

## Central indexes to the citation distribution: a complement to the $h$ -index

Pablo Dorta-González · María-Isabel Dorta-González

Received: 4 November 2010 / Published online: 23 July 2011  
© Akadémiai Kiadó, Budapest, Hungary 2011

**Abstract** The citation distribution of a researcher shows the impact of their production and determines the success of their scientific career. However, its application in scientific evaluation is difficult due to the bi-dimensional character of the distribution. Some bibliometric indexes that try to synthesize in a numerical value the principal characteristics of this distribution have been proposed recently. In contrast with other bibliometric measures, the biases that the distribution tails provoke, are reduced by the  $h$ -index. However, some limitations in the discrimination among researchers with different publication habits are presented in this index. This index penalizes selective researchers, distinguished by the large number of citations received, as compared to large producers. In this work, two original sets of indexes, the central area indexes and the central interval indexes, that complement the  $h$ -index to include the central shape of the citation distribution, are proposed and compared.

**Keywords**  $h$ -index · Citation analysis · Bibliometric indexes · Research career evaluation

**Mathematics Subject Classification (2000)** 62-07 Data analysis

### Introduction

There exists a general consensus among researchers about journal articles which are the most direct results of research, especially when journals with a selective process that

---

P. Dorta-González (✉)  
Departamento de Métodos Cuantitativos en Economía y Gestión,  
Universidad de Las Palmas de Gran Canaria, Gran Canaria, Spain  
e-mail: pdorta@dmc.ulpgc.es

M.-I. Dorta-González  
Departamento de Estadística, Investigación Operativa y Computación,  
Universidad de La Laguna, Tenerife, Spain  
e-mail: isadorta@ull.es

guarantees quality and originality are considered. Although expert opinion is believed to be the most appropriate method of valuing the contribution of an article to a specific field of knowledge, this system presents some limitations, such as the subjective character and its high cost. In this context, bibliometric indexes represent objective evidences that can be used to complement expert opinion.

It is known that some works of limited success are published by the best journals, and some works of great success are published in journals that are not top ranked according to the impact factor. Therefore, there is some rejection to evaluating the impact of a work by the impact factor of the publishing journal.

Most common indexes used to evaluate researchers are based on counting publications and received citations. The *number of publications* ( $N_p$ ) is a quantitative indicator that does not value the scientific advance of the contribution. As qualitative indicators able to assess the impact, influence or visibility of a research, the *total number of citations* ( $N_c$ ) and the *average citations per article* ( $n_c = N_c/N_p$ ) are used. However, although these indicators show the success of a scientific career in many cases, sometimes isolated successes accumulate a high percentage of the total number of citations. In addition, important biases are introduced by large collaborations that collect many citations derived from the work of a large number of researchers.

The *h-index* (Hirsch 2005) tries to solve these limitations. A researcher has an *h-index* when *h* of its publications have each received at least *h* citations, and the rest have *h* or less citations. The number of important articles pertaining to a researcher is estimated by this index, increasing their requirements at the same time as their value rises. Moreover, a lower bound  $h^2$  of  $N_c$  is provided. The value  $N_c$  is generally much greater than  $h^2$  (Hirsch has estimated between three and five times greater). This amount underestimates the citations of the *h* most cited articles (Hirsch core) and ignores articles with less than *h* citations. A correlation between the *h-index* and the success of a researcher appreciated by his peers has been obtained (Hirsch 2005), and the future success of a researcher could be predicted by this value (Hirsch 2007).

The *h-index* has been extensively studied (see reviews by Bornmann and Daniel 2007; Alonso et al. 2009; and the stochastic model by Burrell 2007a, 2009) and important mathematical properties have been fulfilled (Glänzel 2006). However, limitations have been found, some of which are shown below.

*This index depends on the scientific field and the number of collaborations.* It is not appropriate, therefore, to compare researchers from different scientific fields, due to different habits of publication, citation, and collaboration. This problem may be corrected, since the maximum *h* value obtainable in each field strongly correlates with the impact factors of the journals in the field, a reference *h-index* can be estimated in each scientific field (Imperial and Rodríguez-Navarro 2007). The *b-index* (Bornmann et al. 2007) is an alternative that indicates the number of articles in the 10% most cited publications in a field, considering ISI-ESI percentiles for example. Multiple authorships and self-citations have been investigated by Schreiber (2008a, b). Concerning the number of collaborators, the  $h_1$ -index (Batista et al. 2006) obtained dividing *h* by the average number of authors of these *h* articles, can be used.

*This index correlates with the number of publications.* The index tends to favor, therefore, those with more extensive scientific careers and is less effective among those with a low number of publications (Cronin and Meho 2006; Saad 2006; Van-Raan 2006). To differentiate between active and inactive researchers and compare scientists at

different stages of their careers, the growth rate  $h'(t)$  has been proposed, being  $t$  the number of years since the publication of the first article (Liang 2006; Burrell 2007b; Rousseau and Ye 2008). As an alternative, the  $h$ -index can be calculated for a certain period of time, instead of along the professional life of a researcher.

*All citations of the most cited articles are not considered in this index.* These most cited works contribute to the  $h$ -index, but their value is not affected by the number of times these articles are cited, since the tails of the citation distribution are not considered. These tails correspond to those publications that move away from the average impact, either because they have been highly cited (upper tail), or less cited (lower tail). Based on the definition of the Hirsch core, several authors have proposed new indicators. The  $g$ -index (Egghe 2006) considers all citations of the  $g$  most cited articles, and represents an average citation of these  $g$  articles. Once the articles have been sorted in decreasing order of citations,  $g$  is the largest value, such that the first  $g$  articles have at least  $g^2$  citations. As a matter of fact, the  $h$ -index and the  $g$ -index are special cases of a family of Hirsch index variants (Schreiber 2010). Similarly, the  $A$ -index (average citation) and the  $AR$ -index (considering the age of the articles) (Jin et al. 2007) have the particularity of taking into account the citations of the Hirsch core. However, as stated above, a heavy upper tail may correspond to the work of many authors included in large research lines that generate many citations.

*This index penalizes selective researchers*, that is, those producing a moderate number of high impact articles as opposed to large producers of moderate impact articles. Although this index has proven to be useful in identifying relevant researchers in a field, empirical evidence has shown it does not discriminate among researchers situated at intermediate levels and penalizes selective producers versus large producers (Costas and Bordons 2007). Cases with similar values of  $h$ , where citation curves are intersected, are especially questioned due to some researchers presenting higher levels of citations at the beginning of the curve and lower levels at the end. Additionally, this index is not *consistent* (Waltman and Van-Eck 2009). That is, the effect of incorporating a new paper with a given number of citations may be different between researchers, increasing the value of  $h$  in some cases and maintaining its value in others.

Finally, some variants have been published in order to improve the *accuracy* of the  $h$ -index: the tapered  $h$ -index (Anderson et al. 2008), the  $R_m$ -index (Panaretos and Malesios 2009), the  $w$ -index (Wohlin 2009), and the  $e$ -index (Zhang 2009). The  $h^2$  lower,  $h^2$  center, and  $h^2$  upper as well as the sRM value (Bornmann et al. 2010), represent new approaches providing additional information that increase the accuracy of the  $h$ -index.

In this work, a complement to the  $h$ -index that increases the consistency of the indicator and favors selective authors against large producers is presented. This approach also increases the accuracy of the  $h$ -index giving information about the shape of the citation distribution. The main difference with respect to the variants which have been proposed previously ( $g$ -index,  $A$ -index,  $AR$ -index,  $h^2$  upper, tapered  $h$ -index,  $R_m$ -index,  $w$ -index, and  $e$ -index) is that all of them are a function of all citations included in a core of most cited papers. By contrast, we establish an upper limit to the maximum number of citations considered for each publication in order to reduce the effect that isolated successes and/or large collaborations may have on the final result (as was pointed out by Hirsh). This upper limit can be modified without further changing the radius of the central index.

**Central indexes**

Given the published articles of an author in decreasing order of citations, let  $c_i$  be the number of citations received by the publication  $i$  ( $c_1 \geq c_2 \geq \dots \geq c_{N_p}$ ), and let  $N_c^j = \sum_{i=1}^j c_i$  be the aggregated number of citations of the  $j$  most highly cited papers.

The citation distribution is obtained plotting the number of citations versus the position of the articles. Connecting these points, the citation curve is obtained. The  $h$ -index is the largest integer number that satisfies  $c_h \geq h$ , that is,

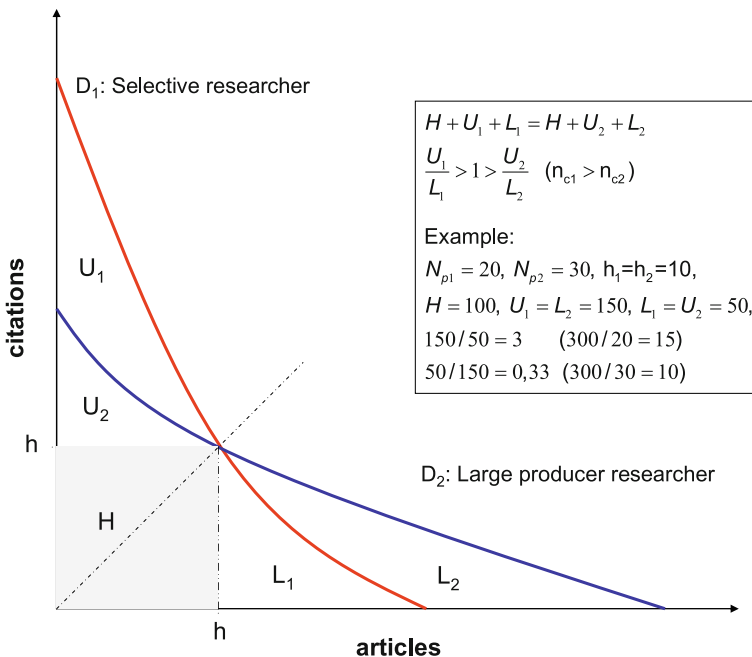
$$h = \max\{i \in Z : c_i \geq i\}.$$

Graphically, the integer part of the intersection point between citation curve and the first quadrant bisector is  $h$ . This is indicated in Fig. 1.

$H = h^2$  is a lower bound for the number of citations of those papers in the Hirsch core. The upper tail  $U$  is the excess citations received by the Hirsch core over the lower bound. The lower tail  $L$  is the number of citations received by those papers outside the Hirsch core. The following relationships are satisfied:

$$\begin{aligned} N_c &= H + U + L, \\ U &= N_c^h - H, \\ L &= N_c - N_c^h. \end{aligned}$$

The relative weight of the citation distribution tails is given by  $N_c/H$ . According to Hirsch (2005) estimations, if  $N_c/H < 3$  the tails of the distribution are light, while if



**Fig. 1** Two citation curves with the same  $h$ -index but different average citations per article

$N_c/H > 5$  the tails are heavy. The  $h$ -index penalizes those researchers who present heavy tails, especially those with a great tail ratio  $U/L$ .

An example with the citation curves of two researchers is shown in Fig. 1. The first researcher presents higher citation levels at the beginning and lower levels at the end of the curve. Therefore, two different profiles of researchers are appreciated, one more selective and another more massive in the production of papers. However, both scientists have the same  $h$ -index. A researcher may present less  $h$ -index than another, although it does not necessarily indicate the former presents a less successful career than the latter. The problem of discriminating between two distributions with similar  $h$ -index but significantly different distribution tail ratios is presented in Fig. 1. As can be appreciated, the higher the rate between tails is presented, the better average citations per article is obtained.

In the above cases, it seems reasonable to measure part of  $U$  and  $L$  in order to complement the  $h$ -index with the area around  $H$ . Thus, the discrimination capacity is increased. This idea allows us to introduce the following index.

Central area index

Let  $E$  ( $F$ ) be the upper (lower) area next to  $H$ , that is, the part of the upper tail  $U$  (lower tail  $L$ ) in the citation distribution closest to  $H$ . The lower area corresponds to those articles that will likely contribute to increasing the value of  $h$  in the future, since they are closer to the Hirsch core. The upper area includes those citations that will form part of  $H$  at the time the  $h$ -index increases its value. Therefore, it seems reasonable to include this area and increase, in this way, the discrimination capacity of the index.

The central area index of radius  $j$  is defined as the citations of the  $h + j$  most cited papers limited to the number of citations of paper  $h - j$ . That is, the citations of those papers in the Hirsch core, restricted by the citations of paper  $h - j$ , jointly with the citations of papers from  $h + 1$  to  $h + j$ . The geometrical representation is showed in Fig. 2.

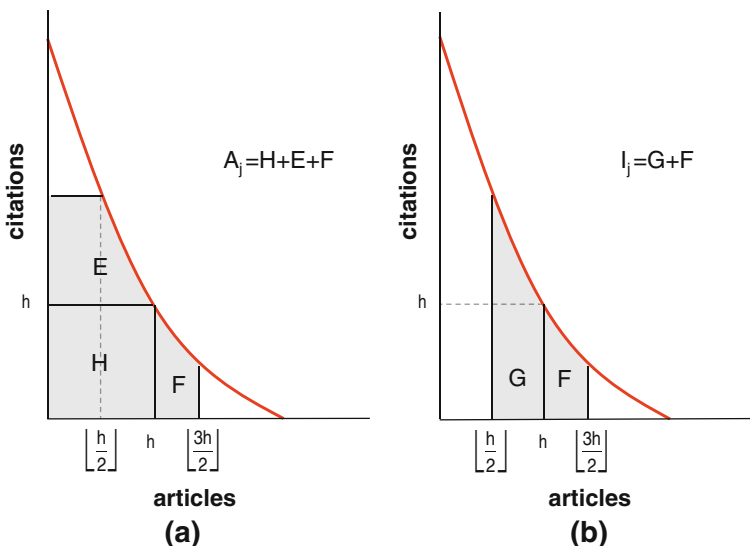


Fig. 2 Central area index (a) and central interval index (b) of radius  $j = \frac{h}{2}$

The arithmetic definition of the index of radius  $j$  is the following:

$$A_j = (h - j) \cdot c_{h-j} + \sum_{i=h-j+1}^{h+j} c_i, \quad j = 1, \dots, h - 1.$$

Note that  $A_{h-1} = N_c^{2h-1}$  includes the total upper tail  $U$ . Although the radius could be defined for  $j \geq h$ , in this case, it would only be adding part of the lower tail.

Figure 3 shows the central area index of radius  $j = \lfloor h/2 \rfloor$  (integer part) for two citation distributions. In the case of distribution  $D_i$  the central area index is  $A_j^i = H + E_i + F_i$ ,  $i = 1, 2$ .

As mentioned above, authors whose citation distributions have heavy tails are penalized by the  $h$ -index. However, the central area index of these authors grows faster than those with less heavy tails, increasing its capacity of discrimination.

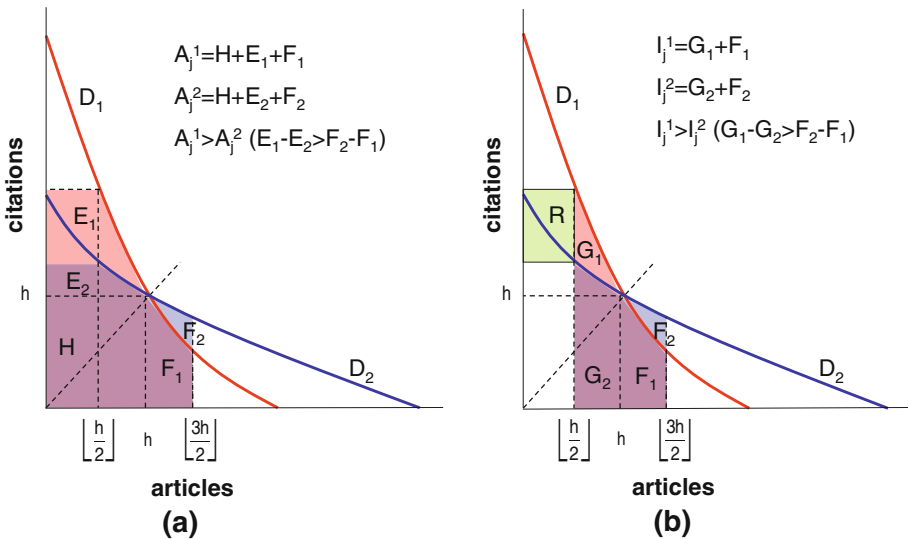
Selective researchers are also penalized by the  $h$ -index. However, the central area index solves this problem. For example, suppose a researcher has 10 publications, the least of which have 20 citations. Then  $h = 10$ , which represents only 100 citations. However,  $A_1 \geq 200$ , that is, twice the number of represented citations.

Central interval index

The central interval index of radius  $j$  is defined as the aggregated citations of the articles from  $h - j$  to  $h + j$ :

$$I_j = \sum_{i=h-j}^{h+j} c_i, \quad j = 1, \dots, h - 1.$$

That is, the citations at the interval  $[h - j, h + j]$ . Note that  $I_{h-1} = A_{h-1} = N_c^{2h-1}$ . The geometrical representation is shown in Fig. 2.



**Fig. 3** Comparison between the central area index (a) and the central interval index (b) of radius  $j = \lfloor \frac{h}{2} \rfloor$ , for two citation curves with the same  $h$ -index

The central interval index of radius  $j = \lfloor h/2 \rfloor$  (integer part) for two citation distributions is shown in Fig. 3. In the case of distribution  $D_i$  the central interval index is  $I_j^i = G_i + F_i$ ,  $i = 1, 2$ .

### Comparison among central indexes

Both indexes have the same area in the lower tail. However, significant differences between both indexes are appreciated in the upper tail. Thus while central interval indexes add citations of articles to the left of  $h$ , the central area indexes add zones of variable size in the upper tail.

Lets now look at differences between both indexes for two authors with the same  $h$ , one more selective than the other. Increasing the radius in one unit and reducing the comparison to the upper tail, where differences exist, the central interval index adds to the selective author the height of the rectangle  $R$ , as shown Fig. 3. However, the central area index also adds to the selective author the area  $R$ . For this reason, the central area index is more beneficial for selective authors.

As an author is more selective, the height of  $R$  increases and the area of  $R$  also increases in a greater proportion.

In the following section, an empirical application determines an optimal radius of the central indexes, obtaining the value of  $j$  that best describes the central shape of the citation distribution.

### Empirical application

The behavior of central indexes for researchers who have received the Price Medal is analyzed in this section. Data about these scientists was obtained from the ISI Web of Science database in February 2010. To estimate the predictive capacity of indexes for 5 and 10 years ahead, and their comparison with the  $h$ -index, the cited articles and the number of citations obtained in 1999, 2004, and 2009, have been considered. In order not to distort further analysis, especially regression analysis, only the 15 existing and currently productive scientists were considered.

The objective consists in obtaining  $A_j, I_j, j = 1, \dots, h - 1$ , at instant  $t$  and estimating the value of  $j$  (optimal radius) most correlated with  $A_k, I_k, k = j, \dots, h - 1$  at instant  $t + 1, t = 1, 2$  and  $t + 2, t = 1$ , that is, the future indexes.

Table 1 shows, for each author, the year of the first article published in the database, the total number of cited articles and the total citations in 1999, 2004, and 2009. This table also shows the evolution of the  $h$ -index.

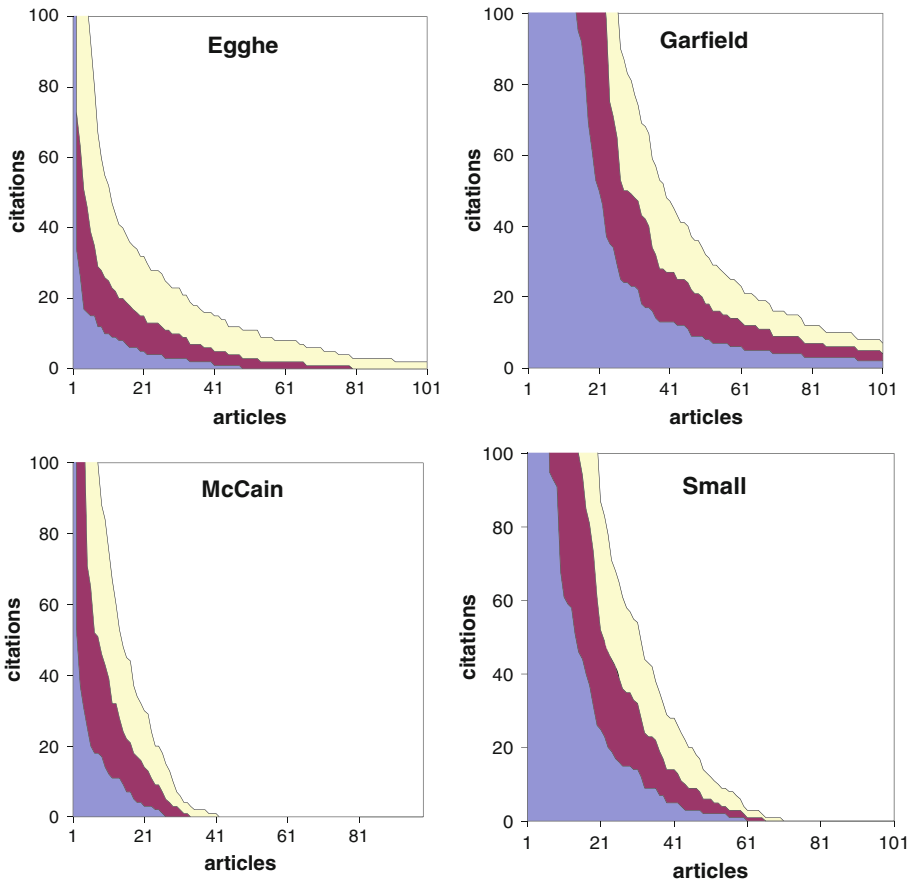
Figure 4 shows the citation curves of four researchers. Three curves are shown, the closest to the origin corresponds to 1999, followed by 2004 and the farthest to 2009. The value 100 has been taken as maximum only for clarity. This plot allows us to observe the evolution of the  $h$ -index, and also to distinguish between selective and large producer researchers. As an example, McCain and Small show a more selective behavior than Egghe and Garfield, respectively.

Production-impact scatter plots are presented in Fig. 5. As shown, linear correlation between the number of articles and the number of citations exists. Authors located above the regression line show a more selective behavior than those below this line. Thus, the more selective authors of the sample are Small and Garfield, respectively.

**Table 1** The production and impact of the researchers

Author	Year of the first article	Cited articles		Citations		<i>h</i> -index		$H = h^2$					
		1999	2004	2009	1999	2004	2009	1999	2004	2009			
Braun, T	1958	135	152	170	1,966	2,498	3,116	24	27	30	576	729	900
Egge, L	1978	47	78	122	299	571	1,277	10	12	18	100	144	324
Garfield, E	1954	163	174	180	3,687	4,298	5,294	25	26	29	625	676	841
Glänzel, W	1983	52	74	112	616	991	2,228	14	18	28	196	324	784
Ingwersen, P	1982	18	27	35	239	686	1,160	7	12	16	49	144	256
Leydesdorff, L	1981	38	54	107	235	477	1,541	9	13	21	81	169	441
McCain, KW	1983	25	32	40	328	761	1,261	11	15	17	121	225	289
Moed, HF	1985	31	50	64	386	804	1,608	12	16	22	144	256	484
Rousseau, R	1986	40	76	122	165	494	1,339	6	11	20	36	121	400
Schubert, A	1981	75	104	121	726	1,126	1,904	14	18	24	196	324	576
Small, H	1961	59	64	69	2,947	3,543	4,296	21	24	25	441	576	625
Van-Raan, AFJ	1976	47	64	78	488	909	1,750	13	17	24	169	289	576
Vinkler, P	1986	20	26	31	149	266	411	7	10	13	49	100	169
Vlady, J	1963	41	42	43	361	374	382	11	11	11	121	121	121
Zitt, M	1991	6	12	23	17	78	267	3	6	10	9	36	100
Average	1977	53.1	68.6	87.8	840.6	1,191.7	1,855.6	12.5	15.7	20.5	194.2	282.3	459.1

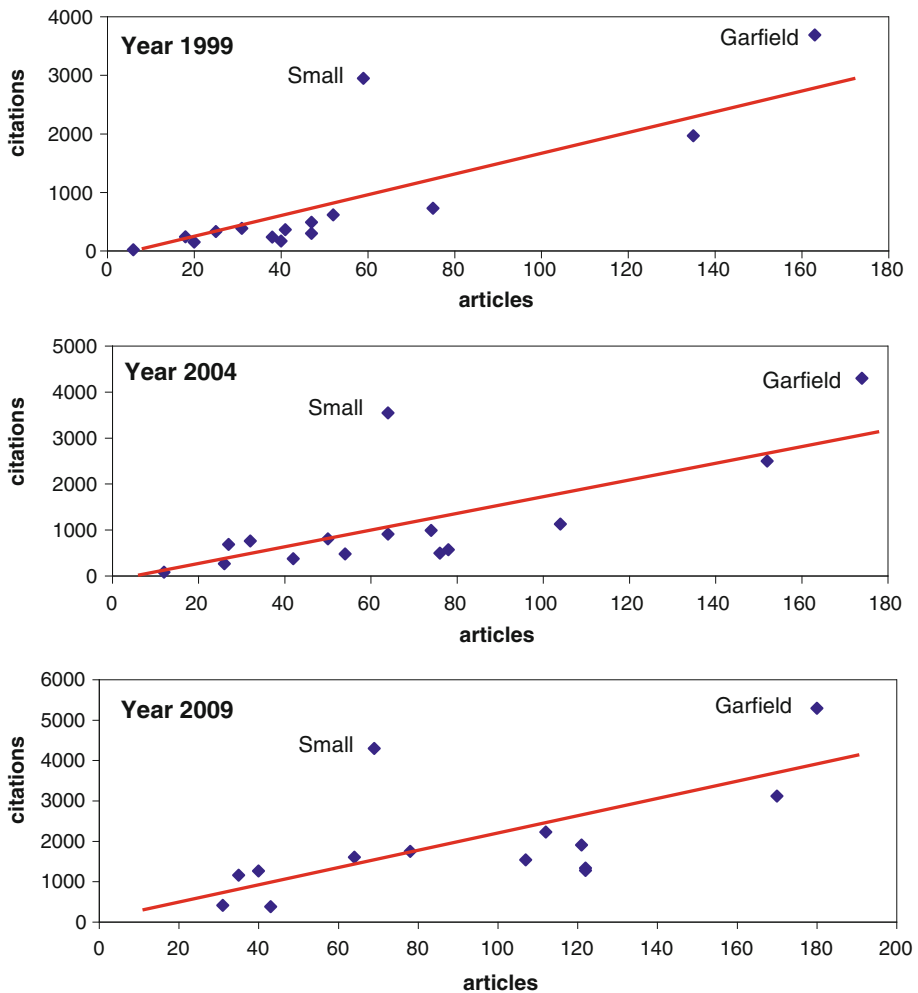




**Fig. 4** Citation curves for some researchers in 1999, 2004, and 2009

Table 2 shows the central indexes of the years considered. These indexes can be obtained up to a radius of 24 for some authors, but as the radius increases the number of data in each column of the table is reduced. Data has been shown until radius 10 to ensure that later the correlation coefficient is calculated with more than half of the sample data (at least nine out of 15). This table is useful in estimating future success. Lets see some examples using radius 7 as reference, approximately half the average  $h$ -index for the first period. It will be shown later that this indicator provides good estimations for 5 year predictions. As can be seen in the case of Leydesdorff, the area index varies from  $A_1 = 97$  to  $A_7 = 171$  in 1999, a significant increment that reveals the evolution of the  $h$ -index in following periods (2004 and 2009). Indeed, this author has an  $h$ -index of 9, 13, and 21, respectively. Something similar can also be seen in the case of McCain, among others. These examples suggest the area indexes obtained in a period, predict the increase in the following period.

Lets now see a comparison between two authors with the same  $h$ . McCain and Vlachy have  $h_{1999} = 11$ .  $A_6$  in the case of McCain (because  $A_7$  is not defined) is greater than the case of Vlachy, which estimates a higher future  $h$ -index; which holds true (15 vs. 11) in Table 1. Something similar happens with Ingwersen and Vinkler, for which  $h_{1999} = 7$ .



**Fig. 5** Production-impact scatter plot

$A_7$  in the case of Ingwersen is greater than in the case of Vinkler, which estimates a higher future  $h$ -index; which also holds true (12 vs. 10) in Table 1. The same conclusion can also be observed for a period of 10 years. Although area indexes have been taken as reference, something similar occurs in the case of interval indexes.

Now, as an example, let's consider a case where the discrimination capacity of the central indexes compared to the  $h$ -index is appreciated. Braun has  $h_{1999}$  greater than Small. However, the central indexes from a certain radius are higher for the later author. Attending to these indicators, the second author seems more selective, which is true according to total citations and the production-impact scatter plot.

Since the central index is an aggregation of citations, its representation with respect to the radius is an increasing function, as can be seen in Fig. 6. The first plot shows that McCain's area indexes are higher than those of Egghe, indicating the first author is more

**Table 2** Central indexes up to radius 10

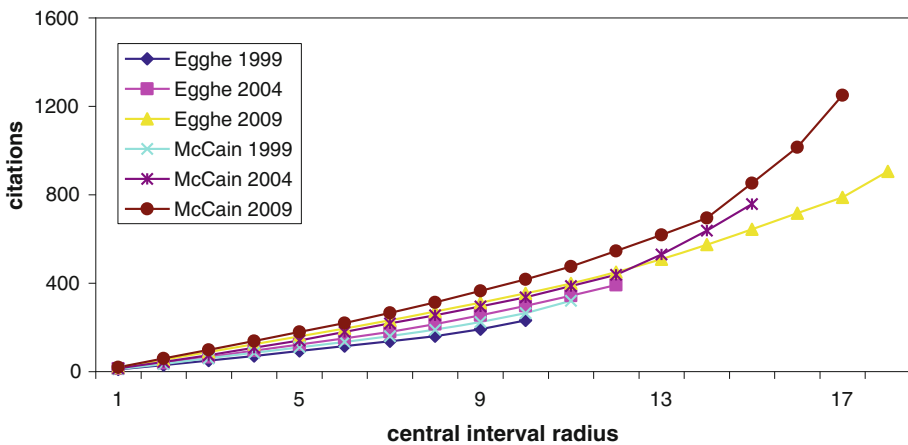
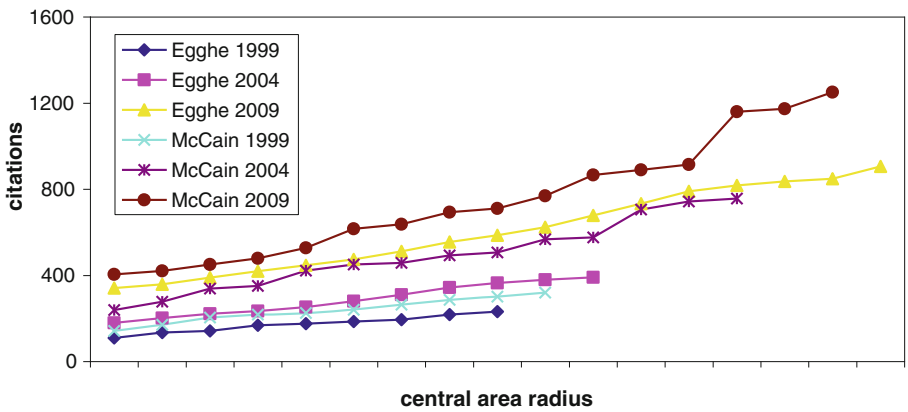
Author	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$
<i>1999</i>										
Braun, T	597	618	660	701	722	797	817	869	903	950
Egghe, L	109	134	142	168	175	185	194	218	231	–
Garfield, E	870	917	985	1,197	1,300	1,379	1,541	1,694	1,935	2,086
Glänzel, W	237	275	298	330	341	352	369	384	398	411
Ingwersen, P	70	106	155	166	211	230	–	–	–	–
Leydesdorff, L	97	112	119	136	155	161	171	179	–	–
McCain, KW	142	171	204	218	225	240	264	286	301	320
Moed, HF	167	186	221	228	249	267	283	293	297	315
Rousseau, R	53	63	69	80	88	–	–	–	–	–
Schubert, A	223	248	282	324	336	355	372	388	403	428
Small, H	543	581	688	806	869	944	987	1,066	1,164	1,187
Van-Raan, AFJ	182	194	226	255	265	274	283	302	315	335
Vinkler, P	87	96	119	124	132	140	–	–	–	–
Vlachy, J	153	162	187	196	222	230	246	262	275	286
Zitt, M	13	15	–	–	–	–	–	–	–	–
<i>2004</i>										
Braun, T	755	805	853	876	921	964	986	1,046	1,085	1,123
Egghe, L	179	201	222	234	252	280	310	344	365	380
Garfield, E	954	1,004	1,099	1,586	1,716	2,001	2,064	2,105	2,259	2,389
Glänzel, W	342	376	393	423	464	491	505	549	571	583
Ingwersen, P	213	224	332	356	369	417	473	575	611	636
Leydesdorff, L	178	209	228	254	262	284	298	306	318	328
McCain, KW	239	278	339	351	422	450	458	493	506	567
Moed, HF	318	348	375	388	434	447	469	513	585	605
Rousseau, R	142	152	162	179	195	209	222	231	248	258
Schubert, A	341	357	373	401	428	465	489	501	522	550
Small, H	623	666	686	726	897	1,041	1,091	1,122	1,211	1,281
Van-Raan, AFJ	306	337	353	394	444	469	482	504	525	545
Vinkler, P	148	163	169	193	198	219	229	247	254	–
Vlachy, J	153	171	188	197	223	236	252	268	279	294
Zitt, M	52	57	64	73	76	–	–	–	–	–
<i>2009</i>										
Braun, T	958	986	1,014	1,067	1,144	1,218	1,266	1,312	1,356	1,379
Egghe, L	341	358	390	419	447	474	511	555	586	623
Garfield, E	985	1,120	1,146	1,222	1,368	1,439	2,124	2,590	2,674	2,733
Glänzel, W	812	839	866	916	964	988	1,033	1,097	1,120	1,196
Ingwersen, P	301	342	394	417	504	523	577	586	649	733
Leydesdorff, L	461	557	577	611	627	703	732	760	785	798
McCain, KW	404	420	450	479	528	616	637	693	711	769
Moed, HF	548	609	629	703	791	827	876	892	919	969
Rousseau, R	438	456	491	508	540	570	586	614	630	666
Schubert, A	623	647	712	734	755	793	813	833	868	901

**Table 2** continued

Author	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$
Small, H	675	721	853	917	959	1,132	1,170	1,309	1,456	1,502
Van-Raan, AFJ	648	671	715	758	856	896	934	955	990	1022
Vinkler, P	219	241	259	276	292	300	324	344	353	361
Vlachy, J	153	173	190	199	225	238	254	270	281	299
Zitt, M	137	161	169	193	204	221	230	237	261	–
Author	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$
<i>1999</i>										
Braun, T	69	114	160	207	254	304	353	404	455	508
Egghe, L	29	50	70	93	115	137	160	192	231	–
Garfield, E	88	147	208	277	350	425	504	590	690	798
Glänzel, W	45	77	108	141	173	205	237	269	302	336
Ingwersen, P	25	46	77	110	163	230	–	–	–	–
Leydesdorff, L	27	46	64	84	107	129	153	179	–	–
McCain, KW	34	59	85	110	135	160	189	224	264	320
Moed, HF	37	60	85	109	135	162	191	221	249	284
Rousseau, R	21	36	51	68	88	–	–	–	–	–
Schubert, A	43	72	102	135	168	201	234	268	303	341
Small, H	68	113	161	214	269	328	389	454	526	597
Van-Raan, AFJ	39	64	91	119	146	172	198	226	255	287
Vinkler, P	27	44	65	86	110	140	–	–	–	–
Vlachy, J	36	58	82	106	132	158	186	216	249	286
Zitt, M	9	15	–	–	–	–	–	–	–	–
<i>2004</i>										
Braun, T	80	133	186	238	291	344	397	452	507	563
Egghe, L	39	66	94	122	150	180	214	254	297	343
Garfield, E	90	153	219	305	396	500	606	711	819	934
Glänzel, W	54	91	127	163	200	238	275	315	355	394
Ingwersen, P	43	71	108	146	183	227	281	359	447	543
Leydesdorff, L	35	59	84	110	136	164	193	222	252	282
McCain, KW	44	74	108	141	179	218	255	295	336	387
Moed, HF	52	88	123	157	194	231	269	310	357	405
Rousseau, R	34	56	78	101	125	149	174	199	228	258
Schubert, A	53	87	121	154	188	223	259	294	330	368
Small, H	73	120	166	213	267	327	387	447	511	579
Van-Raan, AFJ	51	85	119	154	191	229	266	304	343	383
Vinkler, P	36	58	79	103	126	153	181	216	254	–
Vlachy, J	36	59	83	107	133	160	189	220	253	294
Zitt, M	20	33	46	61	76	–	–	–	–	–
<i>2009</i>										
Braun, T	90	149	208	267	328	390	452	514	576	638
Egghe, L	53	88	124	159	195	232	271	312	354	399
Garfield, E	94	158	221	286	356	427	528	650	774	897

**Table 2** continued

Author	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$
Glänzel, W	84	139	194	249	304	358	413	470	526	584
Ingwersen, P	49	82	118	153	194	235	281	327	379	443
Leydesdorff, L	62	107	152	195	237	283	329	376	422	468
McCain, KW	59	98	138	179	220	266	313	365	417	475
Moed, HF	68	115	161	210	263	317	372	424	475	529
Rousseau, R	60	99	139	178	218	258	298	339	380	423
Schubert, A	73	122	172	221	269	317	365	413	462	511
Small, H	77	127	181	237	294	358	422	493	571	648
Van-Raan, AFJ	76	125	175	226	280	335	390	445	500	554
Vinkler, P	43	71	97	124	152	180	209	240	272	305
Vlachy, J	36	61	85	109	135	162	191	222	255	299
Zitt, M	33	56	79	103	128	155	182	211	261	–



**Fig. 6** Comparison of central indexes for two authors



selective than the second. Something similar can be seen with the interval index in the second plot.

Table 3 shows the linear correlation coefficients among indexes for 5 and 10 years. Matrices of order 10 to ensure the correlation coefficient is calculated with more than half of the researchers (at least nine out of 15) are shown. As can be seen, the area indexes for 1999 are strongly correlated with those for 2004, so they look like good estimators for 5 years. In all cases, correlations are higher than 0.94. The strongest correlations are located close to the main diagonal of the matrix. From the fifth element, all coefficients on the diagonal are greater than the correlation between  $h$ -indexes  $\text{corr}(h_{1999}, h_{2004}) = 0.977$ . As can be seen, all elements in column 7 are also higher than this. Therefore,  $A_7$  seems a good estimator for 5 years and the radius is about half the average  $h$ -index of the sample.

Area indexes for 1999 also show high correlations with 2009, although slightly lower than those mentioned in the previous paragraph, making them also good estimators for 10 years. All of the coefficients are greater than the correlation between the  $h$ -indexes  $\text{corr}(h_{1999}, h_{2009}) = 0.812$ . As can be seen, all elements in column 7 are higher than 0.9.

Finally, the area indexes for 2004 also present correlations with the year 2009. Most of the elements (including all of them in column 7) are higher than the correlation between  $h$ -indexes  $\text{corr}(h_{2004}, h_{2009}) = 0.889$ .

With respect to interval indexes, something very similar occurs. Correlations are also high in all cases. In order to better appreciate what indicators provide the best correlations, the differences between correlations for central indexes are also shown in Table 3. As can be seen, most of the elements of these matrices are positive, which means the correlations for the area index are greater than for the interval index (only 10 out of 165 items are negative).

## Conclusions

The  $h$ -index is a bibliometric indicator that attempts to measure the success of a researcher with just a part of the total amount of publications and citations. Due to not considering all production and impact, this index corrects biases of mass collaborations and punctual successes, which may not be significant in the researcher's career as a whole. However, different citation distributions, like those of a selective researcher and a large producer, may cause similar  $h$ -indexes, and in these cases, it is not possible to distinguish between these researchers using the  $h$ -index exclusively.

In this paper two complements to the  $h$ -index, the area and the interval indexes, have been proposed with the aim of increasing the capacity of discrimination among researchers with similar  $h$ , and improving the prediction of future successes. These indicators consider some areas that are larger for selective authors than for large producers. Thus, a problem described in the literature about the  $h$ -index, which penalizes selective researchers compared to large producers, is corrected.

Both central indexes are good estimators and correlations are generally higher for the area index than for the interval index. Moreover, a radius that well describes the shape of the citation distribution has been estimated empirically. This radius is about half the average  $h$ -index of researchers being evaluated.

Finally, we would like to point out that the area index is not considered a substitute, but a complement to the  $h$ -index, especially in an evaluation process where doubts among researchers might exist.

## Materials and methods

### Data collection

The underlying bibliometric data (publications and citations) in the empirical application were obtained from the Web of Science® database (published by Thomson Reuters—ISI, Philadelphia, PA, USA). Searches were conducted in the first week of February 2010. The data gathering about the scientists was as follows. We selected all journal papers of the Price Medallists who are still active (published to December 2009). For this purpose, the “Advanced Search” function to restrict results by author, document type, and time-frame, was used. In order to debug results the “View Distinct Author Sets” tool of the Web of Science database was utilized. This discovery function uses cross-citation data to show sets of papers likely written by the same person. A citations report from the databases Science Citation Index Expanded and Social Sciences Citation Index, was obtained. Then, output records with documents and citations were saved on a text file.

### Data analysis

The text file with documents and citations was imported using Microsoft Office Excel. To estimate the predictive capacity of indexes for 5 and 10 years ahead, three different periods were filtered using this software, 1999, 2004, and 2009. Only the documents with any citation in each period were considered. Tables and figures were obtained using Excel. Finally, indexes were calculated using C++ Programming Language.

**Acknowledgments** This research was partially financed by the Ministerio de Ciencia e Innovación, grant ECO2008-05589.

## References

- Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273–289.
- Anderson, T., Hankin, R., & Killworth, P. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, 76(3), 577–588.
- Batista, P. D., Campiteli, M. G., Kinouchi, O., & Martinez, A. S. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
- Bornmann, L., & Daniel, H. D. (2007). What do we know about the h-index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381–1385.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2007). The b-index as a measure of scientific excellence. A promising supplement to the h-index? *Cybermetrics*, 11(1), paper 6.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2010). The h index research output measurement: Two approaches to enhance its accuracy. *Journal of Informetrics*, 4(3), 407–414.
- Burrell, Q. L. (2007a). Hirsch’s h-index: A stochastic model. *Journal of Informetrics*, 1(1), 16–25.
- Burrell, Q. L. (2007b). Hirsch index or Hirsch rate? Some thoughts arising from Liang’s data. *Scientometrics*, 73(1), 19–28.
- Burrell, Q. L. (2009). On Hirsch’s Egghe’s and Kosmulski’s h(2). *Scientometrics*, 79(1), 79–91.
- Costas, R., & Bordons, M. (2007). The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro-level. *Journal of Informetrics*, 1(3), 193–203.
- Cronin, B., & Meho, L. I. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9), 1275–1278.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.



- Glänzel, W. (2006). On the h-index. A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67(2), 315–321.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Hirsch, J. E. (2007). Does the h-index have predictive power. *Proceedings of the National Academy of Sciences of the United States of America*, 104(49), 19193–19198.
- Imperial, J., & Rodríguez-Navarro, A. (2007). Usefulness of Hirsch's h-index to evaluate scientific research in Spain. *Scientometrics*, 71(2), 271–282.
- Jin, B. H., Liang, L. M., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855–863.
- Liang, L. (2006). h-index sequence and h-index matrix: Constructions and applications. *Scientometrics*, 69(1), 153–159.
- Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, 81(3), 635–670.
- Rousseau, R., & Ye, F. Y. (2008). A proposal for a dynamic h-type index. *Journal of the American Society for Information Science and Technology*, 59(11), 1853–1855.
- Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, 69(1), 117–120.
- Schreiber, M. (2008a). A modification of the h-index: The h(m)-index accounts for multi-authored manuscripts. *Journal of Informetrics*, 2(3), 211–216.
- Schreiber, M. (2008b). The influence of self-citation corrections on Egghe's g-index. *Scientometrics*, 76(1), 187–200.
- Schreiber, M. (2010). A new family of old Hirsch index variants. *Journal of Informetrics*, 4(4), 647–651.
- Van-Raan, A. F. J. (2006). Comparisons of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491–502.
- Waltman, L., & Van-Eck, N. J. (2009). A taxonomy of bibliometric performance indicators based on the property of consistency. Accessed January, 2011, from [http://repub.eur.nl/resource/pub\\_15182](http://repub.eur.nl/resource/pub_15182).
- Wohlin, C. (2009). A new index for the citation curve of researchers. *Scientometrics*, 81(2), 521–533.
- Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE*, 4(5), e5429.