Publication Rates in 192 Research Fields of the Hard Sciences

Ciriaco Andrea D'Angelo[1] and Giovanni Abramo[2]

[1] *dangelo@dii.uniroma2.it*
Department of Engineering and Management
University of Rome "Tor Vergata" – Italy, Via del Politecnico 1, 00133 Rome (Italy)

[2] *giovanni.abramo@uniroma2.it*
Laboratory for Studies of Research and Technology Transfer, Institute for System Analysis and Computer Science (IASI-CNR), National Research Council of Italy, Via dei Taurini 19, 00185 Rome (Italy)

Abstract
Bibliometricians are aware that the citation behavior of scientists varies across fields, and for this they carefully normalize citations by field. They are also aware of the different publication intensities across fields. This imposes that the research performance of a scientist must be compared with that of their colleagues in the same field. Every comparison of scientists in different fields should be preceded by the normalization of the performances, and the same holds for comparing multidisciplinary organizational units. If the Web of Science recognizes 251 subject categories, there should be a somewhat similar number of research fields for the classification of the scientists. The Italian academic system is quite unique in providing a classification of professors, into 370 fields, 192 of them in the hard sciences. In this work we measure the descriptive statistics on annual publication (full and fractional counting) by Italian academics in each of the 192 hard science fields. These statistics help recognize the extent of distortion from failing to normalize the research performance of scientists based in different fields. They could also serve as scaling factors for avoiding distortion in rankings, including in other nations.

Conference Topic
Methods and techniques

Introduction
The purpose of bibliometrics is to provide continuously better support for the policy-makers and administrators of research institutions, in the achievement of their specific objectives, through the provision of methods and indicators for the evaluation of performance that are themselves always more accurate, robust, reliable and functional. The principle obstacle to bibliometrics is the insufficiency of the data to meet such high standards. The practitioner is thus forced to resort to proxies in measurement, which cause varying degrees of distortion in the results.

Research organizations are likened to other productive organizations, but where the product is new knowledge, rather than some other good or service. An organization's performance is then better than that of another one if, at parity of resources, it produces more knowledge or if, at parity of output, it consumes less resources. It is the shortage of information on inputs (production factors) that presents the greatest problem to bibliometricians. The production factors are labor and capital. Capital embeds all those resources other than labor (facilities, technical instruments, materials, databases, etc.). When we wish to measure labor productivity we must thus normalize for capital. But who can really know the financial and technical resources available to all the different institutions, departments, and then individual researchers? The bibliometrician also frequently lacks information on the realities of labor, due to the absence of databases on the researchers, and on their institutional, discipline and field affiliations.

Given these obstacles, practitioners often use indicators that do not relate output to input. This means they produce ranking lists that are highly size-dependent. At that point we cannot know what part of an organization's or nation's rank arises from its performance or is due to size. Examples of this are the CWTS Leiden[1] and SCImago[2] lists, which rank universities by publications and fractional publications. Others have proposed indicators that attempt to get around the problems by relating the impact or excellence of research not to input, but rather to the output itself. Examples of this are the "new crown indicator" (Waltman et al., 2011), which measures the average impact per publication, or the "proportion of highly-cited articles to total publications" (Waltman et al., 2012). However, with this type of indicator, even when the output of the scientist increases, other factors remaining equal, his or her performance could still decrease: a paradox and a violation of the fundamental principle of the measure of efficiency.

In those cases where an indicator does relate output to input, it is still often applied at levels of organizational aggregation that are too high, ignoring the differing intensity of publication across fields. Bibliometricians have been aware of this problem for many years (Butler, 2007; Moed et al., 1985; Garfield, 1979), and are also aware of the distortion that afflicts the resulting aggregate rankings (Abramo, D'Angelo, & Di Costa, 2008). However the task of finer aggregation is difficult to solve without a database that classifies the researchers by field of research. Where they exist, such databases are maintained at central levels. Apart from the Italian one[3], maintained by the Italian Ministry of Education, Universities and Research (MIUR), the only other large-scale one we are aware of is the Norwegian Research Personnel Register[4] compiled by the Nordic Institute for Studies in Innovation, Research and Education (NIFU).

The NIFU system classifies scientists in 58 scientific fields grouped in five main domains. Perhaps the lower number of scientists in Norway works against finer classification: in fact comparing the performance of small numbers of researchers per field creates serious problems of significance. However, on the other hand, the Web of Science (WoS) identifies a full 251 subject categories for the classification of journals. And if there are this many fields for classifying scientific journals, there must be at least that many fields for classifying scientific work, and the scientists. In smaller nations or emerging economies we could expect to see fewer number of these fields present, since research structures will be unable to deal with all the areas, and we would expect to see research in more concentrated fields. However, in larger, developed countries we can expect to see the full spectrum of research fields. In fact in Italy the MIUR manages a system for the classification of all professors into a total of 370 "scientific disciplinary sectors" (SDSs).[5] Each professor belongs to one and only one of the SDSs, which are grouped into 14 university disciplinary areas (UDAs). Further, 192 of the SDSs from 9 of the UDAs fall in the so-called hard sciences. In the following we refer to these SDS by their code or acronym.[6] These 192 SDSs compare to the 176 WoS subject categories identified in the JCR-Science Citation Index (see the Annex 1[7] for a conversion of SDSs to WoS subject categories.

As noted above, the lack of field classification of scientists means that measures of research performance will inevitably be affected by distortions in rankings, due to the different intensity of publication across fields. The higher the level of aggregation, the stronger these distortions become. The corollary is that, rising to international levels, it has been impossible

[1]http://www.leidenranking.com/ranking/2014, last accessed on April 8, 2015.

[2]http://www.scimagoir.com/research.php, last accessed on April 8, 2015.

[3]http://cercauniversita.cineca.it/php5/docenti/cerca.php, last accessed on April 8, 2015.

[4] http://www.nifu.no/en/statistikk/databaser-og-registre/4897-2/ last accessed on April 8, 2015.

[5] The complete list is accessible on attiministeriali.miur.it/UserFiles/115.htm, last accessed April 8, 2015.

[6] The full names can be found in www.iasi.cnr.it/laboratoriortt/TESTI/Altro/ISSI-ANNEX%202_P.pdf, last accessed on April 8, 2015

[7] www.iasi.cnr.it/laboratoriortt/TESTI/Altro/ISSI-ANNEX1.pdf, last accessed on April 8, 2015

to correctly compare institutional or national research performance.
To date, in fact there is no international standard for the classification of scientists. Thus in this work we provide our colleagues and practitioners with descriptive statistics on yearly publications (both full and fractional counting) of Italian academics in each of the 192 hard science SDSs. Our intention is that these statistics might first permit recognition of the extent of distortions that occur when evaluations compare the research performance of scientists within the same discipline, but in different fields. For those nations lacking databases of researchers by field, our statistics could also serve as normalization factors, serving to reduce the distortions when comparing research performance of individuals, groups or entire research organizations.

Data and Methods

In the study we measure "publication rates" in 192 SDSs, meaning average yearly publications of individual scientists, over the period 2009-2013.[8] Data on Italian academics are extracted from the official database maintained by the MIUR. The database indexes the name, academic rank, affiliation, and SDS of all academics in Italian universities. At 31/12/2013 the entire Italian university population consisted of 56,600 professors employed in 96 universities, which are authorized by the MIUR to grant legally recognized degrees. It has been shown (Moed, 2005) that in the so-called hard sciences, the prevalent form of codification for research output is publication in scientific journals. For reasons of robustness, we thus examine only the nine UDAs that deal with the hard sciences,[9] including a total of 192 SDSs. Furthermore, again for reasons of robustness, we calculate the yearly average publication rates only of those professors who have been on staff for at least three years over the observed period.

Table 1. Dataset for the analysis: number of fields (SDSs), universities, research staff and WoS publications in each UDA under investigation

UDA	SDS	Universities	Research staff	Publications*
Mathematics and computer science	10	72	2,930	16,262
Physics	8	65	2,003	22,597
Chemistry	12	60	2,701	26,054
Earth sciences	12	49	974	6,066
Biology	19	67	4,423	34,406
Medicine	50	65	8,998	72,661
Agricultural and veterinary sciences	30	56	2,820	14,951
Civil engineering	9	54	1,394	7,462
Industrial and information engineering	42	73	4,791	40,572
Total	192	86	31,034	207,132†

* Figures refer to publications authored by at least one professor pertaining to the UDA.
† Total is less than the sum of the column data due to double counts of publications co-authored by researchers pertaining to SDSs of more than one UDA.

Publication data are drawn from the Italian Observatory of Public Research (ORP), a database developed and maintained by the authors and derived under license from the WoS. Beginning from the raw data of Italian publications[10] indexed in WoS-ORP, we apply a complex

[8] For the most appropriate publication period to be observed see Abramo et al. (2012b).
[9] Mathematics and computer sciences; Physics; Chemistry; Earth sciences; Biology; Medicine; Agricultural and veterinary sciences; Civil engineering; Industrial and information engineering.
[10] We exclude those document types that cannot be strictly considered as true research products, such as editorial material, meeting abstracts, replies to letters, etc.

algorithm for disambiguation of the true identity of the authors and their institutional affiliations (for details see D'Angelo et al., 2011). Each publication is attributed to the university professors that authored it, with a harmonic average of precision and recall (F-measure) equal to 96 (error of 4%). We further reduce this error by manual disambiguation. Because each professor belongs to one and only one SDS, we can then calculate the distribution of annual publication rates and the relevant descriptive statistics in each SDS.

The dataset for the analysis includes 31,034 professors, employed in 86 universities, authoring over 200,000 WoS publications, sorted in the UDAs as shown in Table 1.

Research projects frequently involve a team of researchers, a fact revealed in the co-authorship of publications. Various performance measures account for the fractional contributions of single co-authors to outputs. The contributions of the individual co-authors to the achievement of the publication are not necessarily equal, and in some fields the authors signal the different contributions through the ordering of the byline. The conventions on the order of authors for scientific papers differ across fields (Pontille, 2004; RIN, 2009), thus in the current study, the fractional contribution of the individuals is weighted accordingly.

Fractional contribution equals the inverse of the number of authors, in those fields where the practice is to place the authors in simple alphabetical order but assumes different weights in other cases, particularly in the life sciences. For these disciplines, we give different weights to each co-author according to their order in the byline and the character of the co-authorship (intra-mural or extra-mural). If first and last authors belong to the same university, 40% of citations are attributed to each of them; the remaining 20% are divided among all other authors. If the first two and last two authors belong to different universities, 30% of citations are attributed to first and last authors; 15% of citations are attributed to second and last author but one; the remaining 10% are divided among all others.[11] Failure to account for the number and position of authors in the byline would result in notable ranking differences, both at the individual level (Abramo, D'Angelo & Rosati, 2013a) and at the institution level (Abramo, D'Angelo & Rosati, 2013b).

Applying the above conventions, for each of the 192 SDS we will provide descriptive statistics on the intensity of annual publication: referred to as P for full counting and FP for fractional counting. We then examine further statistics on P and FP for the SDSs included in each UDA.

Results

Publication rates of professors in a specific field

The publication intensity of professors in a given field is known to be particularly skewed, with a small percentage of individuals authoring a large share of the total papers, and the others authoring a small share (Egghe, 2005; Kyvik, 1989; Lotka, 1926). Figure 1 provides the example of the field of Organic chemistry (SDS CHIM/06), showing the distribution of the average number of publications per year over the period under examination, for each of the 554 professors in the SDS. The distribution fits quite well a logarithmic curve, as indicated by the particularly high value of R^2 (0.974). Here, 10% of the professors have produced on average less than one publication per year, and six were totally unproductive. On the opposite front, we find 20 professors with over 10 publications per year, and one absolute outlier with 25.

The box plot (right side of Figure 1) refers to the same distribution. It shows a median of 3 publications per year and an interquartile range (difference between third and first quartile) of

[11] The weighting values were assigned following advice from senior Italian professors in the life sciences. The values could be changed to suit different practices in other national contexts.

2.6. It also brings out the presence of 30 outliers: hyper-productive professors with a performance that exceeds that of the third quartile by over 1.5 times the interquartile difference.
The distribution of frequencies by class of publication rates (Figure 2) shows a mode between 2 and 3 publications annually and a particularly long right tail, with a final peak for the hyper-productive professors.
The distribution of the average yearly publications measured by fractional counting (FP) shows a very similar situation: in Figure 3 the right tail is actually longer than that for only full counting (Figure 2).
The distributions seen for SDS CHIM/06 show structural elements that recur in the analyses of the other 191 SDSs. Most obvious is the skewness, although there are some interesting exceptions, for example as in VET/04 (Inspection of food products of animal origin). The 77 professors of this SDS have a publication rate that is almost uniform, as illustrated in Figure 4.

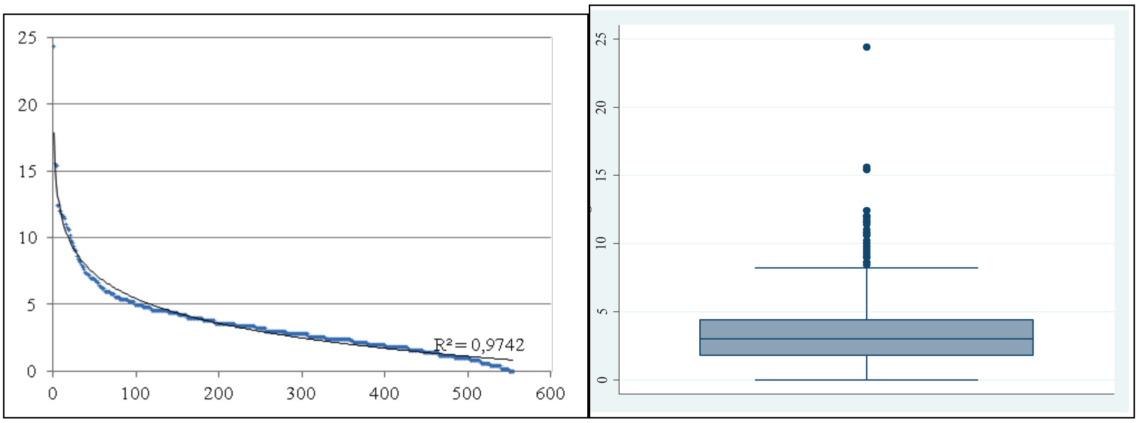

Figure 1. Distribution and box plot of annual publication rate P (full counting, 2009-2013) for 554 Italian professors in Organic chemistry (CHIM/06).

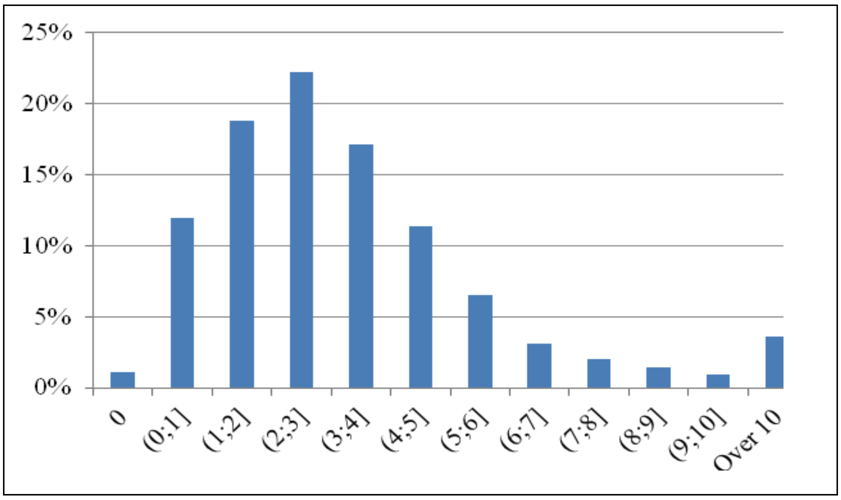

Figure 2. Frequency distribution for classes of annual publication rate P (2009-2013) for the 554 Italian professors in CHIM/06.

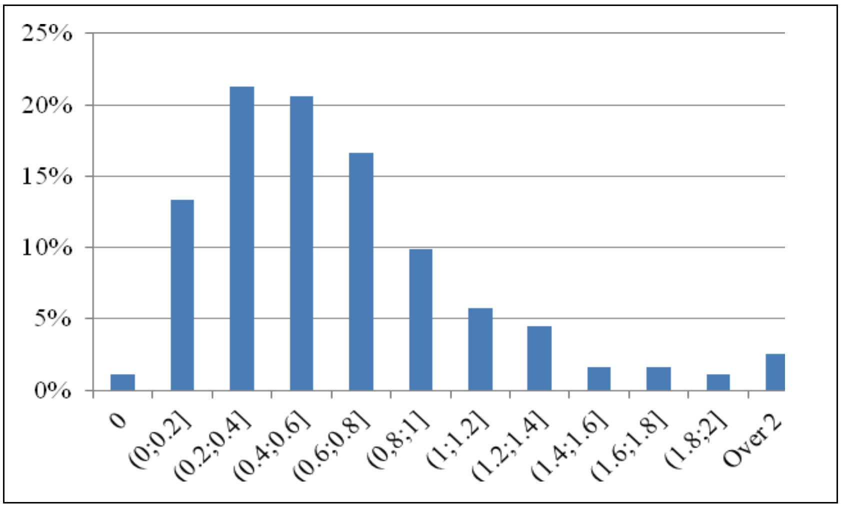

Figure 3. Frequency distribution for classes of annual publication rate FP (fractional counting, 2009-2013) for the 554 Italian professors in CHIM/06.

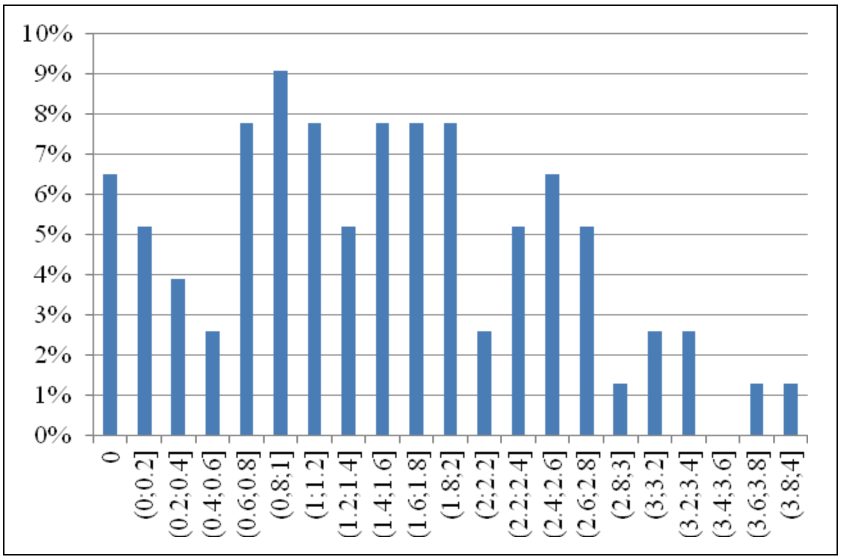

Figure 4. Frequency distribution for classes of annual publication rate P (2009-2013) for Italian professors in Inspection of food products of animal origin (VET/04).

Publication rates of fields within a discipline

As with the two examples above (CHIM/06 and VET/04), the publication rates in the various SDSs are never superimposable. Thus the calculation of the descriptive statistics for the SDSs provides useful benchmarks for the professors that work in them. Table 2 provides the statistics for all the SDSs in the Earth sciences discipline.

This UDA consists of a total of 12 SDSs with very different sizes in terms of national research staff, from a minimum of 17 professors in Applied geophysics (GEO/12) to a maximum of 137, in Palaeontology and palaeoecology (GEO/02). The intensity of publication is structurally very different. In Stratigraphic and sedimentological geology (GEO/03) only 2.2% of the professors (2 of 92) did not produce any publications over the five-year period under examination. On the opposite front there are 19 unproductive professors among the 121 of Physical geography and geomorphology (GEO/05), or 15.7% of the total. This SDS also registers the lowest average annual rate of publication, at 1.12 per year, followed by Structural geology (GEO/04), GEO/02 and Geophysics of solid earth GEO/11 (1.44, 1.48 and 1.49, respectively). In half the SDSs there is an average intensity of publication of 2 per year, with a peak in Applied geology GEO/06 (3.09). Clearly, among all those of the UDA, this SDS has the greatest publication rate: the distribution of the performances shows all values in the highest quartiles. The top 25% of professors (3rd quartile) produce on average more than 4 publications per year, with the absolute record being a professor who produces almost 18. The dispersion of the performances in all the SDSs, indicated by the variation coefficients in the last column of Table 2, results as greatest in GEO/03 and GEO/05, where the coefficient is above 1.

The analyses of the distributions for fractional counting of the publication rate (FP) (Table 3) provide a picture similar to that for full counting. The average intensity of collaboration evidently does not vary in a substantial way between the SDSs, and thus the differential of publication rates between the SDSs does not vary in going from a full counting approach to fractional counting.

Table 2. Descriptive statistics for intensity of annual publication rate P (2009-2013) for the SDSs of Earth sciences.

SDS	Research staff	Unproductive	I quartile	Median	III quartile	Max	Average	Std dev.	Variat. coeff.
GEO/01	93	3.2%	0.8	1.6	2.2	8	1.76	1.40	0.80
GEO/02	137	7.3%	0.6	1	2.20	6.4	1.48	1.25	0.84
GEO/03	92	2.2%	1	1.8	2.8	22	2.40	2.69	1.12
GEO/04	116	6.9%	0.6	1	2	4.8	1.44	1.21	0.84
GEO/05	121	15.7%	0.2	0.8	1.4	8.2	1.12	1.22	1.09
GEO/06	76	1.3%	1.55	2.6	4.05	17.8	3.09	2.51	0.81
GEO/07	82	2.4%	1	1.8	2.75	8.2	1.99	1.46	0.73
GEO/08	67	3.0%	1.3	2.4	3.5	10.6	2.69	2.03	0.75
GEO/09	63	6.3%	0.8	1.8	2.9	11.4	2.21	2.04	0.92
GEO/10	69	4.3%	1.2	1.8	2.4	10.2	2.14	1.82	0.85
GEO/11	41	2.4%	0.6	1.2	2	5.6	1.49	1.12	0.75
GEO/12	17	5.9%	0.8	1.6	2	4.6	1.75	1.34	0.77

Table 3. Descriptive statistics for intensity of annual publication rate FP (2009-2013) for the SDSs of Earth sciences

SDS	I quartile	Median	III quartile	Max	Average	Std dev.	Variat. coeff.
GEO/01	0.20	0.33	0.53	2.61	0.45	0.45	1.00
GEO/02	0.14	0.28	0.45	1.47	0.34	0.29	0.85
GEO/03	0.26	0.43	0.65	2.64	0.53	0.42	0.79
GEO/04	0.14	0.25	0.47	1.81	0.33	0.30	0.91
GEO/05	0.07	0.24	0.42	1.81	0.29	0.31	1.07
GEO/06	0.32	0.56	0.87	3.52	0.71	0.61	0.86
GEO/07	0.20	0.37	0.59	1.62	0.44	0.31	0.70
GEO/08	0.29	0.53	0.72	1.61	0.56	0.39	0.70
GEO/09	0.13	0.39	0.65	3.06	0.48	0.48	1.00
GEO/10	0.29	0.45	0.74	2.44	0.56	0.46	0.82
GEO/11	0.19	0.31	0.61	1.50	0.45	0.38	0.84
GEO/12	0.19	0.32	0.60	0.90	0.38	0.28	0.74

For the descriptive statistics of the full 192 SDSs investigated, we refer the reader to Annex 2[12] for the full counting, and to Annex 3[13] for fractional counting. Below, in Table 4, we show for each UDA the SDSs with minimum and maximum values of some of the above statistics

[12] www.iasi.cnr.it/laboratoriortt/TESTI/Altro/ISSI-ANNEX%202_P.pdf, last accessed on April 8, 2015
[13] www.iasi.cnr.it/laboratoriortt/TESTI/Altro/ISSI-ANNEX%203_FP.pdf, last accessed on April 8, 2015

of P (full counting). The data indicate substantial variability in the intensity of publication between the SDSs in all the UDAs. In Mathematics the percentage of unproductive professors varies from a minimum of 3.9% in MAT/09 (Operations research) and a maximum of 43.2% in MAT/04 (Complementary mathematics). Such substantial variations also occur in Medicine, with 1.1% unproductive professors in MED/08 (Pathological anatomy) and 45.5% in MED/02 (History of medicine). In Agricultural and veterinary sciences, VET/02 (Veterinary physiology) does not have any unproductive professors, while AGR/01 (Rural economics and valuation) registers a share of 45.5%. More contained heterogeneity in unproductive professors is seen in some other UDAs: certainly in Earth sciences, which we have already examined, but also in Biology. In this UDA the maximum incidence of unproductive professors (11.8% of the total professors) is seen in BIO/08 (Anthropology) and the minimum (1.2%) in BIO/15 (Pharmaceutical biology). The median intensity of annual publication also presents high variability between the SDSs of a UDA. In Mathematics the median ranges from 0.2 publications per year in MAT/04 (Complementary mathematics) to 1.8 in MAT/09 (Operations research). In effect the interval of variation of the median values is very substantial in almost all the UDAs. Within Industrial and information engineering, the median intensity of publication registered in ING-INF/06 (Electronic and information bioengineering) and in ING-INF/02 (Electromagnetic fields) is more than 40 times that registered in ING-IND/01 (Naval architecture). In Medicine the two extreme situations concern MED/02 (History of medicine) and MED/16 (Rheumatology): the median intensity of publication registered in the first SDS (0.2) is 1/25th of that for the second (5.0). The differences are more contained in Chemistry (2.0 vs. 3.4), Earth sciences (0.8 vs. 2.6) and Biology (1.1 vs. 3.3). The consistency of the outliers is also significantly different between the SDSs of a given discipline. In the Mathematics UDA, the most productive professor in absolute terms is one in INF/01 (Computer science), with an average of 28.6 publications per year, against the 3.6 of the most productive professor in MAT/04 (Complementary mathematics). In Medicine, a professor in MED/24 (Urology) registers a median of 76 publications per year over the five years examined; the most prolific in MED/47 (Nursing and midwifery) has barely 1.4 publications. In Industrial and information engineering the most prolific professor of ING-IND/01 (Naval architecture) authors an average of 1.4 publications annually, against the 33.2 of the most productive in ING-IND/34 (Industrial bioengineering). Finally, Physics FIS/01 (Experimental physics) includes a professor with an average of over 100 publications per year. In effect, this SDS consists of a range of subfields, including "high energy physics", where scientists regularly author hundreds of publications together with hundreds of co-authors. In this case (but not only in this case) a more opportune benchmark could be the distribution of the publication rate under the fractional counting method. Table 5 shows, for every UDA, the SDS with minimum and maximum values of the main statistics[14] of the fractional counting distributions. We see a level of superimposability with the data of Table 4, both in terms of the SDSs featured and for the intervals of variation in the main statistics of the SDSs, for each UDA.

[14] To avoid pointless duplication, the table does not show the incidence of unproductive professors, and instead provides statistics on average publication rate.

Table 4. SDSs with Min and Max values of descriptive statistics of intensity of annual publication P (2009-2013), for all UDAs.

	Unproductive (%)		Median		Max	
UDA*	Min	Max	Min	Max	Min	Max
1	3.9 (MAT/09)	43.2 (MAT/04)	0.2 (MAT/04)	1.8 (MAT/09)	3.6 (MAT/04)	28.6 (INF/01)
2	2.1 (FIS/04)	37.5 (FIS/08)	0.2 (FIS/08)	5.6 (FIS/01)	4.4 (FIS/08)	102.2 (FIS/01)
3	0.0 (CHIM/04)	8.6 (CHIM/11)	2.0 (CHIM/11)	3.4 (CHIM/02)	7.6 (CHIM/12)	66.2 (CHIM/08)
4	1.3 (GEO/06)	15.7 (GEO/05)	0.8 (GEO/05)	2.6 (GEO/06)	4.6 (GEO/12)	22 (GEO/03)
5	1.2 (BIO/15)	11.8 (BIO/08)	1.1 (BIO/02)	3.3 (BIO/15)	6.4 (BIO/08)	37.6 (BIO/12)
6	1.1 (MED/08)	45.5 (MED/02)	0.2 (MED/02)	5.0 (MED/16)	1.4 (MED/47)	76 (MED/24)
7	0.0 (VET/02)	42.0 (AGR/01)	0.2 (AGR/01)	2.8 (VET/06)	3.2 (AGR/06)	32.6 (VET/06)
8	5.8 (ICAR/03)	29.9 (ICAR/06)	0.2 (ICAR/06)	1.6 (ICAR/03)	2.8 (ICAR/05)	21.2 (ICAR/08)
9	0.0 (ING-IND/18)	50.0 (ING-IND/01)	0.1 (ING-IND/01)	4.4 (ING-INF/02 and ING-INF/06)	1.4 (ING-IND/01)	33.2 (ING-IND/34)

[*] 1 = Mathematics and computer sciences; 2 = Physics; 3 = Chemistry; 4 = Earth sciences; 5 = Biology; 6 = Medicine; 7 = Agricultural and veterinary sciences; 8 = Civil engineering; 9 = Industrial and information engineering

Table 5. SDSs with Min and Max values of descriptive statistics of intensity of annual publication FP (2009-2013), for all UDAs.

	Median		Average		Max	
UDA*	Min	Max	Min	Max	Min	Max
1	0.10 (MAT/04)	0.55 (MAT/09)	0.16 (MAT/04)	0.70 (MAT/07)	1.00 (MAT/04)	6.47 (MAT/02)
2	0.07 (FIS/08)	0.74 (FIS/03)	0.20 (FIS/08)	0.96 (FIS/03)	0.80 (FIS/08)	13.74 (FIS/03)
3	0.35 (CHIM/12)	0.70 (CHIM/02)	0.58 (CHIM/12)	0.83 (CHIM/02)	2.38 (CHIM/12)	17.60 (CHIM/08)
4	0.24 (GEO/05)	0.56 (GEO/06)	0.29 (GEO/05)	0.71 (GEO/06)	0.90 (GEO/12)	3.52 (GEO/06)
5	0.24 (BIO/08)	0.58 (BIO/15)	0.32 (BIO/08)	0.85 (BIO/15)	1.04 (BIO/08)	10.50 (BIO/12)
6	0.01 (MED/02)	0.84 (MED/16)	0.08 (MED/47)	1.18 (MED/16)	0.19 (MED/47)	13.28 (MED/11)
7	0.04 (AGR/01)	0.60 (AGR/15)	0.14 (AGR/01)	0.78 (VET/06)	0.65 (AGR/06)	9.14 (VET/06)
8	0.10 (ICAR/06)	0.48 (ICAR/08)	0.17 (ICAR/06)	0.73 (ICAR/08)	1.27 (ICAR/05)	6.85 (ICAR/08)
9	0.03 (ING-IND/01)	1.08 (ING-INF/02)	0.10 (ING-IND/01)	1.28 (ING-INF/02)	0.54 (ING-IND/02)	9.18 (ING-IND/19)

[*] 1 = Mathematics and computer sciences; 2 = Physics; 3 = Chemistry; 4 = Earth sciences; 5 = Biology; 6 = Medicine; 7 = Agricultural and veterinary sciences; 8 = Civil engineering; 9 = Industrial and information engineering

Conclusions

The great majority of the bibliometric indicators and the relative rankings lack fine-grained normalization of performance to the field to which the scientists belong. While bibliometricians intelligently field-normalize citations to account for the different citation behaviors across fields, they often close an eye when it comes to accounting for the different intensity of publication. At most they distinguish scientists as belonging to a few large disciplines, which cannot be sufficient if we accept the WoS as a true characterization, where scientific work is distinguished in 251 subject categories. Why would we normalize the citations for these 251 subject categories but then the scientists' performance for only a few disciplines? The answer is simple: in most cases the bibliometricians lack information about the field of research of each scientist under observation. Even at the national level the challenge of identifying the scientist's field is daunting, let alone for the task of international comparison.

Taking advantage of a particular feature of the Italian academic system, in this work we have provided descriptive statistics on the yearly publication rates of all Italian professors (over 30,000) in each of the 192 hard sciences fields, with both full and fractional counting method. Although the dataset refers to a specific nation, the very substantial size and the fine-grained field stratification certainly make it a useful reference system for the comparative evaluation of scientists in all the world. The only condition is that scholars recognize in which field of the Italian system the core of their scientific production falls. To this aim, in the Appendix, we have provided the reader with a conversion table, which establishes a link between SDSs and WoS subject categories, based on incidence of publications authored by Italian academics. Through this link, scientists outside Italy, knowing the distribution of their scientific production in the subject categories, can identify the corresponding SDS and select relevant statistic parameters as benchmark for comparative evaluation of their publication rates.

The statistics from the current analyses very clearly demonstrate the heterogeneity of publication rates even in the fields belonging to a single discipline. They help recognize the extent of distortions that occur when comparing the research performance of scientists from different fields, and could then serve as normalization factors to reduce such distortions when comparing the research performance of individuals, groups, or entire research organizations.

In future extensions of this work we could envisage a longitudinal analysis to assess the trends in publication intensity by field. We also know that publication rates of full, associate and assistant professors are different (Abramo, D'Angelo, & Di Costa, 2011). Gender differences in productivity have been demonstrated as well (Abramo, D'Angelo, & Caprasecca, 2009; Leahey, 2006; Fox, 2005; Pripiċ, 2002; Long, 1992). Because the composition of research staff by academic rank and gender varies across fields, a further extension of the analysis may then entail examining the differing publication intensity across fields by academic rank and gender.

References

Abramo, G., D'Angelo, C.A., & Caprasecca, A. (2009). Gender differences in research productivity: a bibliometric analysis of the Italian academic system. *Scientometrics*, *79*(3), 517-539.

Abramo, G., D'Angelo, C.A., & Di Costa, F. (2008). Assessment of sectoral aggregation distortion in research productivity measurements. *Research Evaluation*, *17*(2), 111-121.

Abramo, G., D'Angelo, C.A., & Di Costa, F. (2011). Research productivity: are higher academic ranks more productive than lower ones? *Scientometrics*, *88*(3), 915-928.

Abramo, G., D'Angelo, C.A., & Rosati, F. (2013a). The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences. *Journal of Informetrics, 7*(1), 198–208.

Abramo, G., D'Angelo, C.A., & Rosati, F. (2013b). Measuring institutional research productivity for the life sciences: the importance of accounting for the order of authors in the byline. *Scientometrics*, *97*(3), 779-795.

Butler, L. (2007). Assessing university research: A plea for a balanced approach. *Science and Public Policy*, *34*(8), 565-574.

D'Angelo, C.A., Giuffrida, C., & Abramo, G. (2011). A heuristic approach to author name disambiguation in large-scale bibliometric databases. *Journal of the American Society for Information Science and Technology*, *62*(2), 257–269.

Egghe, L. (2005). Relations between the continuous and the discrete Lotka power function. *Journal of the American Society for Information Science and Technology*, *56*(7), 664–668.

Fox, M.F. (2005). Gender, family characteristics, and publication productivity among scientists, *Social Studies of Science, 35*(1), 131–150.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359-375.

Kyvik, S. (1989). Productivity differences, fields of learning, and Lotka's law. *Scientometrics*, *15*(3-4), 205-214.

Leahey, E. (2006), Gender differences in productivity: research specialization as a missing link, *Gender and Society, 20*(6), 754-780.

Long, J.S. (1992), Measure of sex differences in scientific productivity, *Social Forces, 71*(1), 159-178.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16 (12), 317–324.

Moed, H.F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Moed, H.F., Burger, W.J M., Frankfort, J.G. & Van Raan, A.F.J. (1985). The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics*, *8*(3-4), 177-203.

Pontille, D. (2004). La Signature Scientifique: Une Sociologie Pragmatique de l'Attribution. Paris: CNRS Éditions.

Pripiċ, K. (2002). Gender and productivity differentials in science, *Scientometrics, 55*(1), 27-58.

RIN (Research Information Network) (2009). *Communicating Knowledge: How and Why Researchers Publish and Disseminate Their Findings*. London, UK: RIN. Retrieved April 8, 2015 from www.jisc.ac.uk/publications/research/2009/communicatingknowledgereport.aspx.

Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S.& Van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, *5*(1), 37–47.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., Van Leeuwen, T.N., Van Raan, A.F.J., Visser, M.S., & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*(12), 2419-2432.