Comparing Google Scholar and ISI Web of Science for Earth Sciences

Susanne. Mikki, dr.scient. and Senior Academic Librarian
Bergen University Library, Science Library, Bergen, Norway
E-mail: susanne.mikki@ub.uib.no

Abstract

In order to measure the degree to which Google Scholar can compete with bibliographical databases, search results from this database is compared with Thomson's ISI WoS (Institute for Scientific Information, Web of Science). For earth science literature 85% of documents indexed by ISI WoS were recalled by Google Scholar. The rank of records displayed in Google Scholar and ISI WoS, is compared by means of *Spearman's footrule*. For impact measures the *h-index* is investigated. Similarities in measures were significant for the two sources.

Introduction

The two sources studied are different in goals and content, but keep track of citation data suitable for comparison and scientific metric studies.

Main characteristics can be summarized as followed. The thorough content selection process by ISI WoS [No Date] is based on quality of citation data, publication standards and expert judgement, where regularity and keeping schedule is the most basic assumption [GARFIELD, 1990]. ISI WoS computes the number of citations for each year and depends on regular occurrences to analyse citation trends. Its collection, therefore, is restricted to a set of high-qualified journals (more than 7000) and a minor part of books and proceedings. Citation data is controlled carefully and only exact matches of references contribute to the citation metrics, often leading to under-reported counts. A more complete list may be obtained through the option *Cited Reference Search*.

Google Scholar [No Date] samples a wider variety of publications. Its collection is based on an automate reference recognition and agreements with their partners, the journal publishers, database vendors or scholarly societies, for using their collections and metadata. However, search results include non-scholarly sources, erroneous citation data and items apparently not matching the search expression. In addition, insufficient grouping inflates the result list. The same is true for the "Cited by" feature, over-reporting citation counts [JACSÓ, 2006]. Google

Scholar does not analyse citation data per year, but only totals to date. It is therefore not suitable for analysing trends such as offered by ISI WoS. The service has been criticised for its inefficient use of metadata. It also has been criticised for its inaccurate notification of content.

Despite different database policies, several comparative studies are carried out to assess content of Google Scholar in relation to ISI WoS and other approved databases. Degree of coverage in Google Scholar is one aspect which has been investigated, showing that for multidisciplinary databases coverage range from 68% to 94%, being highest for science and medicine [NEUHAUS, NEUHAUS, ASHER, & WREDE, 2006]. However, considering only subject specific databases omitting life sciences, Google Scholar coverage decreases to less than 50%, for GeoRef [No Date] being 26% only. As described on their home page "*The GeoRef database, established by the American Geological Institute in 1966, provides access to the geoscience literature of the world (…) covers the geology of North America from 1785 to the present and the geology of the rest of the world from 1933 to the present.*" Google Scholar's poor coverage of GeoRef may partly be explained by year of publishing. GeoRef is a database consisting of items published since 1785, where the old ones and the documents they are cited by, are not likely posted on the web and therefore not retrievable by Google. In addition GeoRef includes translated titles and abstracts to publications written in non-English languages, whereas Google Scholar has a pronounced bias towards English language [NEUHAUS et al., 2006; NORUZI, 2005].

The study by Walters [2007] assessing coverage of a specific subject in social science (later-life-migration) shows that Google Scholar in relation to seven scientific databases, ISI WoS included, indexes the largest number of core articles (93%), considering publications from 1990 to 2000 only.

Ranking of search results is another aspect for assessing databases, assuming data for citation counts are available. Studies show that similarity in ranking is quite high between Google Scholar and the approved sources ISI WoS and Scopus[1].

Pauly and Stergiou [2005] compared citation counts between Google Scholar and ISI WoS for a wide range of natural sciences, finding high correlation. They also report less citation counts

[1] Scopus [No Date] is developed by Elsevier and claims to be "the largest abstract and citation database of research literature and quality web sources". References go back to 1996.

for older publications in Google Scholar, suggesting citing articles, probably old themselves, are not posted on the web.

Various studies [BELEW, 2005; MEHO & YANG, 2007; PAULY & STERGIOU, 2005] document similar citation counts for articles indexed in both ISI WoS and Google Scholar. For ranking institutions or scientists, Meho and Yang [2007] state that data retrieved by ISI WoS alone is insufficient for determining science impact, since different databases obviously differ in coverage dependent on research field.

Bar-Ilan et al.[2007] examined rankings by ISI WoS, Scopus and Google Scholar of 22 recognized scientists. Their results show high similarity between the databases. The normalized *Spearman's footrule,* for example, was calculated to 0.884 comparing ISI WoS and Scopus, 0.830 comparing ISI WoS and Google Scholar and 0.780 comparing Scopus and Google Scholar, proving good agreement in ranking.

The *h-index* has recently got attention and is assumed to be a robust measure for scientific performance and impact [HIRSCH, 2005]. It cuts off a long tail of low-cited items and balances inflated citation counts by a small number of publications, which result from co-authorship. Vanclay [2007] investigated data retrieved by ISI WoS and Google Scholar and found similar values for the two services. Same results were obtained by Bar-Ilan [2008], although, discrepancies were higher between ISI WoS and Google Scholar than between ISI WoS and Scopus.

Aims and objectives

This study presents a comparative assessment of Google Scholar and ISI WoS for earth sciences. Content is analysed by overlap, coverage and ranking of the two services. Unique items with high citation counts are investigated more closely to reveal lack or weaknesses of either the one or the other service. To which extent year of publishing is affecting coverage for data presented here is investigated by a single test. Impact studies are carried out in terms of *Spearman´s footrule* the *mean citation count* and the *h-index*.

Methodology

To compare lists of search results, searches were carried out for 29 authors. In Google Scholar searches included in addition to initials also full first names. The majority of scientists are connected to the University of Bergen, mainly working with climate and petroleum geology issues. The oldest document indexed in ISI WoS was published in 1962 (Author 17). Author names that could cause spelling or transcription problems were left out, as inclusion of such authors could introduce a bias in the data. In the case of homonyms three searches were narrowed by discipline. These returned slightly incomplete results and reduced degree of coverage (refer authors 6, 15 and 19, last column in Table 1 and Figure 1). Data were collected in November and December 2007.

The complete result list, including Google Scholar citations marked [BOOK], [CITATION] and [PDF] (representing about 55% of all results) were stored. For ISI WoS, citations were exported to EndNote, and for Google Scholar results were copied from the web and further treated in MS Word and Excel. Calculations were carried out in Matlab. To find identical records title phrases were compared, using no more than 50 characters. Punctuations, blanks and special characters were omitted. The limitation of characters compared was necessary since Google Scholar may display truncated title phrases. 7,7 % of items in Google Scholar, not counting the ones grouped, were identified as multiples and excluded for further investigation. This is also the case for their citations.

Citation data in ISI WoS were obtained through the ordinary search interface, only listing exact matches of references. Citation counts are therefore expected under-reported, a fact to keep in mind when comparing the services.

Citing records in Google Scholar were not individually controlled, but are expected to be erroneous as reported by Jacso [2006] and described in the introductory part.

Comparing search results in ISI WoS and Google Scholar, overlap (O) and, coverage (C) is calculated. Citation counts and the normalized *Spearmans's footrule* (F) is calculated for comparing ranking of the two services. Furthermore the *h-index* and *the mean citation count* are used for impact studies.

Overlap (O) is defined by

$$O = \frac{N_{P(ISI)} \cap N_{P(GS)}}{N_{P(ISI)} \cup N_{P(GS)}},$$

where $N_{P(ISI)}$ is the number of publications retrieved from ISI WoS and $N_{P(GS)}$ is the number of publications retrieved from Google Scholar. If $N_{P(ISI)}$ is equal to $N_{P(GS)}$, O is equal to one, and in case of completely different publication sets O is equal to zero.

Coverage (C) is defined as the recall ISI WoS records in Google Scholar

$$C = \frac{N_{P(ISI)} \cap N_{P(GS)}}{N_{P(ISI)}},$$

where the value of C lies between zero and one, equal to one for 100% coverage, and equal to zero for 0% coverage.

The footrule is a measure for the relative ranking of the results retrieved in the two databases. It is derived by considering only shared references ($Z = N_{P(ISI)} \cap N_{P(GS)}$), and re-ranking search results from 1 to Z, omitting publications unique in either ISI WoS or Google Scholar. The re-ranked series σ_{ISI} is set to (1,2,3, … ,Z) . σ_{GS}, a permutation to σ_{ISI}, is the corresponding rank in Google Scholar. For example if article four in Google Scholar is the same as article one in ISI WoS, the series σ_{GS} starts with 4. The larger the difference (σ_{ISI} - σ_{GS}), the more dissimilar is the ranking.

The normalized *Spearman´s footrule* is defined by

$$F = 1 - \frac{\sum_{i=1}^{Z} | \sigma_{GS\,i} - \sigma_{ISI\,i} |}{F_{\max}},$$

where $F_{\max} = \frac{Z^2}{2}$ for Z even, and $F_{\max} = \frac{(Z+1)(Z-1)}{2}$ for Z odd.

Thus F is ranging form zero to one, being equal to one for identical ranking.

Citation counts are plotted on a log-log scale according to the power-law distribution of citation counts versus citation rank for a given article [BELEW, 2005; REDNER, 1998].

To compare scientific impact two measures are applied:

The *mean citation count*

$$\overline{CC} = \frac{N_C}{N_P},$$

where N_P is the number of publications and N_C is the number of citations.

The *h-index*

“A scientist has index h if *h* of his or her N_P papers have at least *h* citations each and the other ($N_P - h$) papers have ≤ *h* citations each.” [HIRSCH, 2005]. The *mean citation count* and the *h-index* are comparable in size as long as the tails of the power law distribution of citation counts versus citation rank are short, i.e. α the reduction rate ($e^{-\alpha x}$) is big.

Results

Totally 1573 items in ISI WoS and 5048 items in Google Scholar were retrieved. About 55% (2766 items) of these were marked as [BOOK], [PDF] or [CITATION] in the latter. Numbers referred to in this and following paragraphs reflect 26 authors. 27% of database content is overlapping (Figure 1). More than 3 times as many documents were located in Google Scholar making it to the largest repository, but also to a service returning much noise.

Unique items and coverage

4% of all items were unique for ISI WoS. This corresponds to about an eighth part of indexed articles within ISI WoS (Figure 1). Records are not investigated, but assumed to be peer reviewed scientific contributions that are important to be aware of.

69% of all items retrieved were unique for Google Scholar. A total of 107 unique items occurring among the first 20-30 on each result list were investigated more carefully. Of these highly cited documents belonged 55 to journal articles, 22 to conference proceedings, 17 to books or book chapters, reports or, due to poor bibliographic notifications, not identifiable items.

Examples of articles not indexed in ISI WoS belong to the following journals *Climate Dynamics* (finding early issues not indexed), *Tellus (A), Ocean modelling*, *Boreas* and *Catena*. Whereas *Geological Society of America - Special Paper*, *Geological Society of London – Special Publication* were examples of publications not indexed at all in ISI WoS, indicating that certain geological subjects may be underrepresented by this service.

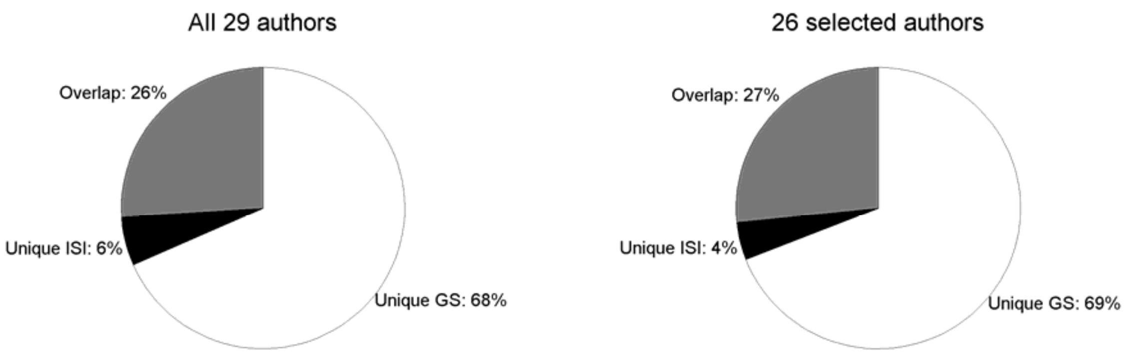

Figure 1: Unique items and overlap of search results. For the selected 26 authors (right figure) three authors are excluded. Their searches were narrowed by discipline and returned incomplete results.

Coverage of ISI WoS in Google Scholar was as high as 0.86. For author 9, 14 and 26 even the total amount of articles were recalled ($C = 1$), confer Table 1, last column.

This high degree of coverage is even higher than reported by Neuhaus et al. and Walters a couple of years earlier. Increasing degree of coverage might be explained by Google Scholar's effort to collect content, and by sample selection, since this study includes publication younger than 1962 only. This finding shows that the academic community within earth sciences is well served by Google Scholar, even better than expected. Still, the missing publications may represent very much recognized sources. To get a more reliable picture of science coverage it would therefore be wise to investigate a third source, for example Georef. As discussed in the introductory section, for this source much more unique items are expected.

To investigate the effect of publishing year on coverage, one test was conducted for publications younger than 2000. In contrast to findings by Neuhaus et al. [2006] and Walters [2007], coverage did not increase as expected, finding that digital-born publications were no more likely to be found in Google Scholar. Coverage in fact decreased by 9% to 77%, confer Table 1, last two rows.

A single case study; coverage, citation counts and ranking

To elucidate method and results, one author (05) is discussed in detail. As seen by Figure 2A, 36% of all items were indexed in both Google Scholar and ISI WoS, 7% were unique in ISI WoS, while the majority of items, 57%, were unique ones in Google Scholar. Coverage of ISI WoS in Google Scholar was 85% ($C = 0.85$), compare also Table 1. Figure 2B shows the first 20 highest ranked items in ISI WoS and their twins in Google Scholar. All the twenty items in ISI WoS were recalled by Google Scholar. Unique items among the first twenty in Google Scholar (item 1, 12 and 18, marked on y-axis) belonged to a book and proceedings. Figure 2C shows the typical power-law distribution for citation counts versus rank for both sources. It reveals that the list is not strictly sorted by the number of citations; confer peaks between rank 30 and 40. It also shows that the tail of rarely cited items is considerably longer in Google Scholar, indicating irrelevant noise.

ISI WoS aggregated 775 citations against 835 in Google Scholar, confer Table 1. The highest citation count (224) was not unexpectedly obtained by the book in Google Scholar. Even ISI WoS received fewer citations considering all items, the citation count of the 44 shared items was higher for ISI Wos (754) than for Google Scholar (461). As illustrated in Figure 2D, the majority of data points is located beneath the dashed line, with higher counts for ISI WoS.

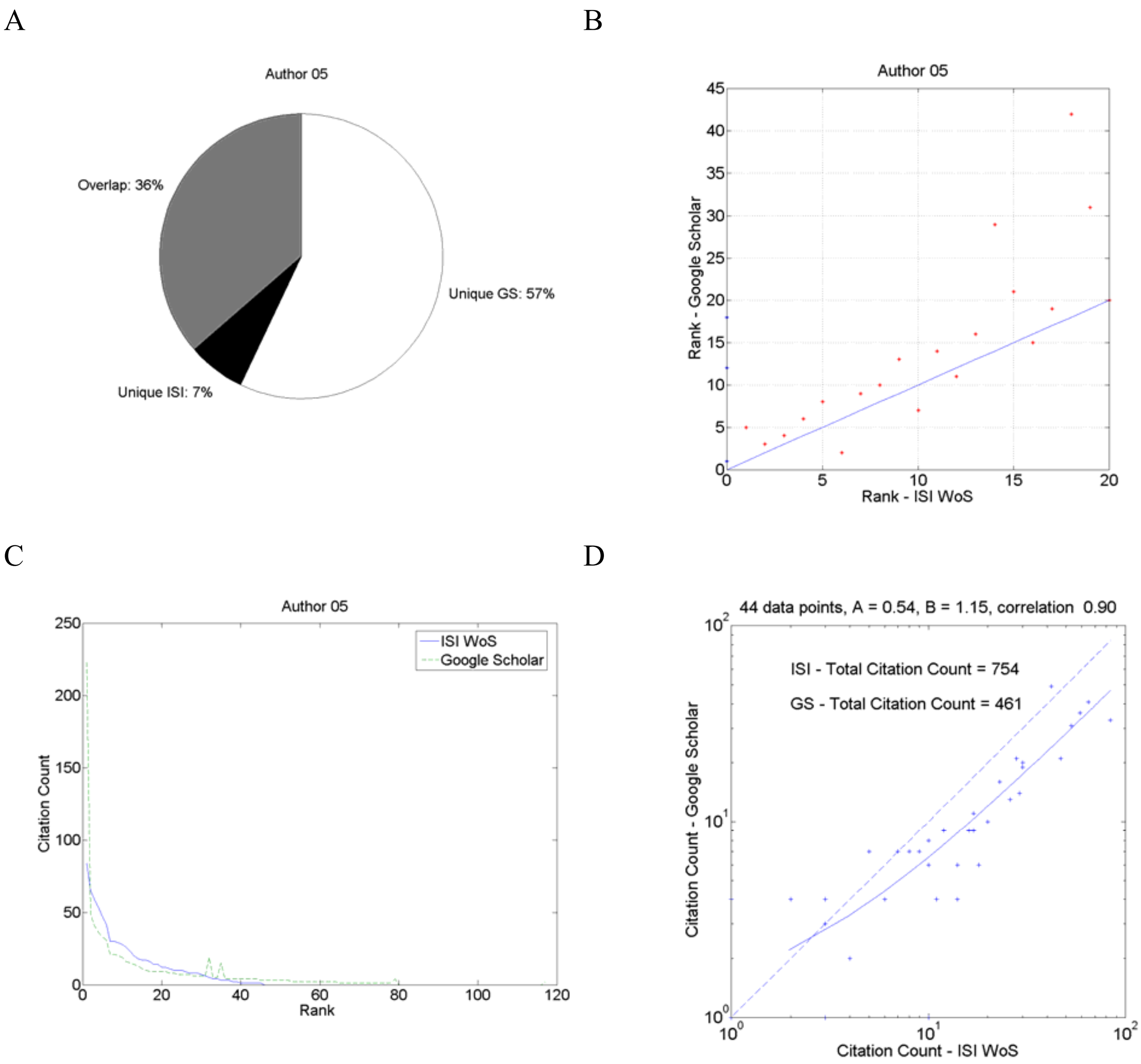

Figure 2: Results for Author 05

The scientific performance of the author is calculated for the *mean citation count* to 14.9 and *the h-index* to16 for ISI WoS. Corresponding values for Google Scholar are 7.1 and 13. Google Scholar's low value for the *mean citation count* is a consequence of the large number of hits causing noise for impact studies. However, the *h-index* is of same size being slightly higher in ISI WoS, with a relative *h-index* of 1.23. The most cited publication, the book, is not given credit by the measures applied. Firstly, it is not indexed by ISI WoS, secondly, it does not effect the *h-index*, and thirdly, it contributes little to the *mean citation count* derived by Google Scholar due to the long tail of low-cited items.

General results for citation counts and ranking

The total number of citation counts in ISI WoS was 27189 against 29053 in Google Scholar (Table 1). Results were quite similar; counts being slightly lower in ISI WoS considering 26 authors only (omitting three discipline-refined searches). As discussed earlier, counts are not verified and presumed under-reported by ISI WoS and over-reported by Google Scholar. However, due to Google's incomplete grouping of identical records citations belonging to duplicates were ignored.

Although citation counts were similar they varied from author to author with a standard deviation of 0.4 for the relative citation count. Author 1 for example received most citations in Google Scholar, about 2.5 times more than in ISI WoS. In contrast Author 7 received highest counts in ISI WoS, 1.76 times more than in Google Scholar (Table 1).

Table 1: ISI WoS and Google Scholar compared for sum publications, citations, h-index, mean citation count, relative citation count and *Spearman's footrule* for the relative ranking. The last two columns present the coverage of ISI WOS articles in Google Scholar, the second-last only considering items published after 2000 (py > 2000). Sum or mean values in parentheses belong to all 29 authors, whilst values in bold belong to 26 authors, omitting author 6, 15 and 19, since their searches also included a discipline which might have returned erroneous results.

author	ISI sum publ	GS sum publ	ISI citation count	GS citation count	ISI h-index	GS h-index	relative h-index	ISI mean citation count	GS mean citation count	relative citation count	footrule	coverage py>2000	coverage
	$N_{P(ISI)}$	$N_{P(GS)}$	$N_{C(ISI)}$	$N_{C(GS)}$	h_{ISI}	h_{GS}	$\frac{h_{ISI}}{h_{GS}}$	$\overline{CC_{ISI}}$	$\overline{CC_{GS}}$	$\frac{N_{C(ISI)}}{N_{C(GS)}}$	F	$C_{py>2000}$	C
(Σ_{29})	(1573)	(5048)	(43028)	(40908)	(16.7)	(16.0)	(1.04)	(20.3)	(6.5)	(0.93)	(0.80)	(0.75)	(0.84)
Σ_{26}	**1264**	**4196**	**27189**	**29053**	**15.3**	**15.0**	**1.03**	**18.3**	**5.9**	**0.90**	**0.81**	**0.77**	**0.86**
1	9	20	43	107	4	3	1.33	4.8	5.4	0.40	0.63	0.50	0.89
2	67	209	350	668	12	13	0.92	5.2	3.2	0.52	0.74	0.94	0.88
3	61	154	1479	1204	23	22	1.05	24.2	7.8	1.23	0.84	0.80	0.86
4	65	265	541	1136	13	14	0.93	8.3	4.3	0.48	0.80	0.73	0.87
5	52	117	775	835	16	13	1.23	14.9	7.1	0.93	0.81	0.75	0.85
6	20	35	421	292	9	8	1.13	21.1	8.3	1.44	0.75	0.40	0.40
7	99	196	1125	639	16	10	1.60	11.4	3.3	1.76	0.78	0.77	0.49
8	20	79	161	134	6	6	1.00	8.1	1.7	1.20	0.86	0.70	0.88
9	26	112	142	296	7	6	1.17	5.5	2.6	0.48	0.67	1.00	1.00
10	55	127	941	918	17	17	1.00	17.1	7.2	1.03	0.85	0.63	0.83
11	42	184	445	884	13	15	0.87	10.6	4.8	0.50	0.79	0.87	0.93
12	60	173	1971	1277	25	20	1.25	32.9	7.4	1.54	0.87	0.86	0.93
13	16	45	88	140	6	7	0.86	5.5	3.1	0.63	0.84	0.71	0.94
14	18	59	186	274	8	10	0.80	10.3	4.6	0.68	0.86	1.00	1.00
15	223	543	13465	9058	54	43	1.26	60.4	16.7	1.49	0.78	0.62	0.69
16	98	259	3723	2729	36	28	1.29	38.0	10.5	1.36	0.73	0.69	0.80
17	101	461	5485	6674	38	42	0.90	54.3	14.5	0.82	0.75	0.55	0.69
18	14	93	467	643	8	12	0.67	33.4	6.9	0.73	0.86	0.67	0.92

19	66	274	1953	2505	23	23	1.00	29.6	9.1	0.78	0.80	0.75	0.79
20	55	179	1716	1478	21	18	1.17	31.2	8.3	1.16	0.84	0.88	0.87
21	70	277	2318	2078	29	23	1.26	33.1	7.5	1.12	0.84	0.44	0.64
22	37	106	353	488	9	12	0.75	9.5	4.6	0.72	0.82	0.80	0.92
23	53	135	473	359	11	9	1.22	8.9	2.7	1.32	0.81	0.89	0.91
24	45	188	606	825	13	13	1.00	13.5	4.4	0.74	0.87	0.80	0.84
25	79	285	692	1983	14	21	0.67	8.8	7.0	0.35	0.78	0.81	0.88
26	10	79	144	293	6	9	0.67	14.4	3.7	0.49	0.76	1.00	1.00
27	51	175	2118	2263	26	27	0.96	41.5	12.9	0.94	0.89	0.75	0.86
28	14	53	273	298	8	8	1.00	19.5	5.6	0.92	0.87	0.67	0.85
29	47	166	574	430	14	12	1.17	12.2	2.6	1.34	0.79	0.80	0.86

For overlapping cited articles (represented by 1221 data points), considering all 29 authors, the citation counts were plotted on a log-log scale (Figure 3). In this figure, the dashed line indicates identical citation counts (A=1). The linear regression line is tipping under the dashed line, A=0.52, which illustrates that ISI WoS accumulates significantly more citing articles, in total 37892 versus 26992 in Google Scholar. These results were not unexpected and in accordance with findings by Belew [2005] and Pauly and Stergiou [2005]. The correlation of citation counts in both datasets was high, with a calculated correlation coefficient of 0.74. For comparison, the correlation coefficient was calculated to 0.5 by Belew [2005] and to 0.83 and higher by Pauly and Stergiou [2005]. All studies prove high similarity in citation counts by the two services.

The scientific performance of the selected scientists is calculated by the *mean citation count* and *the h-index*. On average, values for the *h-indexes* were almost identical, being 1.03 for the *relative h-index* (Table 1), i.e. slightly higher for ISI WoS. The standard deviation for the relative *h-index* was 0.2. For 5 authors (1, 7, 18, 25, 26) the deviation was, without obvious reason, higher than 0.3. These authors could for example not be linked to a distinguished sub-discipline with deviating publishing patterns.

In this study the *h-index* has confirmed to be a robust measure returning similar values for the two databases. The measure ignores the impact of inflated citation counts of single documents, and a long tail of rarely cited documents, as in case of Google Scholar.

The *mean citation count* for ISI WoS may deviate considerable from the *h-index* (Table 1), dependent on the rate of reduction for the power law distribution of citation counts versus citation rank. Deviations from *mean citation count* will indicate whether a particular author has either a considerable higher citation count for his or her production (as author 17, emeritus today), or the production is large, but does not achieve that high impact (author 2).

The *h-index* functions as a rule of thumb for scientific assessment, while the *mean citation count* can be related to it in order to gain a more nuanced view.

The relative ranking of the results by the two databases is calculated by the normalized *Spearman´s footrule (F)*. On average, *F* was 0.81 (Table 1), indicating high similarity in ranking for articles indexed in both databases. This finding is in accordance with results by Bar-Ilan [2007] who calculated the normalized *Spearman´s footrule* to 0.83.

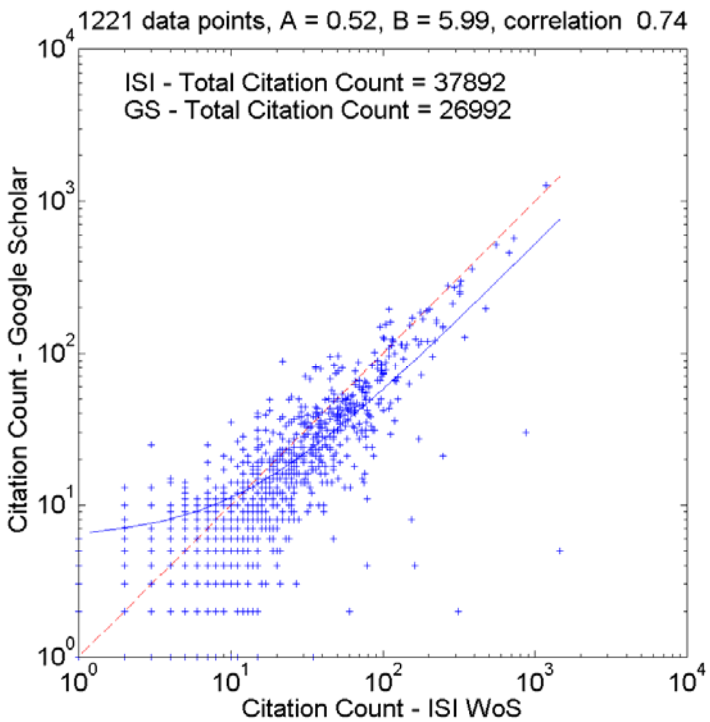

Figure 3: Citation counts for articles indexed in both ISI WoS and Google Scholar.

Even if the citation counts in this study are comparable, the citing documents in Google Scholar and ISI WoS will differ. The validity of the citations was not examined in this study. Since content in ISI WoS is selected carefully and partly controlled manually it provides a more dependable set of citations. In contrast, citations in Google Scholar are extracted only automatically from reference lists of recognized scientific literature. It samples a wider variety of publications, but lacks numerous important publishers and contains a lot of annoying noise. Therefore, ISI WoS still remains the highest acknowledged citation service.

Conclusion

The amount of earth science content is comprehensive in Google Scholar. It covers about 85% of content indexed by ISI WoS. Unexpectedly, for documents published after 2000 the degree of coverage did not increase, due to increased digital publishing and web posting.

For impact studies the *h-index* has proofed to be a robust measure leading to similar values for the two sources and may serve as a rule of thumb for performance assessments.

Similarity in ranking of the two services is significant by terms of *Spearman's footrule*, considering overlapping items. For overlapping items, ISI WoS accumulates significantly more citing articles, by which it confirms its position as the leading citation index.

The number of search results and their citations is otherwise higher in Google Scholar. The service returns highly cited sources not indexed by ISI WoS, but also a long tail of minor relevant items, barely matching the search expression.

Future studies

Google Scholar is still in its beta version and expected to be developed further, both regarding functionalities and amount of content. To survey changes it would be worthwhile to rerun the calculations after a year or so. In this study only the amount of citations are compared. For future studies it would be interesting to investigate the recall of citing records to proof validity of Google Scholar's *Cited by* feature.

In addition to citation metrics used in this study, alternative measures could be applied and compared.

GeoRef is a highly recognized database for geological disciplines. To determine coverage and overlap it would be wise to compare content with this source too.

The effect of publishing year on coverage may be worthwhile to examine for different time spans. In addition cover of non-English literature, being highly relevant for certain earth science disciplines, may be subject for future studies.

Acknowledgements

I am grateful for the programming help I received from Knut Barthel at the Geophysical Institute (University of Bergen), and for fruitful discussions with Hege Folkestad my colleague at the Science Library.

References

BAR-ILAN, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics, 74*(2): 257-271.

BAR-ILAN, J., LEVENE, M., & LIN, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics, 1*(1): 26-34.

BELEW, R. K. (2005). Scientific impact quantity and quality: Analysis of two sources of bibliographic data. *Journal*. Retrieved from http://arxiv.org/abs/cs/0504036

GARFIELD, E. (1990). How ISI selects journals for coverage - Quantitative and qualitative considerations. *Current Contents, 22*: 5-13.

GEOREF. THE AMERICAN GEOLOGICAL INSTITUTE. (No Date). About GeoRef. Retrieved July 11, 2008, from http://www.agiweb.org/georef/about/

GOOGLE SCHOLAR. (No Date). About Google Scholar. Retrieved July 11, 2008, from http://scholar.google.com/intl/en/scholar/about.html

HIRSCH, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*(46): 16569-16572.

ISI WOS. THOMSON SCIENTIFIC. (No Date). Web of Science. Retrieved July 11, 2008, from http://scientific.thomsonreuters.com/products/wos/

JACSÓ, P. (2005). As we may search? Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science, 89*(9): 1537-1547.

JACSÓ, P. (2006). Deflated, inflated and phantom citation counts. *Online Information Review, 30*(3): 297-309.

MEHO, L. I., & YANG, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology, 58*(13): 2105-2125.

NEUHAUS, C., NEUHAUS, E., ASHER, A., & WREDE, C. (2006). The Depth and Breadth of Google Scholar: An Empirical Study. *portal: Libraries and the Academy, 6*(2): 127-141.

NORUZI, A. (2005). Google Scholar: The new generation of citation indexes. *Libri, 55*: 170-180.

PAULY, D., & STERGIOU, K. I. (2005). Equivalence of results from two citation analyses: Thomson ISI's citation index and Google's scholar service. *Ethics in Science and Environmental Politics, 2005*: 33-35.

REDNER, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter, 4*(2): 131-134.

SCOPUS. ELSEVIER. (No Date). Scopus Info - Scopus in Detail - What does it cover? Retrieved July 11, 2008, from http://info.scopus.com/detail/what/

VANCLAY, J. K. (2007). On the robustness of the h-index. *Journal of the American Society for Information Science and Technology, 58*(10): 1547-1550.

WALTERS, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing and Management, 43*(4): 1121-1132.