Publication performance without calculations: an analysis based on the *p*-index

Nadia Simoes and Nuno Crespo*
Instituto Universitário de Lisboa (ISCTE-IUL), ISCTE Business School Economics Department, BRU-IUL (Business Research Unit), Lisboa, Portugal

Bibliometric analysis is increasingly being used for key decisions in science and higher education. Regarding author-level performance evaluation, the *p*-index is a well-established measure. In this study, we propose the *p′*-index, which introduces a small adjustment in the *p*-index. Based on this transformation, we are able to build an index that is almost perfectly correlated with the *p*-index, but does not imply any form of calculations. The score of each researcher is presented in a table as on-line appendix. Only the total number of papers and citations are needed to obtain the final performance score. We discuss the application of this measure with a sample of economists belonging to the top 10 world universities.

Keywords: Citations, economists, *h*-index, *p*-index, publication performance.

BIBLIOMETRIC measures are increasingly used to support a wide range of critical decisions in science and higher education, including hiring, promotions, grants, awards and rankings[1–4]. As mentioned by Hicks *et al.*[5], 'research evaluations that were once bespoke and performed by peers are now routine and reliant on metrics'. The quality of the measures is therefore a critical issue that must be assured. However, no less important than the quality of the measures is their simplicity in terms of calculation and interpretation. This characteristic is essential to promote transparency in research evaluation and is the main reason for the remarkable impact of the *h*-index introduced by Hirsch[6].

Scientific performance is a key dimension of research evaluation. The *p*-index (or mock *h*-index) introduced by Prathap[7,8] aims precisely at the measurement of this concept. As mentioned by Prathap[7], 'it appears to be the ideal performance indicator'.

Taking this context into account, the present study contributes to the literature by proposing a measure of publication performance. This measure introduces a small adjustment in the *p*-index and does not require any form of calculation. We build a table (presented here as on-line appendix) showing the final performance scores for each combination of papers and citations. The evidence shows that this new metric is almost perfectly correlated with the *p*-index (0.999), indicating that we can reach the same conclusions without having to perform any calculations. This can be useful not only for evaluation purposes, but also for the researchers since it provides an immediate idea of their level of performance. While we explore the case of author-level evaluation throughout this study, the method can be easily adjusted to other units of analysis.

Following the traditional bibliometric approach, we assume that papers and citations are the relevant dimensions to take into account. Aiming to measure the level of scientific performance of a given author, Prathap[7,8] introduces the *p*-index in which the '*p*' expresses 'p'erformance. For author *i*, this index is obtained as follows

$$p_i = \left(C_i \cdot \frac{c_i}{p_i} \right)^{\frac{1}{3}}. \quad (1)$$

While the total number of citations (C) reflects activity, the quality dimension is captured through (C/P), where P represents the total number of papers of the author. Therefore, it is clear that the level of performance increases with C. In turn, for a given C, performance decreases with P.

The main goal of the present study is to introduce a novel index of performance (p'-index) that can be obtained without calculations. To that end, only a small adjustment in the *p*-index is needed. More specifically, following the 'rule' popularized by the *h*-index and several of its variants, we replace C by $[\text{int}\sqrt{C}]$, where int expresses the fact that we round the value down to the nearest integer. While this assumption violates the principle that a bibliometric measure of publication output should be 'strictly monotonic', i.e. assign a positive score to each new citation as it occurs[9], it is commonly used by several metrics, including the *h*-index. This adjustment has two main consequences. First, different from the *p*-index but similar to the *h*-index, some citations may become irrelevant to the final performance score of the author. This occurs when the number of citations (ignoring their distribution among the papers) is higher than the minimum value necessary to reach a score h, but lower than the necessary to reach ($h + 1$). Second, the number of different scores that can be obtained is dramatically reduced, making it possible to present the results in a table.

For author i, the p'-index is obtained as

$$p'_i = \begin{cases} \left[(\text{int}\sqrt{C_i}) \cdot \frac{(\text{int}\sqrt{C_i})}{P_i} \right]^{\frac{1}{3}} & \text{if } P_i > 0 \\ 0 & \text{if } P_i = 0. \end{cases} \quad (2)$$

Knowing the total number of papers and citations of the author, the publication performance score can be

*For correspondence. (e-mail: nuno.crespo@iscte-iul.pt)

Table 1. Number of papers, number of citations and p'-index

Number of citations	Number of papers															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1–3	–	1,000	0,794	0,693	0,630	0,585	0,550	0,523	0,500	0,481	0,464	0,450	0,437	0,425	0,415	0,405
4–8	–	2,520	2,000	1,747	1,587	1,474	1,387	1,317	1,260	1,211	1,170	1,133	1,101	1,072	1,046	1,022
9–15	–	4,327	3,434	3,000	2,726	2,530	2,381	2,262	2,163	2,080	2,008	1,945	1,890	1,840	1,795	1,754
16–24	–	6,350	5,040	4,403	4,000	3,713	3,494	3,319	3,175	3,053	2,947	2,855	2,773	2,700	2,635	2,575
25–35	–	8,550	6,786	5,928	5,386	5,000	4,705	4,470	4,275	4,110	3,969	3,844	3,735	3,636	3,547	3,467
36–48	–	10,903	8,653	7,560	6,868	6,376	6,000	5,699	5,451	5,241	5,061	4,902	4,762	4,637	4,524	4,421
49–63	–	13,391	10,628	9,284	8,435	7,831	7,369	7,000	6,695	6,437	6,215	6,021	5,849	5,695	5,556	5,430
64–80	–	16,000	12,699	11,094	10,079	9,357	8,805	8,364	8,000	7,692	7,427	7,194	6,989	6,805	6,639	6,488
81–99	–	18,721	14,859	12,980	11,793	10,948	10,302	9,786	9,360	9,000	8,689	8,418	8,177	7,962	7,767	7,591
100–120	–	21,544	17,100	14,938	13,572	12,599	11,856	11,262	10,772	10,357	10,000	9,687	9,410	9,163	8,939	8,736
121–143	–	24,464	19,417	16,962	15,411	14,307	13,463	12,789	12,232	11,761	11,355	11,000	10,686	10,404	10,150	9,920
144–168	–	27,473	21,805	19,049	17,307	16,066	15,119	14,362	13,737	13,208	12,752	12,353	12,000	11,684	11,399	11,140
169–195	–	30,567	24,261	21,194	19,256	17,876	16,822	15,979	15,284	14,695	14,188	13,744	13,352	13,000	12,683	12,394
196–224	–	33,742	26,781	23,395	21,256	19,732	18,569	17,639	16,871	16,221	15,662	15,172	14,738	14,350	14,000	13,682
225–255	–	36,993	29,362	25,650	23,304	21,634	20,358	19,338	18,497	17,784	17,171	16,634	16,158	15,733	15,349	15,000
256–288	–	40,317	32,000	27,955	25,398	23,578	22,188	21,076	20,159	19,383	18,714	18,129	17,610	17,147	16,728	16,348
289–323	–	43,712	34,694	30,308	27,537	25,563	24,056	22,851	21,856	21,014	20,289	19,655	19,093	18,590	18,137	17,724
324–360	–	47,173	37,442	32,708	29,717	27,587	25,960	24,660	23,587	22,679	21,896	21,211	20,605	20,062	19,573	19,128
361–399	–	50,700	40,240	35,153	31,939	29,649	27,901	26,504	25,350	24,374	23,533	22,797	22,145	21,562	21,036	20,558

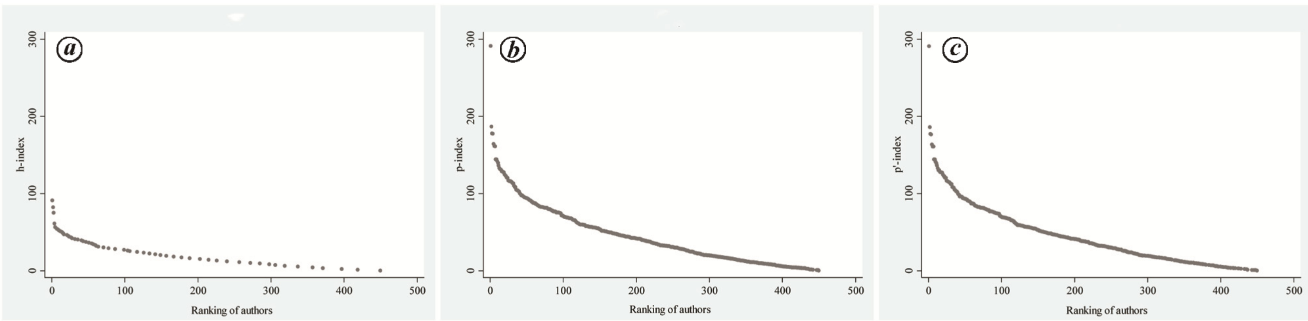

Figure 1. Distribution of the h-index (***a***), p-index (***b***) and p'-index (***c***).

immediately seen in the table presented in the on-line appendix. Table 1 shows a small part of that table, illustrating the basic rationale of the measure.

As already mentioned, the measure introduced in this study (and the p-index) only accounts for the total number of papers and total number of citations. This can be extended in order to also capture the inconsistency (unevenness) of the distribution of citations across the papers of the author, as proposed by Prathap[10].

To illustrate the methodological approach presented above, we use a group of 472 authors. These correspond to the authors with primary appointments in the economics departments of the top-10 world universities, as defined by the QS World University Ranking in the subject of 'economics and econometrics' (Harvard University; Massachusetts Institute of Technology; Stanford University; University of California, Berkeley; Princeton University; University of Chicago; London School of Economics and Political Science; University of Oxford; Yale University and Columbia University). Our sample includes tenure-track or tenured faculty members at the end of 2018, as described in the departmental websites. All data were gathered from the Web of Science (WoS) database (core collection) considering two main criteria: (i) type of paper: articles and reviews, and (ii) language: English. In order to avoid the well-known problems associated with the building of bibliometric databases (for a discussion, see Schreiber[4]), we conduct a comparison, for each author, between the list of papers given by WoS and the information retrieved from departmental and personal websites. Having finalized this process, our sample includes 15,243 papers (published between 1957 and 2018 in 709 different journals) and 1,288,803 citations.

Figure 1 *a–c*, presents the distribution of h, p and p'. We calculate the correlation between the ranking of the authors in the sample using both measures of performance (p and

p') and the h-index. Using the Spearman correlation coefficient, a value of 0.999 is obtained between p and p', indicating an almost perfect correlation between these two measures. Thus, they produce the same general conclusions. In turn, the correlation coefficient between p' and h is 0.942. Despite this high correlation between p' and h, it is important to highlight that the level of granularity of p' (and p) is much higher than in the case of the h-index.

Bibliometric measures are critical inputs for many decisions in universities and research units, including hiring, promotions, research funds allocation and rankings. In this context, the evaluation of performance is critical. In this study, we propose a small adjustment in the p-index advanced by Prathap[7,8]. The main advantage of our measure is that it can be obtained directly from the table presented here, without any calculations. This is an important merit since it improves access and transparency in the bibliometric assessment of publication performance.

1. Perry, M. and Reny, P., How to count citations if you must. *Am. Econ. Rev.*, 2016, **106**, 2722–2741.
2. Barnes, C., The *h*-index debate: an introduction for librarians. *J. Acad. Libr.*, 2017, **43**, 487–494.
3. Hamermesh, D., Citations in economics: measurement, uses, and impacts. *J. Econ. Lit.*, 2018, **56**, 115–156.
4. Schreiber, M., A skeptical view on the Hirsch index and its predictive power. *Phys. Scripta.*, 2018, **93**, 102501.
5. Hicks, D., Wouters, P., Waltman, L., De Rijcke, S. and Rafols, I., Bibliometrics: the Leiden manifesto for research metrics. *Nature News*, 2015, **520**, 429.
6. Hirsch, J., An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 16569–16572.
7. Prathap, G., Is there a place for a mock *h*-index? *Scientometrics*, 2010, **84**, 153–165.
8. Prathap, G., The 100 most prolific economists using the *p*-index. *Scientometrics*, 2010, **84**, 167–172.
9. Anderson, T., Hankin, R. and Killworth, P., Beyond the Durfee square: enhancing the *h*-index to score total publication output. *Scientometrics*, 2008, **76**, 577–588.
10. Prathap, G., The zynergy-index and the formula for the *h*-index. *J. Assoc. Inf. Sci. Technol.*, 2014, **65**, 426–427.

ACKNOWLEDGEMENTS. This work was supported by the Fundação para a Ciência e a Tecnologia under Grant UID/GES/00315/2019. We thank Gangan Prathap for useful discussions.

Received 17 August 2019; accepted 21 August 2019

doi: 10.18520/cs/v117/i11/1872-1874

Comparison of parametric and non-parametric methods for chlorophyll estimation based on high-resolution UAV imagery

Gaurav Singhal[1,2], Babankumar Bansod[1,2,*], Lini Mathew[3], Jonali Goswami[4], B. U. Choudhury[5] and P. L. N. Raju[4]

[1]CSIR-Central Scientific Instruments Organization, Chandigarh 160 030, India
[2]Academy of Scientific and Innovative Research, Ghaziabad 201 002, India
[3]National Institute of Technical Teachers, Training and Research, Chandigarh 160 019, India
[4]North East Space Application Centre, Barapani 793 103, India
[5]ICAR Research Complex for NEH Region, Umiam 793 103, India

The present study provides a systematic comparison of parametric and non-parametric retrieval methods using high-resolution data provided by the unmanned aerial vehicle (UAV). We used turmeric crop reflectance data to evaluate the vegetation index (VI)-based parametric methods and compared them with linear and nonlinear non-parametric methods to build a rigorous LCC estimation model. The study demonstrates that the best-performing VI was the normalized green red difference index (GNRDI), with $R^2 = 0.68$, RMSE = 0.13 and high processing speed of 0.08 s. With regard to non-parametric methods, almost all methods outperformed their parametric counterparts. Particularly, methods such as random forest (RF) and kernel ridge regression (KRR) showed the best performance characterized by $R^2 > 0.72$ and RMSE ≤ 0.12 mg/g of fresh leaf weight. These non-parametric methods possessed the benefit of total spectral information utilization and enabled robust, non-linear relationship between the predictor and target variables, but computational complexity is a major drawback.

Keywords: Chlorophyll, machine learning, unmanned aerial vehicle, vegetation index.

BEING the most important pigment of all photosynthetic cells, variations in leaf chlorophyll concentration (LCC) are an indicator of crop growth and stress status and help in estimating biomass and yield[1–3]. Chlorophyll is best described by its absorption in the red band (600–720 nm) and major reflection from green and NIR wavebands[4].

Remote sensing sensors play an important role in crop health monitoring in terms of large coverage area and fast estimates of crop biophysical and biochemical parameters such as LCC. Remote sensing satellites are capable of providing a vast coverage, but their potential is limited by poor resolution and cloud cover. Flying at low altitude

*For correspondence. (e-mail: csiobabankumar@gmail.com)