Influence of omitted citations on the bibliometric statistics of the major Manufacturing journals

Fiorenzo Franceschini[1], Domenico Maisano[2] and Luca Mastrogiacomo[3]

[1]*fiorenzo.franceschini@polito.it* [2]*domenico.maisano@polito.it* [3]*luca.mastrogiacomo@polito.it*
Politecnico di Torino, DIGEP (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

Abstract

Bibliometrics is a relatively young and rapidly evolving discipline. Essential for this discipline are bibliometric databases and their information content concerning scientific publications and relevant citations. Databases are unfortunately affected by errors, whose main consequence is represented by omitted citations, i.e., citations that should be ascribed to a certain (cited) paper but, for some reason, are lost.

This paper studies the impact of omitted citations on the bibliometric statistics of the major Manufacturing journals. The methodology adopted is based on a recent automated algorithm – introduced in (Franceschini et al., Journal of the American Society for Information Science and Technology, 64(10): 2149-2156, 2013) – which is applied to the Web of Science (WoS) and Scopus database.

Two important results of this analysis are that: (i) on average, the omitted-citation rate (p) of WoS is slightly higher than that of Scopus; and (ii) for both databases, p values do not change drastically from journal to journal and tend to slightly decrease with respect to the issue year of citing papers.

Although it would seem that omitted citations do not represent a substantial problem, they may affect indicators based on citation statistics significantly. This paper analyses the effect of omitted citations on popular bibliometric indicators like the average citations per paper (*CPP*) and its most famous variant, i.e., the ISI Impact Factor, showing that journal classifications based on these indicators may lead to questionable discriminations.

Keywords: Manufacturing journal, Database error, Omitted-citation rate, Citation count, *CPP*, Journal ranking, Confidence interval, Impact Factor.

1. Introduction

Comparing scientific journals on the basis of their citation statistics is a very common operation in bibliometrics. The most popular bibliometric indicator is the ISI Impact Factor (*IF*), which is determined for the journals indexed by the Web of Science (WoS) database – and reported annually in the Journal Citation Report (JCR) by Thomson Reuters (2015). In recent years, other journal indicators have gained a certain importance and diffusion, such as the Source Normalized Impact per Paper (SNIP), the SCImago Journal Ranking (SJR), the Audience Factor, etc. (Falagas et al., 2008; Zitt, 2010; Moed, 2011).

The comparison between scientific journals is often limited to journals in the same discipline. Major stakeholders for these comparisons are:

- Librarians, when selecting the most suitable journals for possible subscriptions.
- Authors, when choosing the journal where to submit their contributions for publication. In this choice, a dominant role is played by the journal reputation, which may depend on several factors, e.g., circulation, prestige of the editorial board, rejection rate, expert-opinion surveys, and, of course, the average citation impact (Lowry et al., 2007).
- Members of (inter)national organizations for research assessment, when evaluating the bibliometric performance of individual scientists or research institutions (Hicks, 2009). Several research assessment exercises evaluate the impact of individual articles (content), and also that of the corresponding journals (container) (ERA, 2010; VQR, 2011). In spite of being questionable, these exercises have a certain diffusion and may have important implications, such as: (i) penalizing articles published by "weaker" and/or younger scientific journals, and (ii) encouraging authors to submit contributions to "dominant" journals.

For almost all bibliometric evaluations of scientific journals, a typical proxy for representing one article's impact on the scientific community is represented by the citations obtained according to a bibliometric database. Currently, the major multidisciplinary databases are Web of Science (WoS), Scopus and Google Scholar (GS); unfortunately, the level of inaccuracy of the latter database makes it still unreliable (Meho and Yang, 2007). Even though the literature contains numerous notifications of blunders (sometimes grotesque!) by GS (Labbé 2010), the errors by WoS and Scopus are almost always ignored. Precalculated citation statistics from the WoS and Scopus databases are often accepted blindly and used to make discriminations between journals. For example, some national research assessment exercises, such as the Australian (ERA, 2010) or the Italian (VQR, 2011), adopt(ed) classifications based on the average citation impact of journals, according to WoS and/or Scopus.

Bibliometric databases, like any database, are not free from errors. The impression of many authors is that the incidence of bibliometric database errors has been gradually declining over the past ten years, although a comprehensive study demonstrating this fact is still lacking. This is probably the effect of the systematic use – by editors and database administrators – of automatic tools for checking/correcting errors in the cited-article lists (Adam, 2002; Neuhaus and Daniel, 2008). Nevertheless, the problem is far from being solved, as proven by (i) several recent articles documenting the existence of different types of errors (Jacsó, 2012) and (ii) the fact that database staff constantly encourage users to report any noticed inaccuracy.

In the literature many authors analysed the presence of database errors. For example, a very popular work by Moed (2005) investigates discrepancies between cited references and cited papers, analyzing the citations received in the year 1999 by documents issued from 1980 to 1999, according to the WoS database. Discrepancies generally originate from different types of errors. A synthetic classification of the major errors is reported in Tab. 1, distinguishing between author and database mapping errors. The contributions by Buchanan (2006), Jacsó (2006), Li et al. (2010), Moed (2005) and Olensky (2013) show that one of the main consequences of these errors is represented by omitted citations, i.e., citations that should be ascribed to a certain (cited) paper but, for some reason, are lost. In other terms, the link between citing and cited article is not established by the database; Fig. 1 contains a schematic representation of the concept of omitted citation.

Tab. 1. Classification of bibliometric database errors according to Buchanan (2006).

Error type	Author errors	Database mapping errors
Definition	Errors made by authors when creating the list of cited articles for their publication.	Failure to establish an electronic link between a cited article and the corresponding citing articles that can be attributed to a data-entry error.
Examples	- Errors in name and initials of the first author, - Errors in publication title, - Errors in publication year, - Errors in volume number, - Errors in pagination.	- Transcription errors, - Target-source article record errors, - Cited article omitted from a cited-article list, - Reason unknown.

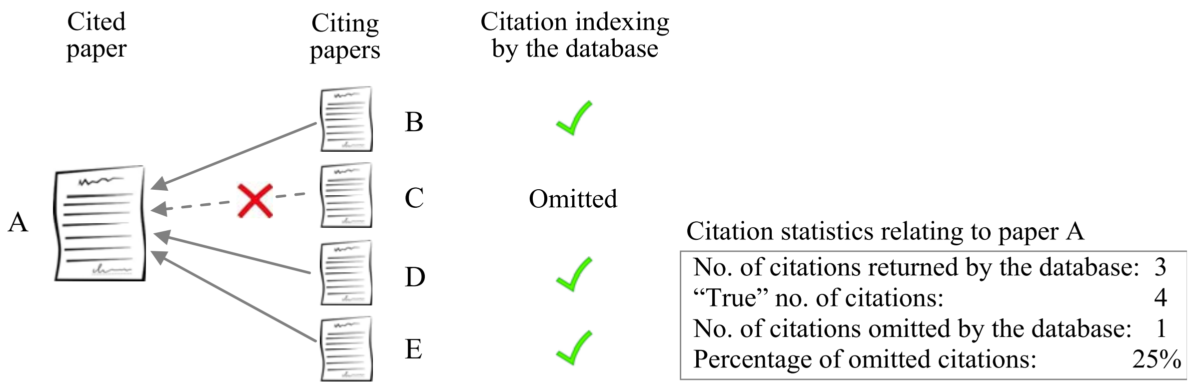

Fig. 1. Schematic representation of the concept of omitted citation.

According to the study by Buchanan (2006), which is based on a limited number of articles, the portion of citations omitted by WoS are likely to be around 5-10% of the “true” number of citations (i.e., the number of citations that would be indexed by the database, in the ideal case of absence of errors. This estimate is in line with the results shown in the aforementioned study by Moed (2005).

A recent article by Franceschini et al. (2013) introduced an automated algorithm for estimating a database’s omitted-citation rate, based on the combined use of two or more bibliometric databases. The basic logic is that the mismatch between the citations occurring in one database and one other is evidence of possible errors/omissions. A preliminary application of the automated algorithm to a set of three journals in the field of Bibliometrics showed that the omitted-citation rate is about 5.6%

for WoS and 3.2% for Scopus. Also, the aforementioned article contains a simplified statistical model that – for a set of papers of interest and a bibliometric database – makes it possible to estimate the "true" number of citations, with an appropriate confidence interval.

This paper presents an extensive application of the automated algorithm to the major scientific journals in the Manufacturing field, with four main aims:

- To estimate the omitted-citation rate of these journals, according to WoS and Scopus;
- To analyse the effect of two factors – i.e., (i) the journal of cited papers and (ii) the age of citing papers – on the omitted-citation rate of the journals examined;
- To estimate the distortions that omitted citations may cause on the average-citations-per-paper (*CPP*) indicator and its extremely popular and (mis)used variant, i.e., the journal *IF*;
- To estimate the "corrected" *CPP* of the journals examined, taking into account the omitted-citation rate of the database(s) in use. This study will also be extended to the estimation of the "corrected" journal *IF*.

The latter two points are particularly interesting since classic *CPP*-based journal rankings do not take database errors into account.

The rest of the paper is organized in four sections. Sect. 2 recalls the procedure for estimating the omitted-citation rate of databases and the statistical model for estimating the "true" number of citations of a set of papers of interest. Sect. 3 illustrates the data collection and the analysis methodology. Sect. 4 presents the analysis results. Sect. 5 summarizes the original contributions of the paper, highlighting the relevant implications, limitations and suggestions for future research. The appendix contains additional material for a deeper investigation.

2. Background Information

The two following subsections recall some basic principles concerning: (1) the automated algorithm for analysing the omitted citations, and (2) a statistical model for estimating the "true" number of citations.

2.1 Automated algorithm for analysing the omitted citations

Before recalling the algorithm, we present an introductory example to illustrate how it works. Let us consider a fictitious paper of interest indexed by WoS and Scopus. The number of citations received by this paper are twelve in WoS and thirteen in Scopus (see Tab. 2).

The union of the citations recorded by the two databases is a total of nineteen citations. Among the citing articles, only nine belong to sources (i.e., journals or conference proceedings) officially covered by both databases (highlighted in grey in Tab. 2). Focusing on these nine "theoretically overlapping" citing articles, one is omitted by WoS (but not by Scopus) and two are omitted by

Scopus (but not by WoS). Therefore, from the perspective of the paper of interest, a rough estimate of the omitted-citation rate is 1/9 ≈ 11.1% in WoS and 2/9 ≈ 22.2% in Scopus. The same reasoning can be extended to multiple papers of interest and more than two bibliometric databases.

Tab. 2. Citation statistics relating to a fictitious article, according to WoS and Scopus. The union of the citations recorded by the two databases (see the first column) is a total of nineteen citations. Among the citing articles, only nine belong to sources officially covered by both databases (highlighted in grey).

Citing Article No.	Citations in WoS	Citations in Scopus
1		✓
2	✓	
3	✓	✓
4	✓	✓
5	✓	✓
6	✓	Omitted
7	✓	
8	✓	✓
9	✓	✓
10		✓
11	✓	Omitted
12	✓	✓
13	✓	
14	Omitted	✓
15		✓
16		✓
17		✓
18	✓	
19		✓
Total	12	13

Let us now focus attention on the automated algorithm, which is based on the combined use of two bibliometric databases (WoS and Scopus) and can be summarised in three steps:

1. Identify a set of (P) papers of interest, indexed by both databases.
2. For each (i-th) paper of the set, identify the “theoretically overlapping” citing papers, defined as the portion of documents issued by journals officially covered by WoS and Scopus. The number of “theoretically overlapping” citing papers (or citations) concerning the i-th paper of interest will be denoted as γ_i.
3. For each database, determine the portion of “theoretically overlapping” citations that do not occur in it and classify them as omitted citations (ω_i). The omitted citation rate (p) relating to the set of papers of interest, according to a database, can be estimated as:

$$p = \sum_{i=1}^{P} \omega_i / \sum_{i=1}^{P} \gamma_i \; . \tag{1}$$

We emphasize that p is estimated on the basis of (i) a set of papers of interests and (ii) a portion of the total citations that they obtained (i.e., that ones related to citing articles purportedly covered by both databases). The extension of p to a wider set of papers represents a very delicate extrapolation, as p could be influenced by factors, such as: (i) journal particularities (even for journals within the same discipline), and (ii) age of the citing papers. These aspects will be investigated individually in

Sects. 4.1 and 4.2.

For a more detailed description of the algorithm, we refer the reader to (Franceschini et al., 2013).

2.2 Statistical model

Considering a relatively large set of scientific articles, with *C* total citations, and assuming that their omitted-citation rate is *p*, the “true” number of citations (C^*), and a relevant 95% confidence interval can be estimated using the approximate relationship:

$$C^* = \frac{C}{1-p} \pm 2 \cdot \sqrt{C \cdot p} \ . \tag{2}$$

For example, assuming that (i) a set of papers of interest obtained $C = 100$ total citations, according to a database, and (ii) the *p* value of the database for these papers is approximately 8%, then C^* and the corresponding 95% confidence interval will be 108.7 ± 5.7.

For a rigorous demonstration of Eq. 2, we refer the reader to (Franceschini et al., 2013). This formula is appropriate in the case *C* is large enough, so that $C \cdot p \geq 5$ (Ross, 2009). This condition is generally satisfied when considering the totality of the articles published by a journal in one or more years.

3. Methodology and data collection

The automated algorithm recalled in Sect. 2.1 was applied to a set of scientific journals in the Manufacturing field. We selected journals (i) included in the ISI Subject Category of *Engineering-Manufacturing* (by WoS[1]) and (ii) covered by Scopus. The fact that journals are covered by (at least) two databases is an essential requirement for applying the algorithm. The authors are aware that the resulting set of journals is not necessarily exhaustive, i.e., it may not include the totality of the journals concerning Manufacturing, such as: (i) journals not indexed by both WoS and Scopus or (ii) journals indexed by both databases but not included in the aforementioned ISI Subject Category. However, at least in a first approximation, this set of journals can be considered as representative of the entire Manufacturing field. The fact remains that the proposed analysis can be extended to any other journal (within or outside Manufacturing).

Journal titles and the corresponding abbreviations are reported in Tab. 3. For each journal, we selected the articles published in the time-window from 2006 to 2012, which are indexed by both databases. This time-window meshes together two partly opposing requirements: (i) articles should not be too recent, so that they have accumulated a certain amount of citations, and (ii) articles should not be too old, so that our analysis can bring out the current error propensity of databases.

[1] According to the 2011 JCR (Thomson Reuters, 2015).

Tab. 3. List of the journals examined. For each journal, it is reported its abbreviation ("Abbr."), title and ISSN code. Journals are sorted alphabetically according to their title.

Abbr.	Journal title	ISSN
J1	AI EDAM - Artificial Intelligence for Engineering Design Analysis and Manufacturing	0890-0604
J2	Assembly Automation	0144-5154
J3	CIRP Annals - Manufacturing Technology	0007-8506
J4	Composites Part A - Applied Science and Manufacturing	1359-835X
J5	Concurrent Engineering - Research and Applications	1063-293X
J6	Design Studies	0142-694X
J7	Flexible Services and Manufacturing Journal	1936-6582
J8	Human Factors and Ergonomics in Manufacturing & Service Industries	1090-8471
J9	IEEE Trasaction on Components Packaging and Manufacturing Technology	2156-3950
J10	IEEE Transactions on Semiconductor Manufacturing	0894-6507
J11	IEEE-ASME Transactions on Mechatronics	1083-4435
J12	International Journal of Advanced Manufacturing Technology	0268-3768
J13	International Journal of Computer Integrated Manufacturing	0951-192X
J14[1]	International Journal of Crashworthiness	1358-8265
J15[2]	International Journal of Design	1991-3761
J16[2]	International Journal of Industrial Engineering - (Theory) Applications and Practice	1072-4761
J17	International Journal of Machine Tools & Manufacture	0890-6955
J18[3]	International Journal of Production Economics	0925-5273
J19	International Journal of Production Research	0020-7543
J20	Journal of Advances Mechanical Design Systems and Manufacturing	1881-3054
J21	Journal of Computing and Information Science in Engineering - Transactions of the ASME	1530-9827
J22	Journal of Intelligent Manufacturing	0956-5515
J23	Journal of Manufacturing Science and Engineering - Transactions of the ASME	1087-1357
J24	Journal of Manufacturing Systems	0278-6125
J25	Journal of Materials Processing Technology	0924-0136
J26	Journal of Scheduling	1094-6136
J27	Machining Science and Technology	1091-0344
J28[4]	Manufacturing Engineering	0361-0853
J29	Materials and Manufacturing Processes	1042-6914
J30	Proceedings of the Institution of Mechanical Engineers Part B - Journal of Engineering Manufacture	0954-4054
J31	Packaging Technology and Science	0894-3214
J32	Precision Engineering - Journal of the International Societies for Precision Engineering and Nanotechnology	0141-6359
J33	Production and Operations Management	1059-1478
J34	Production Planning & Control	0953-7287
J35	Research in Engineering Design	0934-9839
J36	Robotics and Computer-Integrated Manufacturing	0736-5845
J37	Soldering & Surface Mount Technology	0954-0911

(1) this journal was indexed by the 2011 JCR, but no longer by the 2012 JCR, since it has been banned for boosting impact factor with self-citations (Thomson Reuters, 2015; Van Noorden, 2013). However, we included it in our analysis since the corresponding citation statistics were still available in WoS.

(2) these journals include articles whose DOI codes are reported by Scopus but not by WoS. For this reason, they were excluded from the analysis.

(3) this journal was indexed by the 2012 JCR, but no longer by the 2013 JCR, since it has recently been banned for boosting impact factor with "citation stacking" (Thomson Reuters, 2015). However, we included it in our analysis since the corresponding citation statistics were still available in WoS.

(4) this journal includes articles whose DOI codes are not reported by any of the databases in use. For this reason, it was excluded from the analysis.

We excluded articles without the DOI code or whose DOI code is not indexed by both databases, as they would be difficult to disambiguate. In fact, article disambiguation is performed by querying the two databases with the DOI code and/or full title of papers. Since databases rarely make mistakes in the DOI code indexing (see Franceschini et al., 2015), the use of DOI codes generally entails an accurate matching between papers contained in different databases.

We noticed that, at the moment of the analysis, DOI codes of most of the articles issued by J15 and

J16 were reported by Scopus but not by WoS. Also, DOI codes (if present) of the articles issued by J28 are not reported by any of the databases in use. Therefore, these three journals were excluded from the analysis. As regards the remaining thirty-four journals, only a few articles were excluded: mainly editorials, notes and articles on special issues with unindexed DOI code.

For each of the selected articles, we collected relevant data concerning their citing articles (i.e., issue year, article title, author(s), DOI code, journal title, etc.), from WoS and Scopus[2]. Data collection was carried out in June 2013. Data were used for estimating the *p* values of the databases in use and investigating the possible influence of two factors:

(i) *Journal of cited papers*, i.e., we will investigate possible differences between groups of (cited) articles issued by different Manufacturing journals;

(ii) *Age of citing papers*, i.e., we will investigate possible differences between groups of citing articles issued in different years.

Tab. A2, in the appendix, contains details about the number of articles selected for the analysis, depending on journal and issue year. Journals will be compared on the basis of their *CPP* values, calculated with the specific time-windows introduced in Sect. 4.4. Next, *CPP* values will be corrected taking into account the journal's omitted-citation rates. Eq. 3 will be used to estimate the CPP^* of each journal – i.e., the "corrected" *CPP* in the absence of omitted citations – with a relevant 95% confidence interval. This relationship is obtained by dividing both terms of Eq. 2 by *P* (i.e., the total number of papers considered for the Manufacturing journal):

$$CPP^* = \frac{CPP}{(1-p)} \pm 2 \cdot \sqrt{\frac{CPP \cdot p}{P}}, \tag{3}$$

being:

$CPP = C/P$ the average citations per paper of the journal of interest;

CPP^* the estimate of the "corrected" *CPP*, with a 95% confidence interval;

p is the omitted-citation rate of the journal of interest.

A model similar to that of Eq. 3 will be used to estimate the "corrected" journal *IF*, i.e., the most popular variant of *CPP*.

4. Results

Sects. 4.1 and 4.2 discuss the influence of the two aforementioned factors on the estimation of the omitted-citation rate. Sect. 4.3 estimates the omitted-citation rate of individual Manufacturing journals, according to the two databases in use. Sect. 4.4 compares the Manufacturing journals, based on their *CPP* and CPP^* values.

---

[2] The same portfolio of cited/citing papers was used in another work of ours – i.e., (Franceschini et al., 2014) – which demonstrates the link between omitted-citation rate and publishers (e.g., Elsevier, Springer, Taylor & Francis, etc.) of the citing papers.

4.1 Effect of the journal of cited papers

The *p* value of each journal is estimated by using the citations accumulated by the papers of interest, from the time of their publication until the end of 2012, both for WoS and Scopus. To make the study repeatable, we neglected the citations obtained in the year 2013, since it was not yet completed at the moment of the analysis (June 2013). Considering a journal (*J*) and a database of interest, the corresponding omitted-citation rate is defined as:

$$p_J = \sum_{i=1}^{P_J} (\omega_J)_i \Bigg/ \sum_{i=1}^{P_J} (\gamma_J)_i \, , \tag{4}$$

being:

P_J the number of articles of interest, issued by *J*;

$(\omega_J)_i$ the number of citations omitted by the database, concerning the *i*-th article issued by *J*;

$(\gamma_J)_i$ the number of "theoretically overlapping" citations, concerning the *i*-th article issued by *J*.

The resulting p_J values are reported in Tab. 4 and plotted in Fig. 2. We immediately notice that the omitted-citation rates according to WoS are larger than those according to Scopus, for almost all the journals. Moreover, the absence of a correlation between the p_J values of the two databases (very low $R^2 \approx 0.06$) is confirmed, as previously reported by Franceschini et al. (2013) (see Fig. 2).

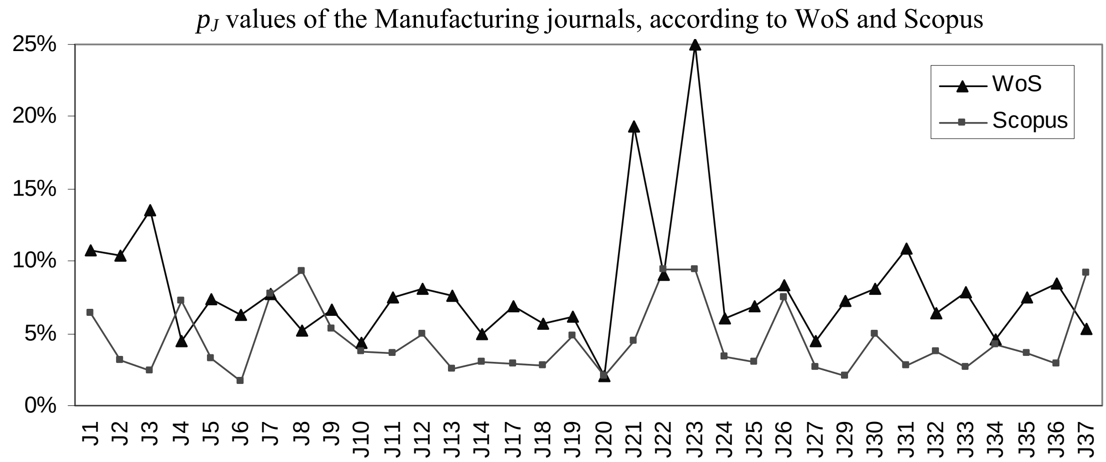

Fig. 2. p_J values of the journals examined, for the WoS and Scopus database respectively. Numerical data are reported in Tab. 4.

Focusing on the statistics related to a certain database, the p_J values of some journals appear very different from those of others; for example, as regards WoS, they switch from 2.1% for J20 to 19.3% for J21 or even 25.0% for J23. However, to understand whether these differences are statistically significant, it is necessary to evaluate the dispersion associated with these estimates. Sect. A.1.1 (in the appendix) illustrates a possible approach for this evaluation, based on the analysis of omitted citations, at the level of individual articles. As regards the distribution of omitted citations, the differences between Manufacturing journals seem insignificant for both WoS and Scopus and it cannot be stated that a database's propensity to omit citations varies significantly from one journal to one other.

Tab. 4. Omitted-citation rates and other statistics, for the scientific journals listed in Tab. 3. Articles were issued from 2006 to 2012 and citations were accumulated in the same period. Statistics are determined both for WoS and Scopus.

		(a) WoS		(b) Scopus	
Journ.	$\sum_{i=1}^{P_J}(\gamma_J)_i$	$\sum_{i=1}^{P_J}(\omega_J)_i$	p_J	$\sum_{i=1}^{P_J}(\omega_J)_i$	p_J
J1	374	40	10.7%	24	6.4%
J2	356	37	10.4%	11	3.1%
J3	5310	716	13.5%	126	2.4%
J4	13551	607	4.5%	990	7.3%
J5	366	27	7.4%	12	3.3%
J6	710	45	6.3%	12	1.7%
J7	52	4	7.7%	4	7.7%
J8	343	18	5.2%	32	9.3%
J9	76	5	6.6%	4	5.3%
J10	1075	47	4.4%	41	3.8%
J11	3496	262	7.5%	127	3.6%
J12	12886	1039	8.1%	647	5.0%
J13	1177	90	7.6%	30	2.5%
J14	840	41	4.9%	25	3.0%
J17	8809	607	6.9%	257	2.9%
J18	11807	677	5.7%	335	2.8%
J19	7640	468	6.1%	368	4.8%
J20	94	2	2.1%	2	2.1%
J21	574	111	19.3%	26	4.5%
J22	1539	138	9.0%	144	9.4%
J23	2364	591	25.0%	222	9.4%
J24	234	14	6.0%	8	3.4%
J25	23627	1626	6.9%	707	3.0%
J26	253	21	8.3%	19	7.5%
J27	418	19	4.5%	11	2.6%
J29	3406	248	7.3%	68	2.0%
J30	2140	174	8.1%	105	4.9%
J31	960	105	10.9%	27	2.8%
J32	1590	101	6.4%	61	3.8%
J33	793	62	7.8%	21	2.6%
J34	855	39	4.6%	36	4.2%
J35	387	29	7.5%	14	3.6%
J36	2168	183	8.4%	63	2.9%
J37	284	15	5.3%	26	9.2%

$\Sigma(\gamma_J)_i$ is the total number of "theoretically overlapping" citations;
$\Sigma(\omega_J)_i$ is the total number of omitted citations, according to a database;
p_J is the omitted-citation rate, according to a database.

Considering the WoS database, two "outlier journals" (i.e., J21 and J23) have an omitted-citation rate significantly higher than the others. A manual examination of the omitted citing-papers showed that, for both journals, database errors are mainly due to the relatively complicated journal title.

Specifically, the full name of J21 is “Journal of Computing and Information Science in Engineering - Transactions of the ASME” but, in the reference lists of citing papers, it is often replaced with several variants, sometimes ignored by WoS, such as, “J. Comput. Inf. Sci. Eng.” or “Journal of Computing and Information Science in Engineering”. The same applies to J23: the journal title is “Journal of Manufacturing Science and Engineering - Transactions of the ASME” but usual variants are “Transactions of the ASME - Journal of Manufacturing Science and Engineering”, “J. Manuf. Sci. E. - Trans. ASME” or “ASME J. Manuf. Sci. Eng.”.

4.2 Effect of the age of citing papers

To study the effect of the age of citing papers on the estimation of the omitted-citation rate, we propose a methodology articulated in the following steps:

- Identify the articles issued from 2006 to 2012 by the Manufacturing journals and classify them as articles of interest.
- Identify the articles issued from 2006 to 2012, which cite the articles of interest. These citing articles are then divided according to their issue year (i.e., from 2006 to 2012). To make the study repeatable, the citing articles issued during 2013 were neglected, as this year was not yet completed at the moment of the analysis.
- For each year (*Y*), calculate the omitted-citation rate (p_Y) of a database of interest as:

$$p_Y = \sum_{i=1}^{P} (\omega_Y)_i \Big/ \sum_{i=1}^{P} (\gamma_Y)_i \, , \tag{5}$$

being:

$(\omega_Y)_i$ the number of citations related to the *i*-th article of interest, which are obtained in the year *Y* and omitted by the database;

$(\gamma_Y)_i$ the number of “theoretically overlapping” citations concerning the *i*-th article of interest, obtained in the year *Y*;

P the total number of articles of interest, issued by the Manufacturing journals in the period from 2006 to 2012.

Tab. 5 contains the p_Y values related to WoS and Scopus. The fact that the $\sum(\gamma_Y)_i$ values (in the second column of Tab. 5) gradually increase with the issue year is not surprising; this originates from the use of a fixed time-window (i.e., from 2006 to 2012) for cited articles and a variable time-window (i.e., from the issue year to 2012) for citing articles. Since the articles of interest were issued during the period from 2006 to 2012, the number of relevant citing articles, issued in the same period, will of course tend to increase over time.

The choice of these time-windows does not undermine the representativeness of our analysis, since the $\sum(\gamma_Y)_i$ values are generally large (it can be seen that they generally are of the order of magnitude of 10^3 or even 10^4).

An interesting result is that, for both databases, the omitted-citation rate tends to decrease over time; this result is even more remarkable, as we considered seven consecutive years only (i.e., from 2006 to 2012). Probably, this is an effect of the increasing care of editors and database administrators in checking and correcting errors in the cited-article lists. The graph in Fig. 3 shows that this reduction is particularly clear in the early years (e.g., according to WoS, p_Y switches from 18.6% in 2006 to 14.9% in 2007 and even 9.9% in 2008) but then tends to "saturate" (according to WoS, there is a slight increase of p_Y from 2011 to 2012). This result foretells that databases will hardly be able to get rid of omitted citations in the years to come. Sect. A.1.2 (in the appendix) goes through this point, showing that the reduction in the p_Y values, although present, is not very significant from a statistical viewpoint.

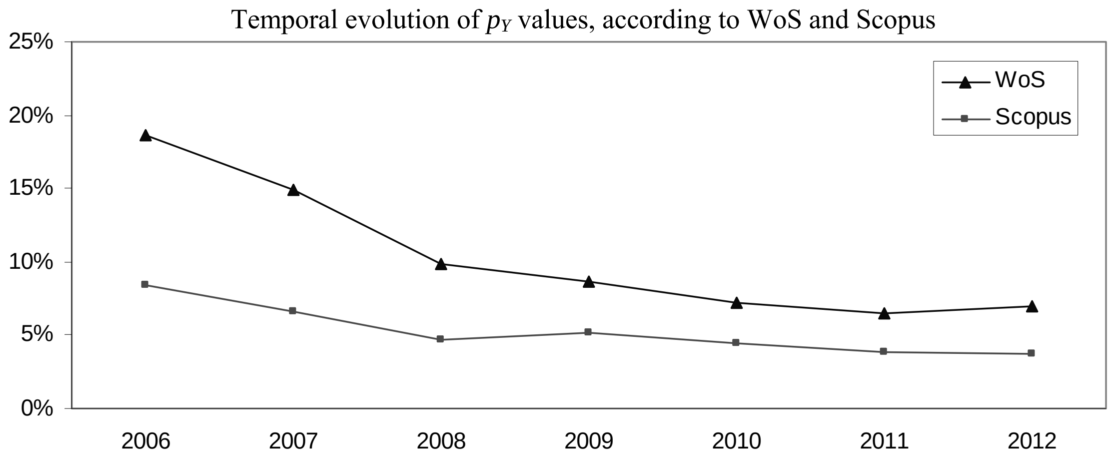

Fig. 3. p_Y values depending on the issue year of the citing papers examined, according to the WoS and Scopus database respectively. Numerical data are reported in Tab. 5.

Tab. 5. Omitted-citation rates related to the Manufacturing journals, depending on the issue year of the citing papers.

Year	$\sum_{i=1}^{P}(\gamma_Y)_i$	(a) WoS $\sum_{i=1}^{P}(\omega_Y)_i$	(a) WoS p_Y	(b) Scopus $\sum_{i=1}^{P}(\omega_Y)_i$	(b) Scopus p_Y
2006	274	51	18.61%	23	8.39%
2007	1927	287	14.89%	128	6.64%
2008	6221	616	9.90%	289	4.65%
2009	13221	1148	8.68%	677	5.12%
2010	20759	1492	7.19%	920	4.43%
2011	30047	1947	6.48%	1160	3.86%
2012	38099	2667	7.00%	1408	3.70%

$\Sigma(\gamma_Y)_i$ is the total number of "theoretically overlapping" citations;

$\Sigma(\omega_Y)_i$ is the total number of omitted citations, according to a database;

p_Y is the omitted-citation rate according to a database; P is the total number of articles of interest, issued by the Manufacturing journals examined in the period from 2006 to 2012.

4.3 Overall estimate of p for the totality of Manufacturing journals

An overall omitted-citation rate for the totality of the Manufacturing journals examined, irrespective of the issue year of the citing papers, can be estimated using the following formula:

$$p = \left(\sum_{J \in S} \sum_{i=1}^{P_J} (\omega_J)_i \right) \Bigg/ \left(\sum_{J \in S} \sum_{i=1}^{P_J} (\gamma_J)_i \right), \tag{6}$$

being

P_J the number of papers of interest, which were issued by the journal J in the period from 2006 to 2012;

$(\omega_J)_i$ the number of citations omitted by the database in use, from the perspective of the i-th paper issued by the journal J, considering the citing papers issued in the period from 2006 to 2012;

$(\gamma_J)_i$ the number of "theoretically overlapping" citations, from the perspective of the i-th paper of the journal J, considering the citing papers issued in the period from 2006 to 2012.

S the set of journals examined. i.e., those listed in Tab. 3 except J15, J16 and J28.

The calculation was performed for both WoS and Scopus; results are reported in Tab. 6.

Tab. 6. Overall omitted-citation rates for the Manufacturing journals examined, regarding WoS and Scopus.

	(a) WoS		(b) Scopus	
$\sum\sum(\gamma_J)_i$	$\sum\sum(\omega_J)_i$	p	$\sum\sum(\omega_J)_i$	p
110548	8208	7.4%	4605	4.2%

$\Sigma\Sigma(\gamma_J)_i$ is the total number of "theoretically overlapping" citations;
$\Sigma\Sigma(\omega_J)_i$ is the total number of omitted citations, according to a database;
p is the omitted-citation rate, according to a database.

It can be noticed that the p of WoS is greater than that of Scopus: 7.4% against 4.2%. The p value of WoS is in line with the results of other studies concerning omitted citations, such as those by Buchanan (2006) and Moed (2005).

4.4 Comparison of journals based on the "corrected" CPP and IF values

The omitted-citation rate estimates can be used to correct journal indicators based on citation statistics. Our initial intention was to focus on the *IF* of each Manufacturing journal, owing to the great popularity and diffusion of this indicator. We recall the definition of the (two-year) *IF* of a journal, related to a certain year *Y*: *the number of citations of articles published in years (Y–1) and (Y–2) in the journal, which appeared in articles published in year Y, divided by the number of "citable documents" published in the journal in years (Y–1) and (Y–2)* (Thomson Reuters, 2015).

As highlighted in the literature (see for example (Bar-Ilan, 2010)), the *IF* definition is somewhat questionable. For example, it is not perfectly clear what the "citable documents" are, although they generally seem to include research articles, proceedings papers, reviews and letters, and exclude editorials, prefaces, corrections or other documents that, nevertheless, can contribute to citation accumulation. Unfortunately, data concerning the total number of citable documents, used for calculating the *IF* values, are not available on the JCRs (Thomson Reuters, 2015). According to some authors, this would be a form of protection adopted by Thomson Reuters to hide possible inaccuracies in the citation count or in the designation of the citable documents (Rossner et al., 2008). Actually, this lack of data makes the *IF* irreproducible, even for who can access the WoS database. Another obvious limitation of the *IF* is that it is constructed using database records from WoS exclusively.

To allow the comparison of statistics from both WoS and Scopus and to avoid the ambiguity concerning the designation of the citable-documents, we defined a "modified *IF*" where the totality of the articles published by a journal and indexed by the database in use are classified as "citable documents". This indicator was generically denominated as *CCP* and calculated for each year (*Y*) of the five consecutive years from 2008 to 2012. Tab. A3 (in the appendix) reports the resulting journal *CPP* values and other statistics, both for WoS and Scopus. The time-windows used for identifying the documents of interest and counting the relevant citations are those specified in the *IF* definition. For the purpose of example, Tab. 7 reports the statistics relating to the year 2012.

It is interesting to examine the variations in *CPP*, when calculated referring to the WoS or the Scopus database. Rank-reversals in the resulting journal ranking are not so rare. For the purpose of example, let us consider J5, J7 and J23: according to the citation statistics by WoS, the relative ranking in the year 2012 is $\mathrm{J7}(0.594) \succ \mathrm{J5}(0.438) \succ \mathrm{J23}(0.379)$, while according to Scopus, the ranking is completely subverted, i.e., $\mathrm{J23}(0.716) \succ \mathrm{J5}(0.652) \succ \mathrm{J7}(0.550)$, where symbol "$\succ$" means "preferred to". This is a first sign of the instability of journal rankings based on *CPP* or its variants – such as the *IF*, with its false impression of precision conveyed by the three decimal points! (Franceschini and Maisano, 2011). Of course, this difference depends on the fact that the two databases have a different coverage and – even when considering the same set of articles – citing papers can be different.

Next, *CPP* values were corrected using the model in Eq. 3. Given a journal (*J*) and a reference year (*Y*), the *p* value was estimated considering the omitted citations of papers issued in the year *Y*, which cite articles published in the years *Y*–1 and *Y*–2. For example, the *p* used for correcting the *CPP* of a journal in the year 2010 is calculated considering the articles issued by that journal in 2008 and 2009, and the citations obtained in 2010. Tab. A3 (in the appendix) reports the corresponding *p* values, CPP^* values and 95% confidence intervals. We note that, for some journals and years, *p* values are not estimated (e.g., see J1, J5 and J7 in 2012). The reason is that the sample of “theoretically overlapping” citations in use was too small (i.e., $\sum_{i=1}^{P_J} \gamma_{J_i} < 30$, see Tab. 7), then not suitable for a robust estimate.

Tab. 7. *CPP*, *CPP** and relevant statistics for the journals examined, relating to the year 2012. Indicators are calculated both for WoS (a) and Scopus (b).

		(a) WoS						(b) Scopus					
Journ.	$\sum_{i=1}^{P_J} \gamma_{J_i}$	*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	*CPP**	95% *CI* limits		*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	*CPP**	95% *CI* limits	
J1	18	0.302	0	-	-	-	-	0.679	0	-	-	-	-
J2	34	0.388	5	14.7%	0.455	0.403	0.507	0.844	0	0.0%	0.844	0.562	0.656
J3	567	1.953	41	7.2%	2.105	2.063	2.147	2.546	18	3.2%	2.629	0.562	0.656
J4	1154	2.386	46	4.0%	2.485	2.456	2.514	3.031	84	7.3%	3.269	0.562	0.656
J5	22	0.438	0	-	-	-	-	0.625	0	-	-	-	-
J6	60	1.017	4	6.7%	1.090	1.022	1.158	1.828	0	0.0%	1.828	0.562	0.656
J7	17	0.594	1	-	-	-	-	0.550	3	-	-	-	-
J8	39	0.467	2	5.1%	0.492	0.459	0.525	0.770	5	12.8%	0.884	0.562	0.656
J9	52	0.599	5	9.6%	0.663	0.631	0.695	1.259	3	5.8%	1.336	0.562	0.656
J10	79	0.719	1	1.3%	0.728	0.711	0.746	1.116	5	6.3%	1.191	0.562	0.656
J11	543	2.171	49	9.0%	2.386	2.328	2.445	3.836	33	6.1%	4.084	0.562	0.656
J12	1431	0.959	123	8.6%	1.049	1.034	1.065	1.596	51	3.6%	1.654	0.562	0.656
J13	138	0.764	12	8.7%	0.836	0.796	0.876	1.200	3	2.2%	1.227	0.562	0.656
J14	94	0.800	4	4.3%	0.836	0.800	0.871	1.000	3	3.2%	1.033	0.562	0.656
J17	463	2.138	25	5.4%	2.260	2.215	2.306	3.009	8	1.7%	3.062	0.562	0.656
J18	1003	1.760	44	4.4%	1.841	1.817	1.864	2.853	36	3.6%	2.959	0.562	0.656
J19	997	1.265	75	7.5%	1.368	1.345	1.390	1.771	50	5.0%	1.864	0.562	0.656
J20	63	0.427	2	3.2%	0.441	0.422	0.459	0.470	1	1.6%	0.478	0.562	0.656
J21	35	0.247	8	22.9%	0.320	0.270	0.371	0.557	3	8.6%	0.609	0.562	0.656
J22	145	0.759	19	13.1%	0.874	0.824	0.924	0.975	24	16.6%	1.169	0.562	0.656
J23	146	0.379	58	39.7%	0.630	0.581	0.678	0.686	27	18.5%	0.842	0.562	0.656
J24	40	0.822	4	10.0%	0.914	0.828	0.999	0.216	2	5.0%	0.228	0.562	0.656
J25	915	1.729	59	6.4%	1.848	1.819	1.877	2.340	22	2.4%	2.398	0.562	0.656
J26	58	0.679	6	10.3%	0.757	0.698	0.816	0.938	3	5.2%	0.989	0.562	0.656
J27	41	0.740	2	4.9%	0.778	0.724	0.832	1.020	1	2.4%	1.046	0.562	0.656
J29	525	1.170	38	7.2%	1.261	1.233	1.289	1.369	4	0.8%	1.379	0.562	0.656
J30	218	0.623	24	11.0%	0.700	0.673	0.728	0.817	2	0.9%	0.825	0.562	0.656
J31	46	0.571	9	19.6%	0.710	0.634	0.787	0.790	2	4.3%	0.826	0.562	0.656
J32	222	1.273	15	6.8%	1.365	1.321	1.409	1.852	22	9.9%	2.056	0.562	0.656
J33	36	1.188	2	5.6%	1.257	1.209	1.306	0.150	3	8.3%	0.163	0.562	0.656
J34	55	0.487	2	3.6%	0.506	0.481	0.530	0.941	6	10.9%	1.056	0.562	0.656
J35	48	1.051	2	4.2%	1.097	1.030	1.164	1.500	0	0.0%	1.500	0.562	0.656
J36	197	0.985	13	6.6%	1.055	1.019	1.091	1.950	2	1.0%	1.970	0.562	0.656
J37	31	0.735	0	0.0%	0.735	0.735	0.735	1.104	1	3.2%	1.141	0.562	0.656

$\sum \gamma_{J_i}$ is the total number of “theoretically overlapping” citations;

CPP is the average citations per papers;

$\sum \omega_{J_i}$ is the total number of omitted citations;

p is the estimated omitted-citation rate of a journal; *p* is not estimated in the case $\sum \gamma_{J_i} \leq 30$.

CPP* is the corrected *CPP* (with the corresponding 95% confidence interval limits in the two columns to the right). *CPP** is not calculated in the case $\sum \gamma_{J_i} < 30$ (i.e., for J1, J5 and J7).

Not surprisingly, *CPP** values are always greater than *CPP* values, since they compensate for omitted citations. Considering the 95% confidence intervals, the journal ranking takes on a different connotation. For some journals, confidence intervals are noticeably overlapped, indicating that the differences in terms of CPP^* are statistically insignificant[3]; for the purpose of example, see the graphical representations in Fig. 4 and Fig. 5, relating to the WoS statistics for the year 2012. This is another evidence of the risk of discriminating scientific journals, when using indicators like the *CPP* and its variants (Franceschini and Maisano, 2011); this applies to Scopus and especially to WoS, because of the higher omitted-citation rate.

Let us make the apparently reasonable exercise of dividing journals into four classes (i.e., A, B, C and D), using the quartiles of the distribution of their *CPP* values. Class A would include about 25% of the journals with the highest *CPP*, class B the second 25%, and so on for classes C and D.

As Fig. 4 and Fig. 5 show, the way journals are divided in quartiles is not always compatible with their *CPP* confidence intervals. For example, journals may move from one class to one other, depending on the database in use (e.g., see J6), and confidence intervals of two journals with different classes are often overlapped! This means that it is not appropriate to discriminate journals by their crude *CPP* values.

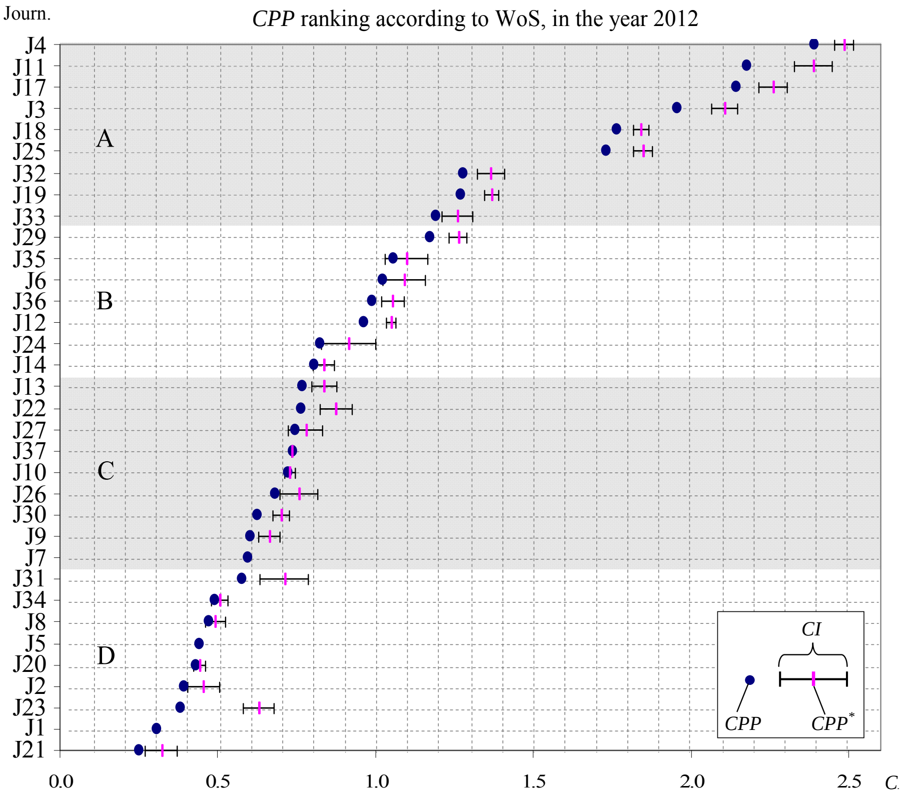

Fig. 4. Journal ranking according to the *CPP* values in the year 2012. Indicators were determined according to the citation statistics returned by WoS. For each journal, the *CPP value and the relevant 95% confidence interval (*CI*) are also represented; numerical values are reported in Tab. 7(a). Journals are grouped into four classes (A, B, C and D), using the quartiles of the distribution of their *CPP* values.**

[3] Authors are aware that a more rigorous testing should be that of the differences between CPP^* values of pairs of journals (Schenker and Gentleman, 2001). The fact remains that the qualitative approach in use is simpler and more straightforward.

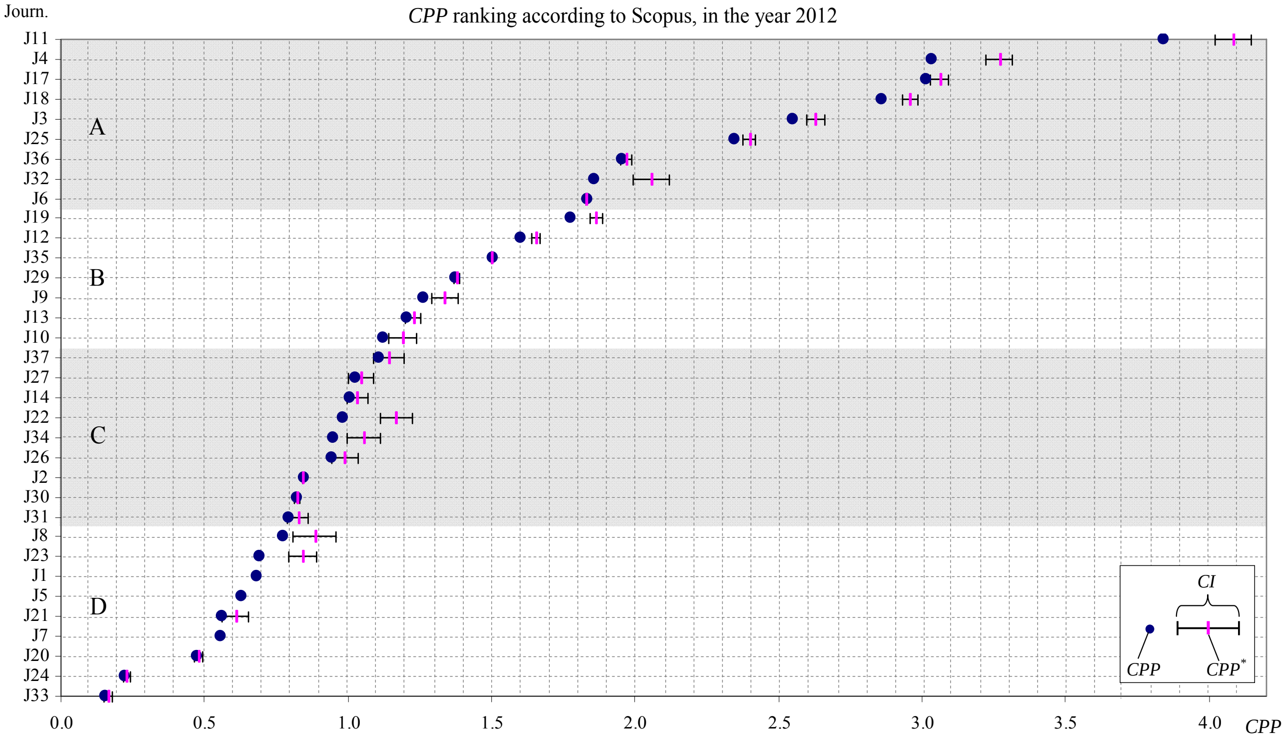

Fig. 5. Journal ranking according to the *CPP* values in the year 2012. Indicators were determined according to the citation statistics returned by Scopus. For each journal, the CPP^* value and the relevant 95% confidence interval (*CI*) are also represented; numerical values are reported in Tab. 7(b). Journals are grouped into four classes (A, B, C and D), using the quartiles of the distribution of their *CPP* values.

Let us now focus our attention on the authentic journal *IF*. The preceding analysis of one journal's *CPP* can be used for providing a rough estimation of the "corrected" *IF* value, which takes omitted citations into account. To this purpose, the model in Eq. 3 can be modified as:

$$IF^* = \frac{IF}{(1-p)} \pm 2 \cdot \sqrt{\frac{IF \cdot p}{P_{cit.}}}, \qquad (7)$$

being:

IF and IF^* respectively the journal impact factor before and after the correction;

p the omitted-citation rate determined for correcting *CPP* values (in Tab. 7(a) and Tab. A3(a)), according to WoS;

$P_{cit.}$ the (unknown) number of citable documents.

For the purpose of example, let us correct the *IF* of the Manufacturing journals, considering the year 2012. The original *IF* values were collected from the 2012 JCR (see the second column of Tab. 8). Interestingly, *IF* values are systematically larger than the corresponding *CPP* values related to WoS (in Tab. 7(a)). These deviations are due to the fact that the number of citable documents is, by definition, always lower than or equal to the number of articles which can contribute to citation accumulation. Moreover, the *IF* journal ranking often subverts that based on *CPP*; for instance, the top three journals according to the *CPP* are respectively J4, J11 and J17, while, according to the *IF*, J14, J11 and J4. This is a further distortion due to the introduction of the so-called citable documents (Rossner et al., 2008).

The (unknown) $P_{cit.}$ values were estimated by querying WoS and counting research articles, reviews, proceedings, letters and notes issued by each journal in the years 2010 and 2011. Other document types – such as bibliographies, editorials, book reviews, biographical items, corrections, etc. – were excluded. The authors are aware of the ambiguity concerning the designation of the citable documents (Rossner et al. 2008); however, "dubious" documents – such as notes – were deliberately included in order to avoid underestimating the $P_{cit.}$ and, consequently, overestimating the resulting 95% confidence interval around it (see $P_{cit.}$ in the denominator of the second term on the right-hand side of Eq. 7).

Resulting confidence intervals are shown in Tab. 8 and represented graphically in Fig. 6. There are various overlappings, even for journals in different *IF* quartile classes (A, B, C and D). This is a further proof that this kind of classification is questionable and may lead to unjustified journal discriminations. Unfortunately, some national research assessment exercises adopted and/or keep adopting them (ERA, 2010; VQR, 2011; Arnold and Fowler, 2011; DORA, 2013).

Tab. 8. *IF*, *IF and relevant statistics for the journals examined, relating to the year 2012.**

Journ.	*IF*	*p*	IF^*	$P_{cit.}$	95% *CI* limits	
J1[1]	0.407	-	-	54	-	-
J2	0.603	14.7%	0.707	78	0.640	0.774
J3	2.251	7.2%	2.426	315	2.381	2.472
J4	2.744	4.0%	2.858	453	2.827	2.889
J5[1]	0.542	-	-	48	-	-
J6	1.545	6.7%	1.655	55	1.569	1.742
J7[1]	0.857	-	-	28	-	-
J8	0.624	5.1%	0.658	85	0.619	0.697
J9	1.261	9.6%	1.395	219	1.348	1.442
J10	0.862	1.3%	0.873	116	0.854	0.892
J11	3.135	9.0%	3.446	222	3.375	3.517
J12	1.205	8.6%	1.318	1269	1.300	1.336
J13	0.944	8.7%	1.034	162	0.989	1.079
J14[2]	-	4.3%	-	-	-	-
J17	2.262	5.4%	2.391	221	2.344	2.438
J18	2.081	4.4%	2.176	541	2.150	2.202
J19	1.460	7.5%	1.579	724	1.554	1.603
J20	0.494	3.2%	0.510	154	0.490	0.530
J21	0.488	22.9%	0.633	86	0.561	0.705
J22	1.278	13.1%	1.471	151	1.404	1.537
J23	0.786	39.7%	1.304	248	1.233	1.375
J24	1.070	10.0%	1.189	43	1.089	1.289
J25	1.953	6.4%	2.088	515	2.056	2.119
J26	0.941	10.3%	1.050	75	0.978	1.122
J27	0.840	4.9%	0.883	50	0.826	0.940
J29	1.297	7.2%	1.398	421	1.368	1.428
J30	0.770	11.0%	0.865	361	0.835	0.896
J31	0.737	19.6%	0.916	76	0.829	1.003
J32	1.393	6.8%	1.494	173	1.447	1.541
J33	1.315	5.6%	1.392	111	1.341	1.444
J34	0.600	3.6%	0.623	105	0.594	0.651
J35	1.562	4.2%	1.630	35	1.544	1.716
J36	1.230	6.6%	1.317	196	1.276	1.358
J37	0.816	0.0%	0.816	49	0.816	0.816

IF is the journal impact factor according to the 2012 JCR (Thomson Reuters 2015);
p is the omitted-citation rate determined in Tab. 7(a);
IF^* is the corrected *IF* (with the corresponding 95% confidence interval limits in the two columns to the right);
P_{cit} is the estimated number of "citable documents";
(1) the IF^* of these journals was not calculated since the corresponding *p* values were not available (see Tab. 7(a));
(2) this journal is not included in the 2012 JCR, since it was banned for boosting impact factor with self-citations (van Noorden, 2013).

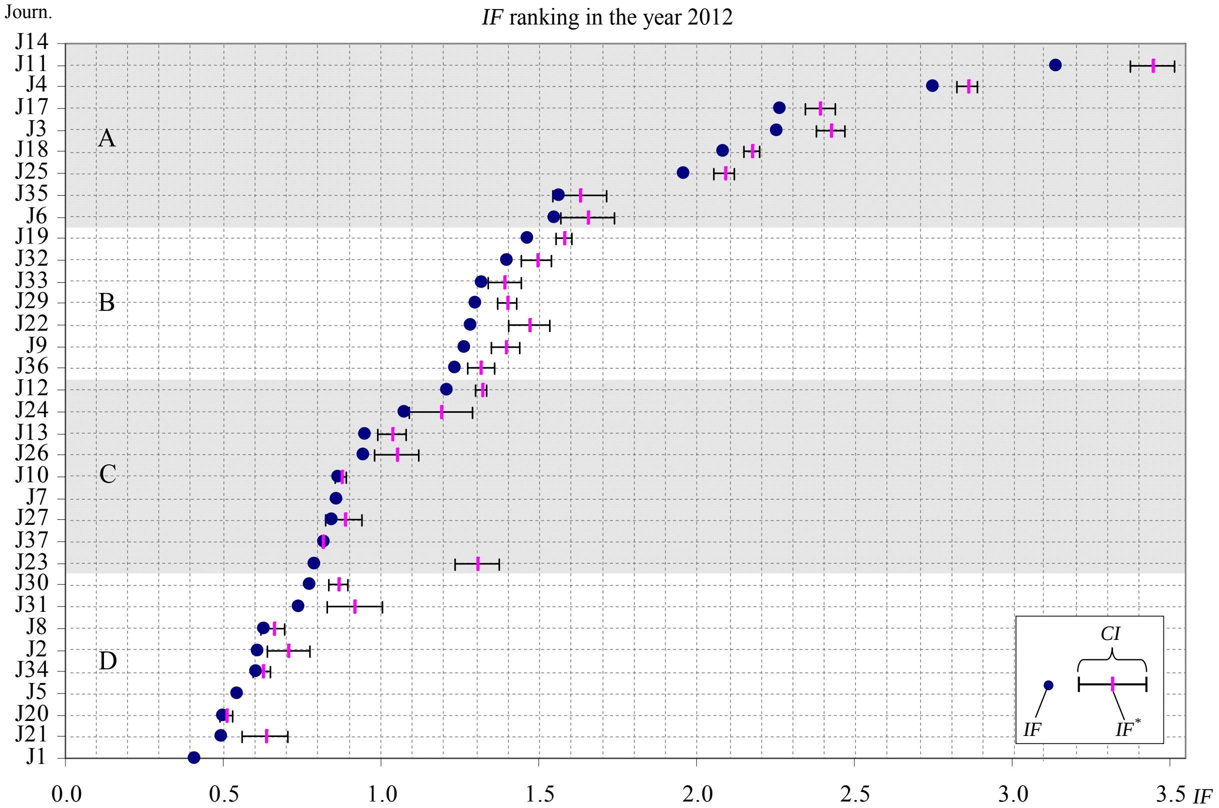

Fig. 6. Journal ranking according to the *IF* values in the year 2012. For each journal, the IF^* value and the relevant 95% confidence interval (*CI*) are also represented; numerical values are reported in Tab. 8. Journals are grouped into four classes (A, B, C and D), using the quartiles of the distribution of their *IF* values.

5. Conclusions

This paper studied the omitted-citation rate of two bibliometric databases (i.e., WoS and Scopus), referring to the articles published by thirty-four Manufacturing journals, in the period from 2006 to 2012.

With a few exceptions – the differences in terms of p among the Manufacturing journals examined are included between 4% and 10% for WoS and between 2% and 8% for Scopus. A practical justification is that, being these journals in the same discipline, the population of the citing papers tends to be quite uniform: in fact, it generally consists of papers from various sources, such as journals or conference proceedings in the Engineering field or neighbouring disciplines, like Materials Science, Physics, etc..

Also, it was noted that the p values for WoS are generally higher than those for Scopus. This result is in line with the output of the preliminary investigation by Franceschini et al. (2013).

In light of the above results, omitted citations can constitute a substantial problem: it was shown that they may significantly affect indicators based on citation statistics, such as *CPP* or *IF*, which may be severely underestimated. For example, the corrected *IF* (i.e., IF^*) related to J23 in 2012 is larger than the 65% of the *IF* (i.e., $IF = 0.786$ while $IF^* = 1.304$, see Tab. 8). Our caveat is therefore to "handle with care" journal classifications based on the *IF* or similar indicators, as they may lead to questionable discriminations; in fact, the problem of omitted citations is subtle and tricky and is often neglected, even by bibliometricians.

It is curious that a pioneering article by Schubert and Glänzel (1983) – although based on a very different research approach – comes to the same conclusions that *CPP* values are not adequate for ranking journals.

Another interesting finding is that *p* values tend to gradually decrease with the issue year of citing papers. This is probably due to the growing attention of editors and database administrators in checking/correcting errors in cited-article lists, which facilitates database indexing and error reduction.

The proposed method has the great advantage of being automated, i.e., it does not require manual analysis of cited/citing papers. The price to pay for this advantage is the possibility of small distortions in the results, as discussed in (Franceschini et al., 2013). Another limitation is that our analysis focussed on (i) journals confined within the area of Manufacturing and (ii) articles issued in the period from 2006 to 2012.

A starting point for future research could be extending the analysis to journals in other scientific fields. Also, the ability of databases to correct errors in previously indexed records can be investigated by repeating this study (i.e., using the same portfolio of cited articles and the same time-windows for citation count) in multiple different-time sessions (e.g., after 6 months, 1 year, 2 years, etc.). Finally, it would be interesting to identify appropriate analytical models for representing the distribution of omitted citations, which can be seen as relatively rare events.

References

Adam, D. (2002). Citation analysis: the counting house. Nature, 415(6873), 726-729.

Arnold, D.N., Fowler, K.K. (2011). Nefarious Numbers, Notices Amer. Math. Soc., 58(3): 434-437.

Bar-Ilan, J. (2010). Ranking of information and library science journals by JIF and by h-type indices. Journal of Informetrics, 4(2): 141-147.

Buchanan, R.A. (2006). Accuracy of Cited References: The Role of Citation Databases. College & Research Libraries, 67(4), 292-303.

DORA (2013). San Francisco Declaration on Research Assessment, http://am.ascb.org/dora/ [20 May 2014].

ERA (2010). Excellence in Research for Australia Initiative. http://www.arc.gov.au/era/era_2010/era_2010.htm [20 May 2014].

Falagas, M.E., Kouranos, V.D., Arencibia-Jorge, R., Karageorgopoulos, D.E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. The FASEB Journal, 22(8): 2623-2628.

Franceschini, F., Maisano, D. (2011) Influence of database mistakes on journal citation analysis: remarks on the paper by Franceschini and Maisano, QREI (2010). Quality and Reliability Engineering International, 27(7), 969-976.

Franceschini, F., D. Maisano and L. Mastrogiacomo (2013). A novel approach for estimating the omitted-citation rate of bibliometric databases. Journal of the American Society for Information Science and Technology, 64(10): 2149-2156.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2014). Scientific journal publishers and omitted citations in bibliometric databases: Any relationship? Journal of Informetrics, 8(3), 751-765.

Franceschini, F., Maisano, D., Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. To appear in Scientometrics, DOI: 10.1007/s11192-014-1503-4.

Hicks, D. (2009). Evolving regimes of multi-university research evaluation. Higher Education, 57: 393-404.
Jacsó, P. (2006). Deflated, inflated and phantom citation counts. Online Information Review, 30(3): 297-309.
Jacsó, P. (2012). Grim Tales about the impact factor and the *h*-index in the Web of Science and the Journal Citation Reports databases: Reflections on Vanclay's criticism. Scientometrics, 92(2), 325-354.
Li, J., Burnham, J.F., Lemley, T., Britton, R.M. (2010). Citation analysis: comparison of Web of Science, Scopus, Scifinder, and Google Scholar. Journal of Electronic Resources in Medical Libraries 7(3), 196-217.
Lowry, P.M., Humpherys, S.L., Malwitz, J., Nix, J. (2007). A Scientometric Study of the Perceived Quality of Business and Technical Communication Journals. IEEE Transactions on Professional Communication. 50(4): 352-378.
Labbé, C. (2010). Ike Antkare, one of the great stars in the scientific firmament. ISSI Newsletter, 6(2): 48-52.
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.
Moed, H.F. (2005). Citation analysis in research evaluation. Information Sciences and knowledge Management: Vol. 9. Dordrecht: Springer. http://dx.doi.org/10.1007/1-4020-3714-7. ISBN: 978-1-4020-3713-9.
Moed, H.F. (2011) The Source-Normalized Impact per Paper (SNIP) is a valid and sophisticated indicator of journal citation impact. Journal of the American Society for Information Science and Technology, 62(1): 211-213.
Neuhaus, C., Daniel, H.D. (2008) Data sources for performing citation analysis: an overview. Journal of Documentation, 64(2), 193-210.
Olensky, M. (2013) Accuracy Assessment for Bibliographic Data. Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI), vol. 2, pp. 1850-1851, Vienna, Austria.
Ross, S.M. (2009). Introduction to probability and statistics for engineers and scientists. Academic Press.
Rossner, M., Van Epps, H., Hill, E. (2008) Irreproducible results—A response to Thomson Scientific. The Journal of general physiology, 131(2), 183-184.
Schenker, N., Gentleman, J.F. (2001) On judging the significance of differences by examining the overlap between confidence intervals. The American Statistician, 55(3): 182-186.
Schubert, A., Glänzel, W. (1983) Statistical reliability of comparisons based on the citation impact of scientific publications. Scientometrics, 5(1), 59-74.
Scopus Elsevier (2015). Scopus Content Coverage. Available at http://www.scopus.com [20 May 2014].
Thomson Reuters (2015) http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/ [20 May 2014].
Van Noorden, R. (2013) New record: 66 journals banned for boosting impact factor with self-citations. Nature News Blog, http://blogs.nature.com/news/2013/06/new-record-66-journals-banned-for-boosting-impact-factor-with-self-citations.html [20 May 2014].
VQR (2011). Italian Quality research Evaluation VQR 2004–2010. http://www.anvur.org/anvur/ [20 May 2014].
Zitt, M. (2010). Citing-side normalization of journal impact: A robust variant of the Audience Factor. Journal of Informetrics, 4(3): 392-406.

Appendix

A.1 Analysis of the distribution of omitted citations

A.1.1 Study at the level of the journal of cited papers

The dispersion related to the p_J value of each journal (defined in Sect. 4.1) can be roughly estimated through an expedient. Each p_J value can be expressed as:

$$p_J = \sum_{i=1}^{P_J} \left[(p_J)_i \cdot (\gamma_J)_i \right] \Big/ \sum_{i=1}^{P_J} (\gamma_J)_i \, , \qquad \text{(A1)}$$

being $(p_J)_i = (\omega_J)_i / (\gamma_J)_i$ the percentage of citations omitted by the database of interest, referring to the *i*-th article published by *J*.

Eq. A1 shows that $(p_J)_i$ can be seen as a weighted average of the omitted-citation rates of individual papers (i.e., $(p_J)_i$ values). These contributions have a variable weight, represented by the number of "theoretically overlapping" citations of each *i*-th article of interest (i.e., $(\gamma_J)_i$). Of course, articles with no citation will have a zero weight.

Being p_J a weighted quantity, one can represent the distribution of $(p_J)_i$ values by a special box-plot based on *weighted quartiles*, defined as ${}^{w}Q_J^{(1)}$, ${}^{w}Q_J^{(2)}$ and ${}^{w}Q_J^{(3)}$, i.e., the weighted first, second (or weighted median) and third quartile of the $(p_J)_i$ values. Weighted quartiles are reported in Tab. A1. These indicators are obtained by ordering in ascending order the $(p_J)_i$ values of the articles of interest and considering the values for which the cumulative of weights is equal to respectively the 25%, 50% and 75% of their sum.

Tab. A1. Weighted quartiles related to the distributions of the $(p_J)_i$ values, for the scientific journals listed in Tab. 3. ${}^{w}Q_J^{(1)}$, ${}^{w}Q_J^{(2)}$ and ${}^{w}Q_J^{(3)}$ are the first, second and third weighted quartile respectively. Statistics are determined both for WoS and Scopus.

	(a) WoS			(b) Scopus		
Journ.	${}^{w}Q_J^{(1)}$	${}^{w}Q_J^{(2)}$	${}^{w}Q_J^{(3)}$	${}^{w}Q_J^{(1)}$	${}^{w}Q_J^{(2)}$	${}^{w}Q_J^{(3)}$
J1	0.0%	0.0%	9.5%	0.0%	0.0%	0.0%
J2	0.0%	0.0%	14.3%	0.0%	0.0%	0.0%
J3	0.0%	9.3%	23.1%	0.0%	0.0%	2.3%
J4	0.0%	0.0%	6.3%	0.0%	5.6%	11.8%
J5	0.0%	0.0%	3.8%	0.0%	0.0%	0.0%
J6	0.0%	0.0%	11.1%	0.0%	0.0%	0.0%
J7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J10	0.0%	0.0%	4.5%	0.0%	0.0%	3.4%
J11	0.0%	0.0%	9.5%	0.0%	0.0%	4.8%
J12	0.0%	0.0%	11.1%	0.0%	0.0%	2.4%
J13	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%
J14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J17	0.0%	2.0%	9.1%	0.0%	0.0%	2.1%
J18	0.0%	0.0%	6.7%	0.0%	0.0%	1.6%
J19	0.0%	0.0%	9.1%	0.0%	0.0%	0.0%
J20	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J21	0.0%	12.5%	33.3%	0.0%	0.0%	0.0%
J22	0.0%	0.0%	12.5%	0.0%	0.0%	0.0%
J23	0.0%	14.3%	44.4%	0.0%	0.0%	9.1%
J24	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J25	0.0%	0.0%	9.1%	0.0%	0.0%	1.8%
J26	0.0%	0.0%	12.5%	0.0%	0.0%	0.0%
J27	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J29	0.0%	0.0%	9.1%	0.0%	0.0%	0.0%
J30	0.0%	0.0%	12.5%	0.0%	0.0%	5.6%
J31	0.0%	4.0%	14.3%	0.0%	0.0%	0.0%
J32	0.0%	0.0%	8.3%	0.0%	0.0%	0.0%
J33	0.0%	0.0%	11.1%	0.0%	0.0%	5.3%
J34	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
J35	0.0%	0.0%	11.1%	0.0%	0.0%	0.0%
J36	0.0%	0.0%	10.0%	0.0%	0.0%	0.0%
J37	0.0%	0.0%	0.0%	0.0%	0.0%	15.4%

Box-plots relating to weighted quartiles are represented in Fig. A2 and Fig. A3, for WoS and Scopus respectively.

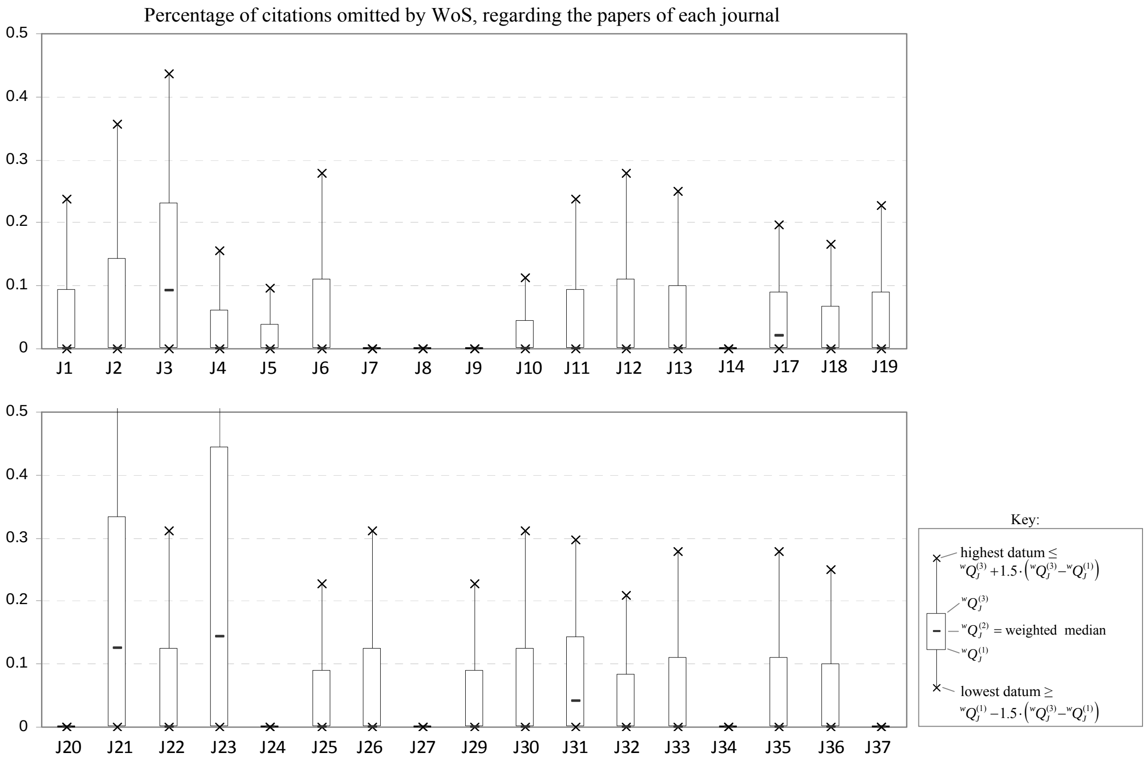

Fig. A1. "Weighted" box-plot of the $(p_J)_i$ values relating to the papers in each journal (J), according to the WoS database. ${}^wQ_J^{(1)}$, ${}^wQ_J^{(2)}$ and ${}^wQ_J^{(3)}$ are the first, second and third weighted quartile of the distributions of interest. Journal abbreviations are reported in Tab. 3.

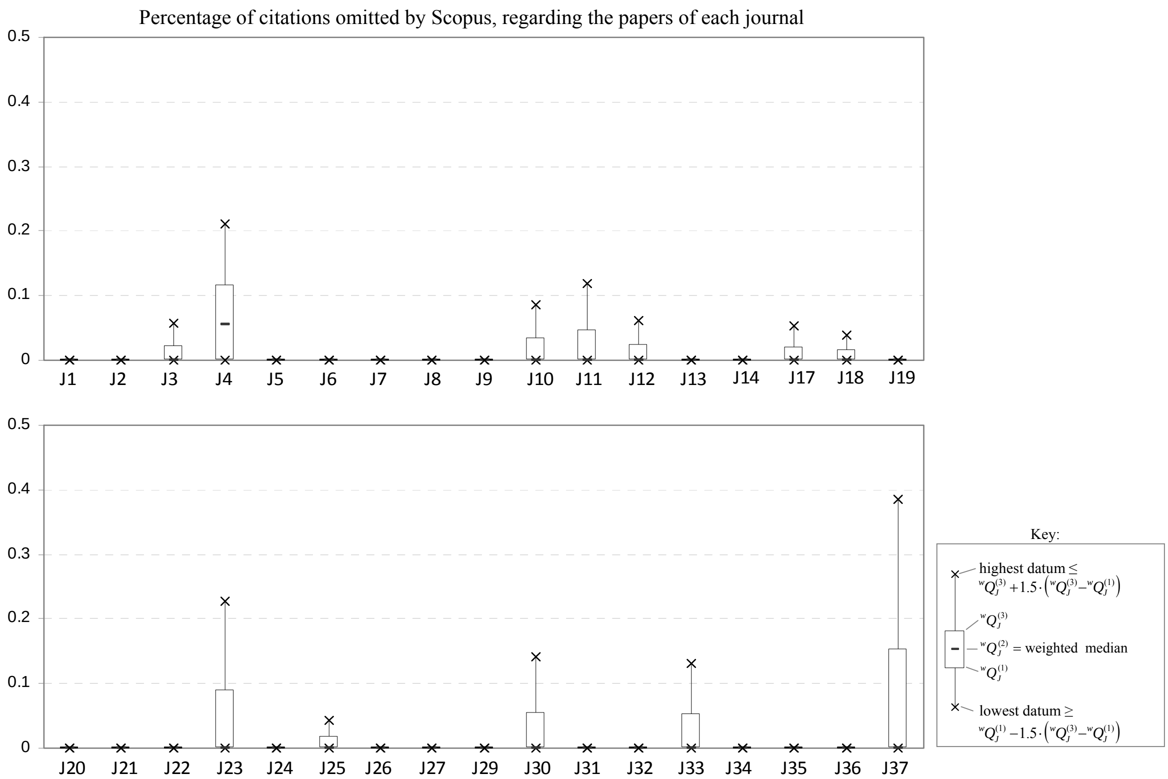

Fig. A2. "Weighted" box-plot of the $(p_J)_i$ values relating to the papers in each journal (J), according to the Scopus database. ${}^wQ_J^{(1)}$, ${}^wQ_J^{(2)}$ and ${}^wQ_J^{(3)}$ are the first, second and third weighted quartile of the distributions of interest. Journal abbreviations are reported in Tab. 3.

The differences between the $(p_J)_i$ distributions of the Manufacturing journals seem insignificant for both WoS and Scopus. The reason is that the notches related to the majority of the journals are overlapped. In particular, we note that most of the notches are “collapsed” on the line corresponding to $(p_J)_i = 0$ and all ${}^{w}Q_J^{(1)}$ values are zero, as well as almost all of ${}^{w}Q_J^{(2)}$ values, both for WoS and Scopus. This result is very interesting because it tells us that omitted citations are generally concentrated into a relatively small number of articles. To confirm this, we can see that – for each of the journals analyzed – the weighted median of the $(p_J)_i$ values (i.e. ${}^{w}Q_J^{(2)}$, in Tab. A1) is systematically lower than the weighted average, i.e. p_J.

A.1.2 Study at the level of the age of citing papers

The dispersion related to the p_Y values of each journal (defined in Sect. 4.2) can be roughly estimated through an expedient, similarly to that presented in Sect. A.1.1. Each p_Y value can be expressed as:

$$p_Y = \sum_{i=1}^{P}\left[(p_Y)_i \cdot (\gamma_Y)_i\right] \Big/ \sum_{i=1}^{P}(\gamma_Y)_i \, , \tag{A2}$$

being $(p_Y)_i = (\omega_Y)_i / (\gamma_Y)_i$ the percentage of citations omitted by the database of interest, among those obtained in the year *Y*, referring to the *i*-th article examined.

Eq. A2 shows that the p_Y value relating to a database can be seen as a weighted average of the omitted-citation rates of individual papers ($(p_Y)_i$). These contributions have a variable weight, given by the number of theoretically overlapping citations ($(\gamma_Y)_i$).

The dispersion of the $(p_Y)_i$ values can be roughly estimated by examining the relevant *weighted quartiles*, defined as ${}^{w}Q_Y^{(1)}$, ${}^{w}Q_Y^{(2)}$ and ${}^{w}Q_Y^{(3)}$. The construction of these indicators is analogous to that described in Sect. 4.1.

The surprising result is that the totality of the weighed quartiles are zero for both databases. This result is not incompatible with the fact that the weighted quartiles seen for individual journals (in Tab. A1) were not necessarily all zero. In this new case, we used time-windows of a single year when counting the (omitted) citations of citing papers; the incidence of articles with zero omitted citations is therefore greater than in the previous case. The practical consequence is that all non-zero $(p_Y)_i$ values fall beyond the third weighted quartile of the corresponding (weighted) distribution. As an example, the graph in Fig. A4 represents the weighted cumulative distribution relating to the $(p_Y)_i$ values for the year 2012, according to WoS. It can be noticed that the first seventy-six weighed percentiles are all zeros. Similar results can be found considering the remaining years.

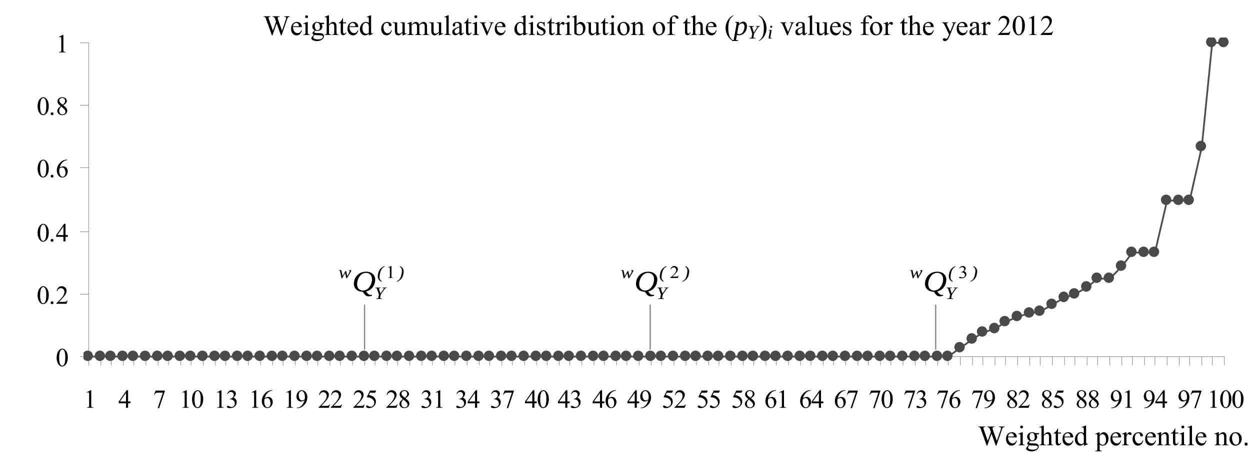

Fig. A3. Weighted cumulative distribution relating to the $(p_Y)_i$ values for the year 2012, according to WoS.

This result confirms the fact that, although the p_Y values of the two databases tend to decrease over time, these variations are quite weak from a statistical viewpoint.

A.2 Additional tables

See Tab. A2 and A3.

Tab. A2. Annual number of articles (*P*) issued by each of the journals analyzed and indexed by both WoS and Scopus. Journal abbreviations are introduced in Tab. 3.

Journ.	Year	*P*	Journ.	Year	*P*	Journ.	Year	*P*	Journ.	Year	*P*	Journ.	Year	*P*
J1	2006	24	J4	2009	229	J7	2012	28	J11	2007	55	J14	2010	54
J1	2007	34	J4	2010	217	J7	Overall	70	J11	2008	81	J14	2011	56
J1	2008	27	J4	2011	238	J8	2006	19	J11	2009	75	J14	2012	56
J1	2009	29	J4	2012	250	J8	2007	28	J11	2010	102	J14	Overall	400
J1	2010	35	J4	Overall	1634	J8	2008	36	J11	2011	126	J17	2006	227
J1	2011	21	J5	2006	30	J8	2009	38	J11	2012	124	J17	2007	247
J1	2012	35	J5	2007	22	J8	2010	29	J11	Overall	645	J17	2008	167
J1	Overall	205	J5	2008	26	J8	2011	43	J12	2006	555	J17	2009	143
J2	2006	39	J5	2009	24	J8	2012	46	J12	2007	479	J17	2010	121
J2	2007	37	J5	2010	23	J8	Overall	239	J12	2008	504	J17	2011	103
J2	2008	36	J5	2011	25	J9	2006	0	J12	2009	705	J17	2012	103
J2	2009	38	J5	2012	25	J9	2007	0	J12	2010	641	J17	Overall	1111
J2	2010	36	J5	Overall	175	J9	2008	0	J12	2011	758	J18	2006	219
J2	2011	39	J6	2006	31	J9	2009	0	J12	2012	631	J18	2007	223
J2	2012	34	J6	2007	33	J9	2010	0	J12	Overall	4273	J18	2008	373
J2	Overall	259	J6	2008	32	J9	2011	139	J13	2006	64	J18	2009	330
J3	2006	151	J6	2009	38	J9	2012	239	J13	2007	65	J18	2010	247
J3	2007	145	J6	2010	30	J9	Overall	378	J13	2008	75	J18	2011	314
J3	2008	142	J6	2011	28	J10	2006	51	J13	2009	65	J18	2012	360
J3	2009	138	J6	2012	31	J10	2007	65	J13	2010	80	J18	Overall	2066
J3	2010	156	J6	Overall	223	J10	2008	83	J13	2011	85	J19	2006	264
J3	2011	161	J7	2006	0	J10	2009	74	J13	2012	83	J19	2007	286
J3	2012	157	J7	2007	0	J10	2010	59	J13	Overall	517	J19	2008	332
J3	Overall	1050	J7	2008	4	J10	2011	62	J14	2006	51	J19	2009	342
J4	2006	238	J7	2009	6	J10	2012	59	J14	2007	62	J19	2010	363
J4	2007	251	J7	2010	12	J10	Overall	453	J14	2008	62	J19	2011	374
J4	2008	211	J7	2011	20	J11	2006	82	J14	2009	59	J19	2012	465

Journ.	Year	*P*
J19	Overall	2426
J20	2006	0
J20	2007	0
J20	2008	0
J20	2009	0
J20	2010	123
J20	2011	29
J20	2012	115
J20	Overall	267
J21	2006	47
J21	2007	45
J21	2008	37
J21	2009	43
J21	2010	48
J21	2011	39
J21	2012	42
J21	Overall	301
J22	2006	56
J22	2007	56
J22	2008	62
J22	2009	65
J22	2010	76
J22	2011	82
J22	2012	211
J22	Overall	608
J23	2006	116
J23	2007	123
J23	2008	133
J23	2009	107
J23	2010	121
J23	2011	113
J23	2012	96
J23	Overall	809
J24	2006	18
J24	2007	20
J24	2008	20
J24	2009	17
J24	2010	19
J24	2011	26
J24	2012	54
J24	Overall	174
J25	2006	706
J25	2007	764
J25	2008	925
J25	2009	765
J25	2010	273
J25	2011	244
J25	2012	295
J25	Overall	3978
J26	2006	4
J26	2007	2
J26	2008	13
J26	2009	3
J26	2010	33
J26	2011	48

Journ.	Year	*P*
J26	2012	60
J26	Overall	163
J27	2006	25
J27	2007	30
J27	2008	31
J27	2009	31
J27	2010	25
J27	2011	25
J27	2012	35
J27	Overall	202
J29	2006	119
J29	2007	152
J29	2008	137
J29	2009	224
J29	2010	214
J29	2011	209
J29	2012	230
J29	Overall	1285
J30	2006	186
J30	2007	159
J30	2008	151
J30	2009	135
J30	2010	138
J30	2011	199
J30	2012	158
J30	Overall	1126
J31	2006	30
J31	2007	36
J31	2008	44
J31	2009	42
J31	2010	34
J31	2011	38
J31	2012	40
J31	Overall	264
J32	2006	48
J32	2007	47
J32	2008	38
J32	2009	67
J32	2010	97
J32	2011	79
J32	2012	76
J32	Overall	452
J33	2006	0
J33	2007	0
J33	2008	46
J33	2009	25
J33	2010	20
J33	2011	18
J33	2012	2
J33	Overall	111
J34	2006	72
J34	2007	65
J34	2008	72
J34	2009	65
J34	2010	59

Journ.	Year	*P*
J34	2011	61
J34	2012	66
J34	Overall	460
J35	2006	13
J35	2007	11
J35	2008	15
J35	2009	24
J35	2010	21
J35	2011	18
J35	2012	22
J35	Overall	124
J36	2006	52
J36	2007	71
J36	2008	76
J36	2009	99
J36	2010	91
J36	2011	109
J36	2012	67
J36	Overall	565
J37	2006	21
J37	2007	16
J37	2008	19
J37	2009	23
J37	2010	25
J37	2011	23
J37	2012	24
J37	Overall	151

Tab. A3. *CPP*, *CPP and relevant statistics for each of the journals examined, in the years from 2008 to 2012. Indicators are calculated both for WoS (a) and Scopus (b).**

			(a) WoS						(b) Scopus					
Journ.	Year	$\sum_{i=1}^{P_J}\gamma_{J_i}$	*CPP*	$\sum_{i=1}^{P_J}\omega_{J_i}$	*p*	CPP^*	95% *CI* limits		*CPP*	$\sum_{i=1}^{P_J}\omega_{J_i}$	*p*	CPP^*	95% *CI* limits	
J1	2008	11	0.410	3	-	-	-	-	0.817	0	-	-	-	-
J1	2009	29	0.508	1	-	-	-	-	1.246	4	-	-	-	-
J1	2010	30	0.554	1	3.3%	0.573	0.536	0.609	1.357	4	13.3%	1.566	1.452	1.680
J1	2011	38	0.594	3	7.9%	0.645	0.591	0.699	1.844	2	5.3%	1.946	1.868	2.024
J1	2012	18	0.302	0	-	-	-	-	0.679	0	-	-	-	-
J2	2008	15	0.313	1	-	-	-	-	0.512	0	-	-	-	-
J2	2009	31	0.383	4	12.9%	0.439	0.390	0.489	0.589	3	9.7%	0.652	0.596	0.708
J2	2010	38	0.476	5	13.2%	0.548	0.492	0.603	0.797	1	2.6%	0.819	0.785	0.853
J2	2011	37	0.482	4	10.8%	0.540	0.490	0.590	0.920	2	5.4%	0.973	0.921	1.024
J2	2012	34	0.388	5	14.7%	0.455	0.403	0.507	0.844	0	0.0%	0.844	0.844	0.844
J3	2008	246	0.684	100	40.7%	1.153	1.092	1.214	1.319	5	2.0%	1.346	1.327	1.365
J3	2009	354	1.217	74	20.9%	1.539	1.481	1.598	1.980	5	1.4%	2.008	1.989	2.028
J3	2010	435	1.426	72	16.6%	1.708	1.651	1.766	2.592	2	0.5%	2.604	2.591	2.617
J3	2011	511	1.861	25	4.9%	1.956	1.921	1.991	2.633	16	3.1%	2.718	2.684	2.751
J3	2012	567	1.953	41	7.2%	2.105	2.063	2.147	2.546	18	3.2%	2.629	2.597	2.661
J4	2008	881	1.518	48	5.4%	1.605	1.576	1.634	2.174	81	9.2%	2.394	2.354	2.433
J4	2009	976	1.636	43	4.4%	1.712	1.684	1.740	2.930	73	7.5%	3.167	3.124	3.210
J4	2010	977	1.715	45	4.6%	1.798	1.768	1.827	2.751	61	6.2%	2.934	2.894	2.973
J4	2011	1130	2.146	41	3.6%	2.226	2.199	2.254	3.351	73	6.5%	3.583	3.539	3.627
J4	2012	1154	2.386	46	4.0%	2.485	2.456	2.514	3.031	84	7.3%	3.269	3.225	3.313
J5	2008	22	0.638	2	-	-	-	-	1.207	1	-	-	-	-
J5	2009	25	0.604	1	-	-	-	-	1.302	2	-	-	-	-
J5	2010	24	0.600	2	-	-	-	-	1.140	4	-	-	-	-
J5	2011	23	0.447	2	-	-	-	-	0.894	0	-	-	-	-
J5	2012	22	0.438	0	-	-	-	-	0.625	0	-	-	-	-
J6	2008	35	0.866	1	2.9%	0.891	0.853	0.930	1.415	2	5.7%	1.501	1.431	1.572
J6	2009	33	0.667	2	6.1%	0.710	0.660	0.759	2.092	1	3.0%	2.158	2.095	2.220
J6	2010	59	0.886	1	1.7%	0.901	0.872	0.930	2.643	0	0.0%	2.643	2.643	2.643
J6	2011	45	0.971	1	2.2%	0.993	0.957	1.028	2.147	0	0.0%	2.147	2.147	2.147
J6	2012	60	1.017	4	6.7%	1.090	1.022	1.158	1.828	0	0.0%	1.828	1.828	1.828
J7	2008	0	-	0	-	-	-	-	-	0	-	-	-	-
J7	2009	0	-	0	-	-	-	-	-	0	-	-	-	-
J7	2010	3	0.364	0	-	-	-	-	0.250	0	-	-	-	-
J7	2011	3	0.167	0	-	-	-	-	0.222	0	-	-	-	-
J7	2012	17	0.594	1	-	-	-	-	0.550	3	-	-	-	-
J8	2008	13	0.479	1	-	-	-	-	0.358	3	-	-	-	-
J8	2009	20	0.354	2	-	-	-	-	0.478	0	-	-	-	-
J8	2010	20	0.280	2	-	-	-	-	0.413	4	-	-	-	-
J8	2011	40	0.471	1	2.5%	0.483	0.459	0.506	0.971	1	2.5%	0.995	0.958	1.033
J8	2012	39	0.467	2	5.1%	0.492	0.459	0.525	0.770	5	12.8%	0.884	0.810	0.957
J9	2008	0	-	0	-	-	-	-	-	0	-	-	-	-
J9	2009	0	-	0	-	-	-	-	-	0	-	-	-	-
J9	2010	0	-	0	-	-	-	-	-	0	-	-	-	-
J9	2011	0	-	0	-	-	-	-	-	0	-	-	-	-
J9	2012	52	0.599	5	9.6%	0.663	0.631	0.695	1.259	3	5.8%	1.336	1.290	1.382
J10	2008	68	0.897	2	2.9%	0.924	0.893	0.956	1.145	2	2.9%	1.180	1.147	1.213
J10	2009	95	1.110	5	5.3%	1.172	1.127	1.216	1.103	7	7.4%	1.190	1.145	1.236
J10	2010	76	0.810	1	1.3%	0.821	0.802	0.839	1.093	3	3.9%	1.138	1.105	1.171
J10	2011	67	0.574	5	7.5%	0.620	0.583	0.658	1.053	1	1.5%	1.069	1.047	1.090
J10	2012	79	0.719	1	1.3%	0.728	0.711	0.746	1.116	5	6.3%	1.191	1.143	1.239
J11	2008	148	1.509	8	5.4%	1.595	1.551	1.639	2.216	4	2.7%	2.277	2.238	2.317
J11	2009	247	2.245	24	9.7%	2.487	2.414	2.560	3.103	8	3.2%	3.206	3.156	3.257
J11	2010	295	2.032	13	4.4%	2.126	2.078	2.173	3.337	13	4.4%	3.491	3.432	3.550
J11	2011	382	2.141	23	6.0%	2.278	2.224	2.332	3.898	10	2.6%	4.003	3.957	4.050

			(a) WoS						(b) Scopus					
Journ.	Year	$\sum_{i=1}^{P_J} \gamma_{J_i}$	*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	CPP^*	95% *CI* limits		*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	CPP^*	95% *CI* limits	
J11	2012	543	2.171	49	9.0%	2.386	2.328	2.445	3.836	33	6.1%	4.084	4.021	4.148
J12	2008	500	0.656	60	12.0%	0.746	0.728	0.763	0.809	16	3.2%	0.836	0.826	0.845
J12	2009	838	1.040	78	9.3%	1.146	1.127	1.166	1.228	33	3.9%	1.278	1.265	1.291
J12	2010	1140	0.984	98	8.6%	1.077	1.060	1.094	1.418	90	7.9%	1.539	1.520	1.558
J12	2011	1424	1.044	100	7.0%	1.123	1.108	1.138	1.499	64	4.5%	1.569	1.555	1.583
J12	2012	1431	0.959	123	8.6%	1.049	1.034	1.065	1.596	51	3.6%	1.654	1.641	1.668
J13	2008	59	0.821	5	8.5%	0.897	0.853	0.942	0.842	2	3.4%	0.871	0.843	0.900
J13	2009	72	0.573	12	16.7%	0.688	0.636	0.740	0.901	1	1.4%	0.914	0.895	0.933
J13	2010	75	0.513	9	12.0%	0.583	0.543	0.622	0.863	4	5.3%	0.912	0.876	0.947
J13	2011	150	1.000	6	4.0%	1.042	1.010	1.073	1.444	3	2.0%	1.473	1.446	1.501
J13	2012	138	0.764	12	8.7%	0.836	0.796	0.876	1.200	3	2.2%	1.227	1.202	1.252
J14	2008	47	0.474	4	8.5%	0.518	0.480	0.555	0.561	4	8.5%	0.614	0.573	0.655
J14	2009	35	0.371	2	5.7%	0.393	0.367	0.420	0.573	3	8.6%	0.626	0.586	0.666
J14	2010	64	0.562	4	6.3%	0.599	0.565	0.634	0.694	2	3.1%	0.717	0.690	0.743
J14	2011	83	0.735	5	6.0%	0.782	0.742	0.821	1.071	0	0.0%	1.071	1.071	1.071
J14	2012	94	0.800	4	4.3%	0.836	0.800	0.871	1.000	3	3.2%	1.033	0.999	1.067
J17	2008	591	1.739	55	9.3%	1.917	1.881	1.954	1.880	7	1.2%	1.903	1.889	1.916
J17	2009	638	1.799	61	9.6%	1.989	1.948	2.029	2.600	13	2.0%	2.654	2.631	2.676
J17	2010	550	1.875	41	7.5%	2.027	1.984	2.069	2.760	15	2.7%	2.838	2.807	2.869
J17	2011	609	2.523	27	4.4%	2.640	2.599	2.681	3.140	10	1.6%	3.193	3.165	3.221
J17	2012	463	2.138	25	5.4%	2.260	2.215	2.306	3.009	8	1.7%	3.062	3.031	3.092
J18	2008	594	1.690	34	5.7%	1.793	1.764	1.822	2.463	20	3.4%	2.549	2.522	2.576
J18	2009	923	1.886	57	6.2%	2.010	1.982	2.038	2.746	21	2.3%	2.810	2.789	2.830
J18	2010	1111	1.590	70	6.3%	1.697	1.673	1.721	2.837	27	2.4%	2.908	2.888	2.928
J18	2011	874	1.517	57	6.5%	1.623	1.597	1.649	2.642	25	2.9%	2.720	2.697	2.743
J18	2012	1003	1.760	44	4.4%	1.841	1.817	1.864	2.853	36	3.6%	2.959	2.932	2.986
J19	2008	273	0.588	18	6.6%	0.629	0.613	0.646	1.009	14	5.1%	1.064	1.044	1.083
J19	2009	349	0.688	27	7.7%	0.746	0.727	0.764	1.122	13	3.7%	1.165	1.149	1.181
J19	2010	596	0.866	56	9.4%	0.955	0.933	0.977	1.513	50	8.4%	1.651	1.624	1.679
J19	2011	718	1.028	52	7.2%	1.109	1.088	1.129	1.566	52	7.2%	1.688	1.663	1.714
J19	2012	997	1.265	75	7.5%	1.368	1.345	1.390	1.771	50	5.0%	1.864	1.842	1.886
J20	2008	0	0.125	0	-	-	-	-	-	0	-	-	-	-
J20	2009	0	0.162	0	-	-	-	-	-	0	-	-	-	-
J20	2010	0	0.287	0	-	-	-	-	-	0	-	-	-	-
J20	2011	19	0.182	0	-	-	-	-	0.260	1	-	-	-	-
J20	2012	63	0.427	2	3.2%	0.441	0.422	0.459	0.470	1	1.6%	0.478	0.464	0.492
J21	2008	52	0.645	11	21.2%	0.818	0.742	0.895	1.160	1	1.9%	1.182	1.152	1.213
J21	2009	38	0.443	7	18.4%	0.543	0.482	0.604	1.146	3	7.9%	1.244	1.181	1.308
J21	2010	35	0.291	10	28.6%	0.407	0.345	0.469	1.068	1	2.9%	1.100	1.062	1.137
J21	2011	38	0.275	9	23.7%	0.360	0.307	0.413	1.140	2	5.3%	1.203	1.152	1.254
J21	2012	35	0.247	8	22.9%	0.320	0.270	0.371	0.557	3	8.6%	0.609	0.562	0.656
J22	2008	40	0.575	5	12.5%	0.657	0.607	0.708	1.168	2	5.0%	1.230	1.184	1.275
J22	2009	56	0.636	7	12.5%	0.726	0.675	0.778	1.346	6	10.7%	1.508	1.441	1.574
J22	2010	93	0.709	7	7.5%	0.766	0.725	0.807	1.245	12	12.9%	1.429	1.363	1.495
J22	2011	90	0.539	12	13.3%	0.622	0.577	0.667	1.033	13	14.4%	1.208	1.145	1.270
J22	2012	145	0.759	19	13.1%	0.874	0.824	0.924	0.975	24	16.6%	1.169	1.112	1.225
J23	2008	179	0.713	42	23.5%	0.931	0.879	0.983	1.037	11	6.1%	1.105	1.073	1.137
J23	2009	170	0.485	59	34.7%	0.742	0.691	0.793	1.162	10	5.9%	1.234	1.202	1.267
J23	2010	169	0.353	89	52.7%	0.745	0.690	0.801	0.938	32	18.9%	1.157	1.103	1.211
J23	2011	132	0.395	37	28.0%	0.549	0.505	0.592	0.783	31	23.5%	1.023	0.966	1.079
J23	2012	146	0.379	58	39.7%	0.630	0.581	0.678	0.686	27	18.5%	0.842	0.796	0.888
J24	2008	4	0.189	0	-	-	-	-	0.051	1	-	-	-	-
J24	2009	4	0.150	0	-	-	-	-	0.088	1	-	-	-	-
J24	2010	9	0.378	0	-	-	-	-	0.205	0	-	-	-	-
J24	2011	17	0.472	1	-	-	-	-	0.122	0	-	-	-	-
J24	2012	40	0.822	4	10.0%	0.914	0.828	0.999	0.216	2	5.0%	0.228	0.216	0.240
J25	2008	1326	1.148	93	7.0%	1.234	1.219	1.249	1.341	41	3.1%	1.384	1.373	1.394
J25	2009	1954	1.278	152	7.8%	1.385	1.370	1.401	1.682	47	2.4%	1.723	1.714	1.733
J25	2010	2346	1.450	181	7.7%	1.571	1.555	1.587	2.073	64	2.7%	2.131	2.119	2.142

			(a) WoS						(b) Scopus					
Journ.	Year	$\sum_{i=1}^{P_J} \gamma_{J_i}$	*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	*CPP**	95% *CI* limits		*CPP*	$\sum_{i=1}^{P_J} \omega_{J_i}$	*p*	*CPP**	95% *CI* limits	
J25	2011	1931	1.812	122	6.3%	1.935	1.914	1.956	2.467	59	3.1%	2.545	2.528	2.562
J25	2012	915	1.729	59	6.4%	1.848	1.819	1.877	2.340	22	2.4%	2.398	2.377	2.419
J26	2008	6	1.000	1	-	-	-	-	0.209	0	-	-	-	-
J26	2009	8	1.188	0	-	-	-	-	0.288	1	-	-	-	-
J26	2010	22	1.294	1	-	-	-	-	0.531	0	-	-	-	-
J26	2011	13	0.417	0	-	-	-	-	0.287	1	-	-	-	-
J26	2012	58	0.679	6	10.3%	0.757	0.698	0.816	0.938	3	5.2%	0.989	0.944	1.034
J27	2008	33	0.810	0	0.0%	0.810	0.810	0.810	0.672	1	3.0%	0.693	0.656	0.731
J27	2009	20	0.311	3	-	-	-	-	0.426	0	-	-	-	-
J27	2010	26	0.403	3	-	-	-	-	0.677	3	-	-	-	-
J27	2011	40	0.786	0	0.0%	0.786	0.786	0.786	0.929	2	5.0%	0.977	0.920	1.035
J27	2012	41	0.740	2	4.9%	0.778	0.724	0.832	1.020	1	2.4%	1.046	1.001	1.090
J29	2008	193	0.724	14	7.3%	0.780	0.753	0.807	0.787	4	2.1%	0.803	0.788	0.818
J29	2009	250	0.835	16	6.4%	0.892	0.865	0.919	1.041	5	2.0%	1.062	1.046	1.079
J29	2010	266	0.749	12	4.5%	0.784	0.765	0.803	0.865	6	2.3%	0.885	0.871	0.900
J29	2011	447	0.977	31	6.9%	1.050	1.025	1.075	1.205	5	1.1%	1.218	1.207	1.229
J29	2012	525	1.170	38	7.2%	1.261	1.233	1.289	1.369	4	0.8%	1.379	1.369	1.389
J30	2008	109	0.340	19	17.4%	0.412	0.386	0.437	0.393	15	13.8%	0.456	0.432	0.481
J30	2009	133	0.452	10	7.5%	0.489	0.468	0.510	0.685	26	19.5%	0.851	0.810	0.892
J30	2010	181	0.670	18	9.9%	0.744	0.714	0.775	0.924	7	3.9%	0.961	0.939	0.984
J30	2011	203	0.722	17	8.4%	0.788	0.759	0.817	1.018	6	3.0%	1.049	1.028	1.070
J30	2012	218	0.623	24	11.0%	0.700	0.673	0.728	0.817	2	0.9%	0.825	0.815	0.834
J31	2008	54	1.221	4	7.4%	1.318	1.245	1.391	1.046	0	0.0%	1.046	1.046	1.046
J31	2009	76	1.098	6	7.9%	1.192	1.127	1.257	1.136	5	6.6%	1.216	1.155	1.277
J31	2010	103	1.253	16	15.5%	1.483	1.389	1.578	1.528	3	2.9%	1.574	1.529	1.619
J31	2011	70	0.926	6	8.6%	1.013	0.950	1.075	1.076	1	1.4%	1.092	1.066	1.118
J31	2012	46	0.571	9	19.6%	0.710	0.634	0.787	0.790	2	4.3%	0.826	0.789	0.863
J32	2008	58	0.773	3	5.2%	0.815	0.775	0.856	1.208	1	1.7%	1.230	1.200	1.259
J32	2009	87	1.247	7	8.0%	1.356	1.288	1.425	1.953	3	3.4%	2.023	1.966	2.079
J32	2010	151	1.524	8	5.3%	1.609	1.554	1.665	2.676	6	4.0%	2.787	2.723	2.851
J32	2011	196	1.293	8	4.1%	1.348	1.312	1.384	1.817	7	3.6%	1.884	1.845	1.924
J32	2012	222	1.273	15	6.8%	1.365	1.321	1.409	1.852	22	9.9%	2.056	1.991	2.121
J33	2008	0	-	0	-	-	-	-	-	0	-	-	-	-
J33	2009	69	4.037	6	8.7%	4.422	4.260	4.583	0.344	1	1.4%	0.349	0.340	0.359
J33	2010	162	1.457	21	13.0%	1.674	1.589	1.759	0.860	9	5.6%	0.911	0.880	0.942
J33	2011	18	1.152	1	-	-	-	-	0.122	1	-	-	-	-
J33	2012	36	1.188	2	5.6%	1.257	1.209	1.306	0.150	3	8.3%	0.163	0.148	0.178
J34	2008	43	0.410	3	7.0%	0.441	0.412	0.470	0.700	2	4.7%	0.734	0.704	0.765
J34	2009	74	0.647	5	6.8%	0.694	0.659	0.730	0.929	1	1.4%	0.941	0.922	0.960
J34	2010	53	0.384	3	5.7%	0.407	0.382	0.432	0.797	2	3.8%	0.828	0.799	0.858
J34	2011	63	0.500	8	12.7%	0.573	0.527	0.618	0.935	2	3.2%	0.966	0.935	0.997
J34	2012	55	0.487	2	3.6%	0.506	0.481	0.530	0.941	6	10.9%	1.056	0.998	1.115
J35	2008	16	1.080	0	-	-	-	-	2.040	0	-	-	-	-
J35	2009	21	1.154	3	-	-	-	-	2.423	0	-	-	-	-
J35	2010	33	0.821	3	9.1%	0.903	0.815	0.990	2.462	7	21.2%	3.124	2.893	3.356
J35	2011	39	1.000	1	2.6%	1.026	0.979	1.074	2.848	2	5.1%	3.002	2.889	3.115
J35	2012	48	1.051	2	4.2%	1.097	1.030	1.164	1.500	0	0.0%	1.500	1.500	1.500
J36	2008	95	1.109	12	12.6%	1.270	1.204	1.336	1.866	2	2.1%	1.906	1.871	1.942
J36	2009	175	1.455	18	10.3%	1.621	1.559	1.684	2.704	7	4.0%	2.817	2.763	2.870
J36	2010	176	1.022	17	9.7%	1.132	1.085	1.178	2.447	5	2.8%	2.519	2.479	2.558
J36	2011	205	0.968	23	11.2%	1.091	1.043	1.139	2.037	7	3.4%	2.109	2.071	2.147
J36	2012	197	0.985	13	6.6%	1.055	1.019	1.091	1.950	2	1.0%	1.970	1.950	1.990
J37	2008	18	0.590	0	-	-	-	-	0.821	4	-	-	-	-
J37	2009	24	0.914	0	-	-	-	-	0.800	10	-	-	-	-
J37	2010	25	0.591	1	-	-	-	-	0.810	1	-	-	-	-
J37	2011	17	0.392	1	-	-	-	-	0.521	0	-	-	-	-
J37	2012	31	0.735	0	0.0%	0.735	0.735	0.735	1.104	1	3.2%	1.141	1.087	1.196

$\Sigma(\gamma_J)_i$ is the total number of "theoretically overlapping" citations;

CPP are the average citations per papers;

$\sum(\omega_J)_i$ is the total number of omitted citations;

p is the estimate of the omitted-citation rate of a journal; p was not estimated in the case $\sum \gamma_{J_i} < 30$;

CPP^* is the corrected CPP (with the corresponding 95% confidence interval limits in the two columns to the right). CPP^* was not calculated in the case $\sum \gamma_{J_i} < 30$.