes, we may conclude that NOVA can be considered a member of the world-leading institutes on astronomy.

We would like to thank Edwin Henneken for his continuous support for using NASA's ADS service. Michael Maier is thanked for his feedback on an earlier manuscript and his support on the Python ads module. Peter Kamphuis is thanked for kindly providing us with material and codes from his earlier study.

References

Kamphuis, P., & van der Kruit, P. C. 2010, ArXiv e-prints, 1011.5311

Kinney, A. L. 2008, ArXiv e-prints, 0811.0311

A Top institutes

A.1 Institutes in the United States

The US top institutes are based on a ranking by Kinney (2008), and two added institutes, as also used in Kamphuis & van der Kruit (2010). This report dates

from 2008, but these numbers should still reflect mostly the situation in astronomy today. The list of astronomy departments is:

- Caltech
- UC Santa Cruz
- Princeton University
- Harvard University
- Colorado
- SUNY Stony Brook
- Johns Hopkins Univ.
- Penn. State Univ.
- Univ. Michigan
- Univ. Hawaii
- Univ. Wisconsin
- UC Berkeley
- Michigan State Univ.
- UCLA
- Univ. Texas

A.2 Institutes in Europe

The top European institutes are not a full representation of European astronomy, but includes the Max-Planck institutes of Germany and the Institute of Astronomy in Cambridge, which form a short list of high quality research institutes in astronomy within Europe.

- MPIA Heidelberg
- MPA Garching
- MPE Garching
- IoA Cambridge