This article is a translation of an article published in the *Chinese Journal of Science and Technical Periodicals* (2010), 21(1):33-37, (with an updated reference list).

A Perfect Hirsch-type Index? Experiences Using the Tapered *h*-Index (h_T)

HUA Ping-huan, WAN Jin-kun, WU Jun-hong

花平寰, 万锦堃, 伍军红

China Scientometrics and Bibliometrics Research Center of Tsinghua University, Beijing 100084, China

Abstract: The h_T -index proposed by Anderson *et al.* includes all citations, yet scoring them in a way that is comparable with the *h*-index. This new index is superior to the *h*-index proposed by Hirsch because: 1) it takes all citations into account; 2) it has no problems with special cases, such as situations with only a few highly cited papers; or the case that, because of a lack of publications the *h*-index is set artificially equal to the number of publications; 3) it has a higher resolution than the *h*-index. In order to illustrate our point the h_T – indices of more than 6,000 Chinese academic journals have been calculated and results are discussed.

Keywords: *h* - index, tapered *h* -index, evaluation, journals, bibliometric indicators, informetrics

Introduction

Many bibliometric indicators are used to evaluate the achievement of an individual or a group of researchers. Among them the *h*-index proposed by Hirsch (2005) has many advantages, because it combines quantity (number of papers) and quality (citations) in an intelligent way and is moreover easy to calculate. For these reasons the *h*-index immediately attracted a lot of attention (Ball, 2005; Panaretos & Malesios, 2009).

Another advantage of *h*-index is its robustness, i.e. it is insensitive to one or two highly cited articles. However, some scholars consider this a disadvantage and consequently a long series of “improvements” have been proposed such as the g-index (Egghe, 2006), the *A*-index (Jin, 2006), the *R*-index (Jin et al. 2007) and the hg-index (Alonso et al., 2010). Recently Anderson et al. (2008) proposed a new Hirsch-type index called the tapered *h*-index, denoted as h_T. This index takes all citations into account, yet the contribution of the *h*-core is not changed.

In the case that it is possible to collect all sources and all items (for instance: all articles and all their citations) we consider the h_T-index the perfect Hirsch-type index. As we were in the position to have this information we calculated the h_T-index of more than 6,000 Chinese academic journals. In this paper, our results, as a practical example, are used to discuss the advantages of the h_T-index.

The concept and calculation of the h_T -index

According to Anderson et al. (2008), the calculation process of the h_T-index is performed as follows. Consider a scientist who has 5 publications which, when ranked, have 6, 4, 4, 2, 1 citations. This publication output can be represented by a Ferrers tableau, where each row represents a partition of the total 17 citations received by the 5 publications papers (Figure 1). The largest completed (filled in) square of points in the upper left hand corner of a Ferrers tableau is called the Durfee square. The *h*-index is equal to the length of the side of the Durfee square (in the case of Figure 1, h = 3). The Durfee square is the 3-by-3 square indicated by a dashed line. The tapered *h*-index, denoted as h_T, is the sum of all values in the Ferrers tableau. Note that we prefer the term Ferrers tableau to the term Ferrers graph, as a Ferrers tableau as represented in Fig. 1, is not a graph in the graph or network theoretic sense. Note also that publications with no citations are not represented in a Ferrers tableau. Recently Egghe (2010b) studied Ferrers graphs and Durfee squares in an informetric setting.

If an author has a single paper that has one citation, his h-index value is h = 1. Subsequently, h = 2 is achieved with two papers each with two citations. To move from h = 1 to h = 2, an additional 3 citations are required, one for the first paper and two for the second paper. In turn, moving from h = 2 to h = 3 requires a further 5 citations, reaching a 3, 3, 3 partitioning of the nine citations in the Ferrers tableau. This articles and their citations form a square in the Ferrers tableau, called the Durfee square. Following this scheme illustrated in Fig. 1, one obtains the h-index and this in such a way that each citations really contributes. Thus, the single citation in the Durfee square of side one has a score of 1, the three additional citations in the Durfee square of side 2 each score 1/3, and the five additional citations in the Durfee square of side 3 each score 1/5. Summing the relevant citations, scores of 1, 2, 3 are achieved for Durfee squares whose width is 1, 2, 3, matching the *h*-index.

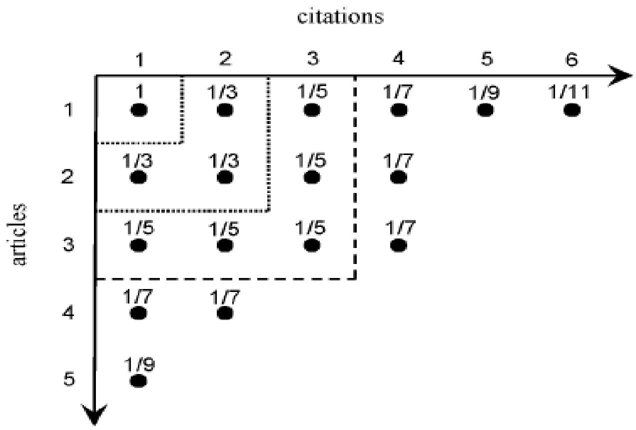


Figure 1 Example of a Ferrers tableau of an author's citations

This notation immediately suggests a new index, denoted as h_T, which has the property that each additional citation increases the total score, whether or not it lies within the *h*-index Durfee square. The score of any citation in a Ferrers tableau is now given by $1/(2L - 1)$, where L is the length of side of a Durfee square whose boundary includes the citation in question. The additional citations that fall outside the Durfee square (of side 3) in Figure 1 can now be counted, the five papers contributing scores of 1.88, 1.01, 0.74, 0.29 and 0.11, leading to a total score for h_T of 4.03.

In mathematical terms, the most cited paper in a given list, with n_1 citations, generates a score, $h_{T(1)}$, of:

$$h_{T(1)} = \sum_{i=1}^{n_1} \frac{1}{2i-1} \tag{1}$$

The resulting sum is 2.13 for 10 citations, 3.28 for 100 citations, 4.44 for 1,000 citations and 5.59 for 10,000 citations.

The paper ranked second in the list scores 1/3 for its first citation, and then 1/3, 1/5, 1/7 etc., for further citations as for the top-ranked paper. Now, if an author has N papers with associated citations n_1, n_2, n_3, …, n_N (ranked in descending order as in a Ferrers tableau), the h_T score for any single paper ranked j in the list (with n_j citations), denoted as $h_{T(j)}$, is:

$$h_{T(j)} = \frac{n_j}{2j-1}, n_j \le j \quad (2) \quad and \quad h_{T(j)} = \frac{j}{2j-1} + \sum_{i=j+1}^{n_j} \frac{1}{2i-1}, n_j > j \tag{3}$$

The tapered *h*-index, h_T, for the whole list is then calculated by summing over all papers:

$$h_T = \sum_{j=1}^{N} h_{T(j)} \tag{4}$$

Because the scoring method for h_T is basically the same as that for h, a direct comparison of the two measures of output is allowed. We further note that $h_T = h + \Delta h$, where Δh is the sum of all citation values outside the Durfee square in the Ferrers tableau.

The h_T-index of Chinese academic journals

The CNKI (China National Knowledge Infrastructure) is an electronic platform launched in June 1999 by the Tsinghua Tongfang Knowledge Network Technology Company, a spin-off company of Tsinghua University, located in Beijing. CNKI contains several databases such as the China Academic Journals Full-text Database (CAJ), the China Proceedings of Conference Full-text Database (CPCD) and the China Doctor/Master Dissertations Full-text Database (CDMD).

CNKI moreover publishes the *Chinese Academic Journals Comprehensive Citation Report* on the basis of the *Chinese Academic Journals Comprehensive Evaluation Database* in which more than 6,000 Chinese academic journals are incorporated. In this report twelve bibliometric indicators are provided, including the total number of citations, the journal impact factor, the number of published articles, the journal immediacy index, the *h*-index, and the download immediacy index (Wan & Xue, 2008; Wan et al., 2010).

Based on the data of the *Chinese Academic Journals Comprehensive Evaluation Database* of the CNKI, the h_T-indices of more than 6,000 Chinese academic journals have been calculated. The citation count is received in the duration of 2003-2007 by Chinese academic journals published in the same years. This is another illustration of the fact that h-type indices can be calculated for all possible citation and publication windows (Liang & Rousseau, 2009).The results are shown in Figure 2. It was found that the peak value for the h_T-index appears in the interval 10-15. The average h_T value is 17.13 and the median is 14.92.

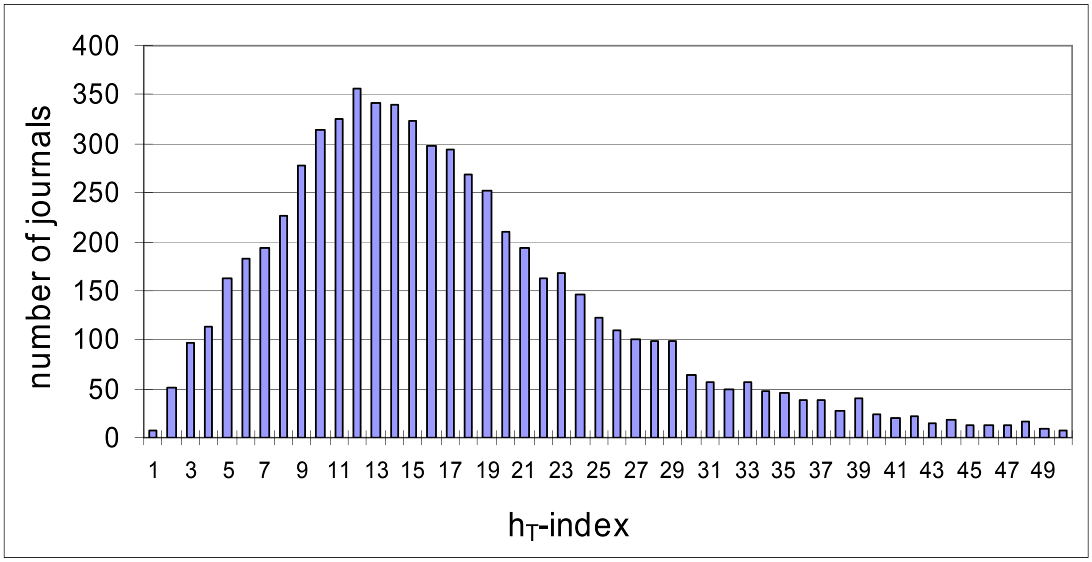


Figure 2. The h_T-index distribution of 6,573 Chinese academic journals Discussion

1) A plot of h_T versus h is shown in Figure 3. The Pearson correlation coefficient being equal to 0.99 indicates that the correlation between the two indices is very high. Using a least-squares straight line fit, we obtained the following equation:

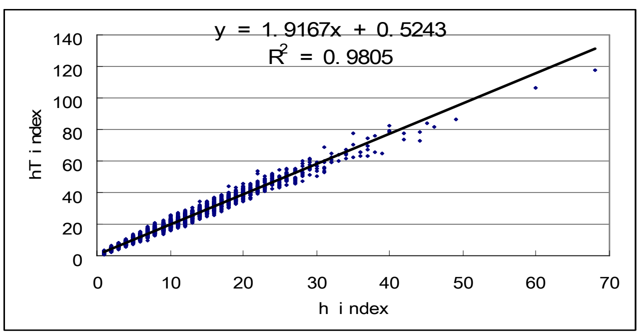


Figure 3 Chinese academic journals: plot of h_T versus h

$$h_T = 1.9167\,h + 0.5324 \qquad (5)$$

The slope of this regression line is 1.92, which is higher than Anderson's (1.73). This means that there are more citations beyond the Durfee square in the case of Chinese academic journals.

2) Moreover, we find the following relationship between the total number of citations ($N_{c,tot}$) and the h_T-index:

$$N_{c,tot} = 2.4(h_T)^2 \qquad (6)$$

This fact means that. Although the h_T -index includes all citations, it is not obtained by simply moving the citations beyond the Durfee square to form a Durfee square with side length equal to h_T. If this were the case $N_{c,tot}$ would be equal to $(h_T)^2$.

We further observe that the relation between $N_{c,tot}$ and the h-index is

$$N_{c,tot} = 9.8h^2 \qquad (7)$$

In connection with equation (7) Hirsch (2005) wrote: '*The relation between $N_{c,tot}$ and h will* *depend on the detailed form of the particular distribution, and it is useful to define the proportionality constant α as*

$$N_{c,tot} = \alpha h^2 \quad (8)$$

I find empirically that α ranges between 3 and 5.' In our case, the value of α is obviously much higher than Hirsch's, meaning that there are more citations located in the 'long tail' otherwise truncated in the calculation of *h*-index.

Bringing our results together we see that from formula (6) we can get $(h_T)^2 = N_{c,tot}/2.4 \approx 0.4\, N_{c,tot}$; and from formula (7) we can get $h^2 = N_{c,tot}/9.8 \approx 0.1\, N_{c,tot}$. Assume that $N_{c,tot}$ has 1,000 units, and that 100 units (10% of $N_{c,tot}$) are used to form the Durfee square h^2 (h = 10), then only 400 units are needed to form an equivalent square with area equal to $(h_T)^2$ (see Figure 4). This illustrates the charm of tapered scoring.

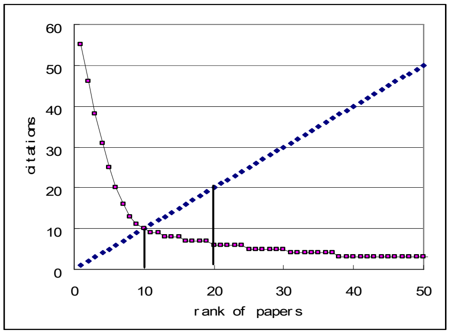
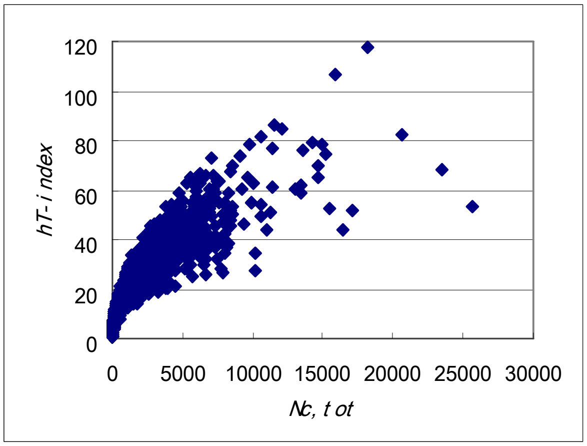


Figure 4. The relationship between *h*, h_T and citations

3) The relationship between $N_{c,tot}$ and h_T-index of more than 6,000 Chinese academic journals is shown in Figure 5 (a), from which we can see that the same value for $N_{c,tot}$, may lead to a whole series of h_T-values. This can be explained by the fact that different journals have different citation distributions.

The relationship between $N_{c,tot}$ and the h_T-index for 398 Chinese academic journals all with an *h*-index equal to ten, is shown in Figure 5 (b). This figure illustrates that a range of h_T-index values can be obtained from the same $N_{c,tot}$, even though their *h*-indices are also the same. For the same *h*-indices (namely equal to ten), the calculated h_T-index ranged between 16.26 and 25.92 (accurate to two decimal places), with a corresponding range in $N_{c,tot}$ of 437 to 6,634. For the 398 Chinese academic journals, the h_T-indices have 288 different values. This fact indicates that the h_T-index has a higher resolving power than the original *h*-index.



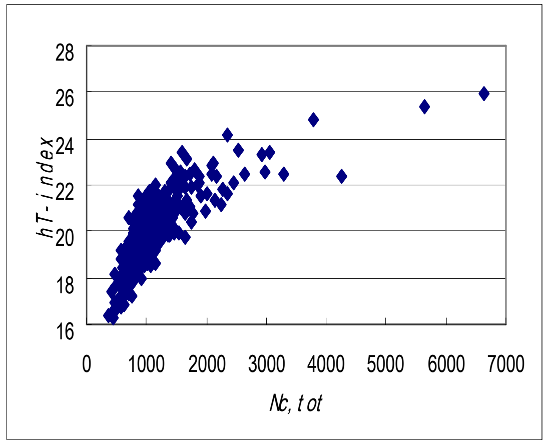


Figure 5. The relationship between $N_{c,tot}$ and h_T-index

Figure 6 compares the citation distribution curves of the journals *Special Wild Economic Animal and Plant Research* (*SWEAPR*) and *China Science and Technology Information* Fhirsch wrote

(*CSTI*). These two journals have the same *h*-index, but their h_T-indices are 16.26 and 25.92 respectively, with corresponding $N_{c,tot}$ values of 437 and 6,634. We consider that, using the h_T-index rather than the *h*-index, to be a more reasonable way of representing the impact of these two journals

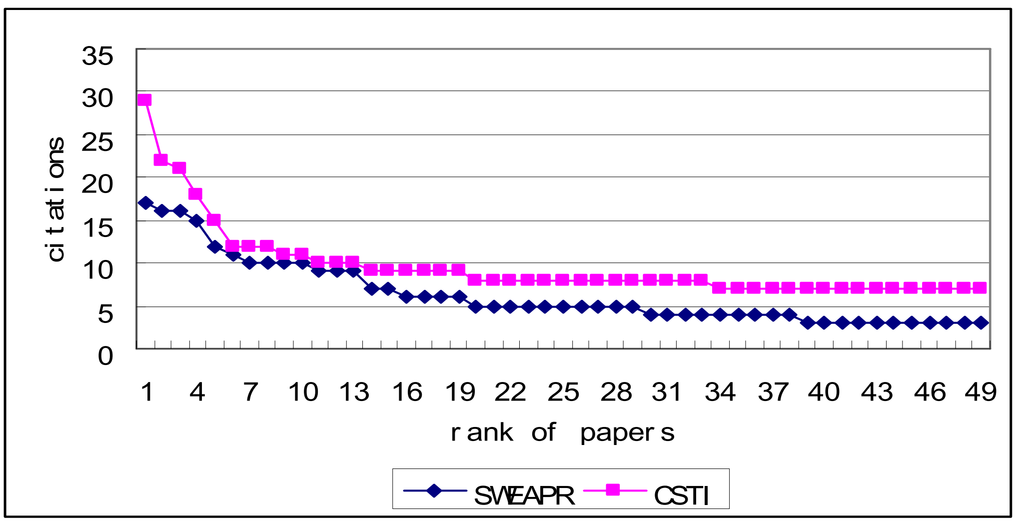


Figure 6. A comparison between the citation distribution curves of the journals *SWEAPR* and *CSTI*

4) Anderson et al. (2008) wrote: "*Perhaps the greatest criticism that can be levelled at the h-index is that an individual's score cannot exceed their number of papers (and will usually be much less), in which case the thousands of citations that typically accompany leading articles in a given field are left effectively unscored*".

Recall that in case the *h*-index is equal to the number of published articles, the number of citations of the h^{th} ranked article is called a pseudo *h*-index (Rousseau *et al.*, 2008), see the data of E and F in Table 1 of (Vinkler, 2007**)** as an example. In such cases, the h_T-index can give better evaluation results.

Table 1. Theoretical examples: sets of papers published by authors A, B, etc.

Paper no	A	B	C	D	E	F
1	100	9	10	50	9	120
2	98	8	10	50	8	110
3	98	8	10	50	7	100
4	97	6	10	50	6	90
5	96	5	10	50	5	80
6	4	4	10	50		
7	3	4	10	50		
8	2	3	10	50		
9	1	2	10	50		
10	1	1	10	50		
P	10	10	10	10	5	5
$N_{c,tot}$	500	50	100	500	35	500
h	5	5	10	10	5	5
h_T	13.27	6.89	10	18.05	5.79	12.46

5) Like other bibliometric indicators such as *h*-index, impact factor, *etc.*, the values of the h_T-index are subject dependent. The average h_T-indices of some subject areas are shown in Figure 7. From this figure we can see that the h_T-index of psychology journals is the highest (39.29), while the music and arts journals have the lowest average h_T-index (only 8.34). Therefore, comparisons of various journals based on the h_T-index must be restricted to journals in the same field. As an example we listed some indicators of mathematics journals in Table 2. In this Table, journals are ranked according the values of their *h*-index. Clearly the h_T-index has a higher resolving power than the *h*-index. Remarkably, the rank according to the *h*-index and according to the h_T-index may be different (see e.g. the *Journal of Mathematics* and the *Chinese Journal of Engineering Mathematics*).

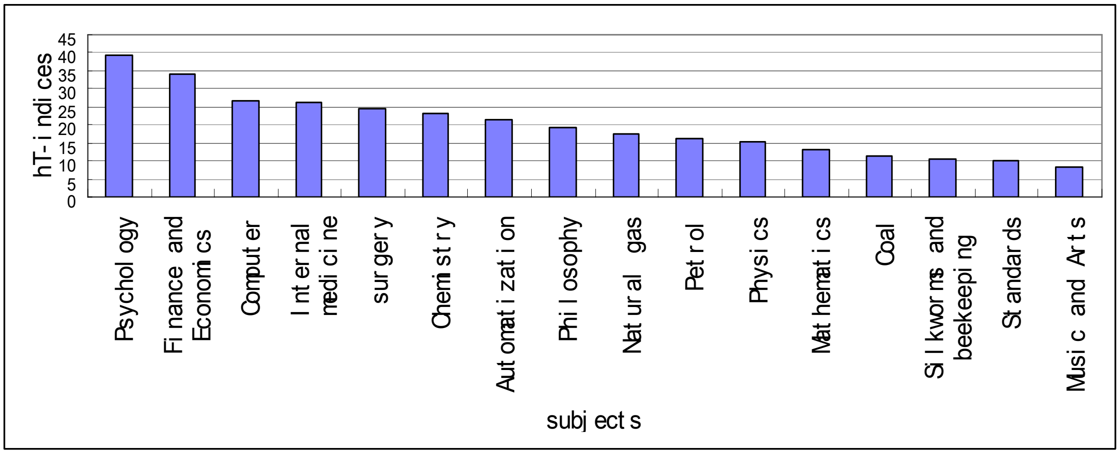


Figure 7 The average h_T-index in various subjects

Table 2. Indicator values for mathematical journals

Journal name	$N_{c,tot}$	h	h_T
Mathematics in Practice and Theory	1218	16	30.72
Acta Mathematica Sinica	1516	12	23.10
Acta Mathematicae Applicatae Sinica	705	11	19.04
Fuzzy Systems and Mathematics	721	11	18.94
Chinese Journal of Engineering Mathematics	571	10	20.07
Journal of Mathematics For Technology	580	10	18.82
Mathematica Numerica Sinica	441	10	17.63
Journal of Biomathematics	650	9	17.81
Chinese Journal of Applied Probability and Statistics	383	9	16.59
Science in China (Series A)	1017	9	16.53
Journal of Mathematics	342	9	14.99
Acta Mathematica Scientia	523	8	17.57
Bulletin of Mathematics	474	8	16.33
Chinese Annals of Mathematics	503	8	16.30
Mathematical Theory and Applications	211	8	14.96
Applied Mathematics A Journal of Chinese Universities	286	8	14.92
Numerical Mathematics A Journal of Chinese Universities	287	8	14.68
Mathematics in Economics	166	7	14.06
Mathematica Applicata	475	7	14.04
Advances in Mathematics	420	7	13.86
Science in China (Series A)	752	7	13.68
Studies In College Mathematics	228	6	12.20
Journal of Mathematical Research and Exposition	355	6	11.94
Journal of Mathematical Study	150	6	9.29
Acta Mathematica Scientia	221	5	10.03

Acta Mathematicae Applicatae Sinica	225	5	9.70
OR Transactions	160	5	9.68
Pure and Applied Mathematics	163	5	9.58
Annals of Differential Equations	130	5	9.34
Acta Analysis Functionalis Applicata	95	5	9.11
Communication on applied mathematics and computation	121	5	8.80
Chinese Annals of Mathematics	206	5	8.73
Journal of Computational Mathematics	235	4	8.95
Applied Mathematics A Journal of Chinese Universities	102	3	6.79
Numerical Mathematics A Journal of Chinese Universities English Series	24	3	4.47
Northeastern Mathematical Journal	136	2	4.76
Chinese Quarterly Journal of Mathematics	127	2	3.46
Approximation Theory and Its Applications	31	2	3.17

6) About the shortcomings, Anderson (2008) said: "*h has its advantages, most notably ease of determination and conceptualisation. An h-index of 15 immediately conveys that an individual has 15 papers, each with at least 15 cites. Yet if instead one proudly announces an h_T score of 30, it is much less clear what this means* "

According to our experience, the calculation of h_T-indices is not an impossible task: a dedicated FoxPro program needed just one day for the calculation of more than 6,000 Chinese academic journals.

About the conceptualization, we suggest to describe the h_T-index as follows: an individual has h_T index equal to h_T if he/she has h_T papers that each have h_T citations (even though the h_T is not an integer), where the area under the citation distribution curve (Figure 8(a)) has been redrawn as the area of an equivalent Durfee square (Figure 8(b)).

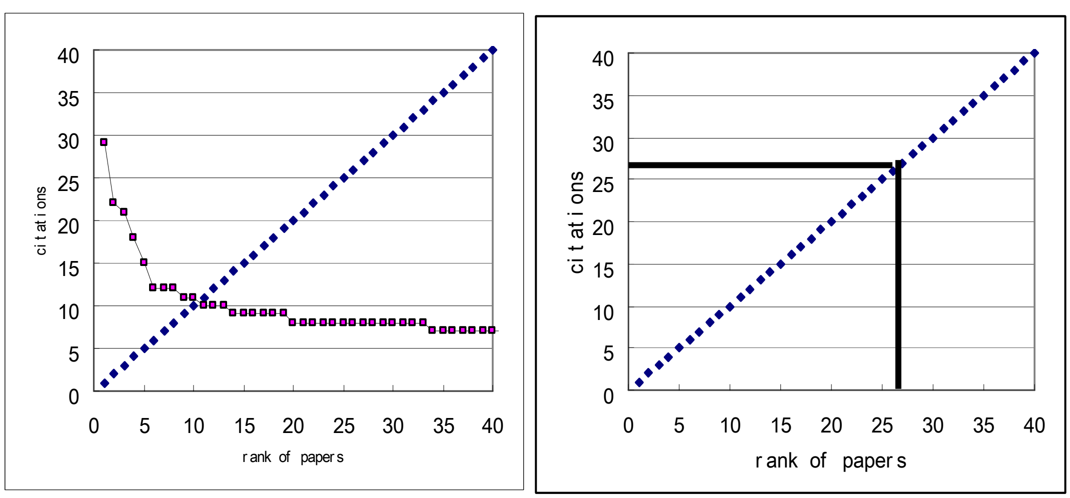


Figure 8. The citation distribution curve (a) and a corresponding equivalent Durfee square (b)

Conclusions

We recall from Liu and Rousseau (2009) and Egghe (2010a) the following list of advantages and disadvantages of the h-index and compare with the h_T - index.

Advantages of the h-index

- It is a mathematically simple index.

  This is not so for the h_T index, but a dedicated software program can take care of this problem.
- It encourages high quality (or at least highly visible) work.

  This is somewhat less the case as all publications are taken into account.
- The *h*-index can be applied to any level of aggregation.

  This also holds for the h_T index.
- It combines two types of activity (in the original setting this is citation impact and publications).

  Also true for the h_T index.
- Increasing the number of publications alone does not have an immediate effect on this index.

  This may have an effect if these publications are cited. Yet because of the way the h_T index is calculated, this effect is very small.
- It is a robust indicator in the sense that small errors in data collection have no or little effect.

  Errors have a somewhat larger effect on the h_T index.
- Single peaks (top publications) have hardly any influence on the h-index.

  They have a larger effect on the h_T index. In this sense the h_T index behaves more like the *g*-index (Egghe, 2006).
- In principle, any document type can be included.

  This is the same for the h_T index
- Publications that are hardly ever cited do not influence the h-index.

  Such publications have a negligible influence on the h_T index.

Disadvantages which the *h*-index has in common with other citation measures.

- Like most pure citation measures it is field-dependent, and may be influenced by self-citations.
- There is a problem finding reference standards.
- There exist many more versatile indicators for research evaluation.
- It is rather difficult to collect all data necessary for the determination of the h-index. Often a scientist's complete publication list is necessary in order to discriminate between scientists with the same name and initial. We refer to this problem as the precision problem.

By definition the same problems apply to the h_T index. Yet, because of the precision problem we propose to apply the h_T index only to journals, where this problem does not occur (or only in a minor way, i.e. when the name of the journal or its publication year has been misprinted).

Disadvantages typical for the *h*-index.

- The *h*-index, in its original setting, puts newcomers at a disadvantage since both publication output and observed citation rates will be relatively low. In other words, it is based on long-term observations.

  It has already been observed that the *h*-index and the h_T index can be applied to any publication and citation window. Then this disadvantage disappears for both.
- The index allows scientists to rest on their laurels since the number of citations received may increase even if no new papers are published.

  Again this is not a disadvantage as this shows the continuing influence of that scientist (or journal). A same remark holds for the h_T index.
- The *h*-index is only useful for comparing the better scientists in a field. It does not discriminate among average scientists.

  As the h_T index takes all publications and citations into account it does not have this disadvantage. It has indeed a higher resolution.
- The *h*-index lacks sensitivity to performance changes: it can never decrease and is only weakly sensitive to the number of citations received.

  Also the h_T index can never decrease, but it is sensitive to each publication and citation.

Concluding we may say that the h_T index solves many of the problems related to the *h*-index. If applied to journals or other situation where there is no precision problem, and when using a dedicated software program then, in our opinion, the h_T index becomes an almost perfect *h*-type indicator.

Acknowledgements

The authors thank Du Jian, Sun Xiu-kun, Huo Dao-fu, Chai Jing-chun, Zhang Yi-chuan for data support. They thank Ronald Rousseau for helpful suggestions and for making this translation.

References

Alonso S., Cabrerizo F.J., Herrera-Viedma E. & Herrera F. (2010). *hg*-index: A new index to characterize the scientific output of researchers based on the *h*- and *g*- indices. *Scientometrics* , 82(2), 391-400.

Anderson T. R., Hankin R. K. S., Killworth P. D. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics,* 76: 577-588.

Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, 436, pp. 900.

Egghe L. (2006). Theory and practise of the *g*-index. *Scientometrics*, 69:131–152.

Egghe, L. (2010a). The Hirsch-index and related impact measures. *Annual Review of Information Science and Technology (ARIST)*, 44, 65-114.

Egghe, L. (2010b). Conjugate partitions in informetrics: Lorenz curves, h-type indices, Ferrers graphs and Durfee squares in a discrete and continuous setting. *Journal of Informetrics* (to appear)

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

Jin, BH. (2006). H-index: An evaluation indicator proposed by scientist. *Science Focus*, 1(1), 8-9 (in Chinese).

Jin, BH., Liang, LM., Rousseau, R. Egghe, L. (2007). The *R*- and *AR*-indices: Complementing the *h*-index. *Chinese Science Bulletin*, 52:855-863.

Liang, LM and Rousseau, R. (2009). A general approach to citation analysis and an h-index based on the standard impact factor framework. *Proceedings of ISSI 2009* (Larsen & Leta, eds.), 143-153.

Liu, YX and Rousseau, R. (2009). Properties of h-type indices: the case of library classification categories. *Scientometrics*, 79, 235-248.

Panaretos, J. & Malesios, C.C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics,* 81(3), 635-670.

Rousseau, R., Guns, R. & Liu YX. (2008). The *h*-index of a conglomerate. *Cybermetrics*, 12(1), Paper 2.
http://www.cindoc.csic.es/cybermetrics/papers/v12i1p2.html

Vinkler P. (2007), Eminence of scientists in the light of the *h*-index and other scientometric indicators. *Journal of Information Science*, 33, 481–491.

Wan, Jk., Hua, PH., Rousseau, R, and Sun, XK. (2010). The download immediacy index (DII): Experiences using a Chinese full-text database. *Scientometrics* (to appear). DOI: 10.1007/s11192-010-0171-2

Wan, JK., Xue, FY.(2008). *Chinese Academic Journals Comprehensive Citation Report*. Beijing: Science Press.