The *h* index

Budiman Minasny, Alex. McBratney & Alfred Hartemink

Introduction

Recently Hirsch (2005) proposed the *h* (or Hirsch) index as an assessment of the research performance of individual scientists. Hirsch defined it as:

> "a scientist has index *h* if *h* of his/her N_p papers have at least *h* citations each, and the other (N_p – *h*) papers have no more than *h* citations each."

Although the paper by Hirsch was published as recently as 2005, it is now being used in several disciplines for ranking or assessing scientists' performance. It is also used as one of the criteria for staff promotion in some departments. The *h* index is a single number that represents both productivity (number of papers) and their impact (number of citations). Here we look at the distribution of the *h* index for pedometrics and pedometricians.

The algorithm to calculate *h* index is as follows: all papers are ranked based on their number of citations, from the most, to the least, cited (See Fig. 1). The paper rank which equals the number of citations is the *h* index. Graphically, as in Fig. 1, the intersection of the 45 degree line with the curve of the number of citations versus the rank gives the *h* index. The total number of citations is the area under the curve.

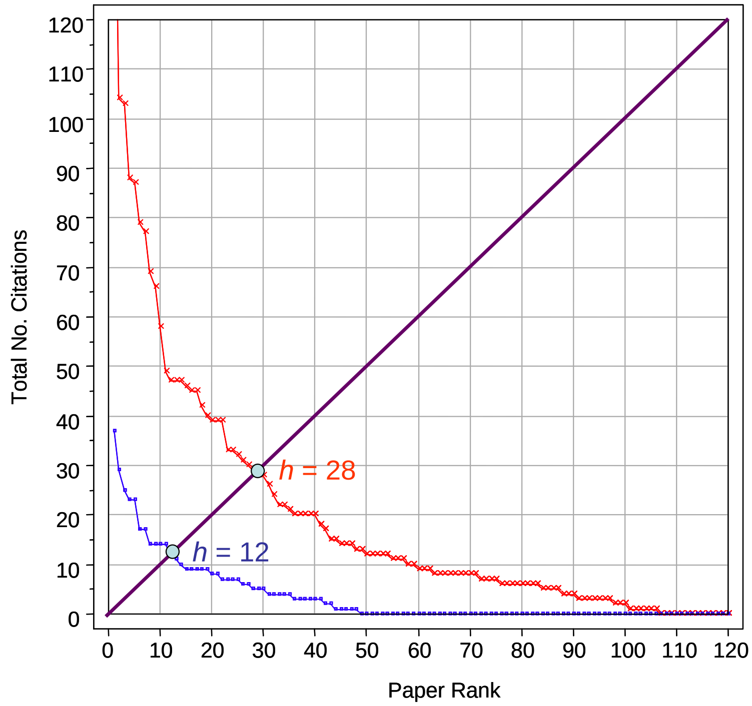

Figure 1. Total no. citations vs. paper rank, the intersect with a 1:1 line is the *h* index.

The distribution of the number of citations (Fig. 1) usually can be modelled as a stretched exponential function (Hirsch, 2005; Laherrere and Sornette, 1998):

$$N_c(y) = N_0 \exp\left[-\left(\frac{y}{y_0}\right)^{\beta}\right] \quad (1)$$

where $N_c(y)$ is the number of citations of the *y*-th paper (ordered from the most to least cited), N_0 is maximum no. citations, and β, and y_0 are empirical parameters with $\beta \le 1$. When plotted on a semi-log plot it gives a straight line, and in a log-log plot shows curvature. The larger the value of β and y_0, the larger *h* will be. The total no. of citations for N_p papers can be found by:

$$N_{c,tot} = \int_1^{N_p} N_c(y)\,dy \quad (2)$$

In practice, it is much easier to count the number of citations than doing the integral.

Data synthesis

We selected 35 pedometricians randomly using the ISI web of science database, accessed in November 2006. The following parameters were recorded: the number of papers N_p, total number of citations $N_{c,tot}$, year of the first paper published by the scientist, the average citations per paper ($N_{c,tot} / N_p$), and the *h* index.

The *h* index for the 35 pedometricians ranges form 1 to 32, with a median of 7. The scientific age (no. years since the first paper is published) ranges from 1 to 41 years representing early to mature researchers.

Results and Discussion

Figure 2 shows the relationship between *h* and total no. of citations. *h* increases with the square root of the number citations, following the diffusion or sorptivity process:

$$h = 0.57\sqrt{N_{c,tot}}\ . \qquad (3)$$

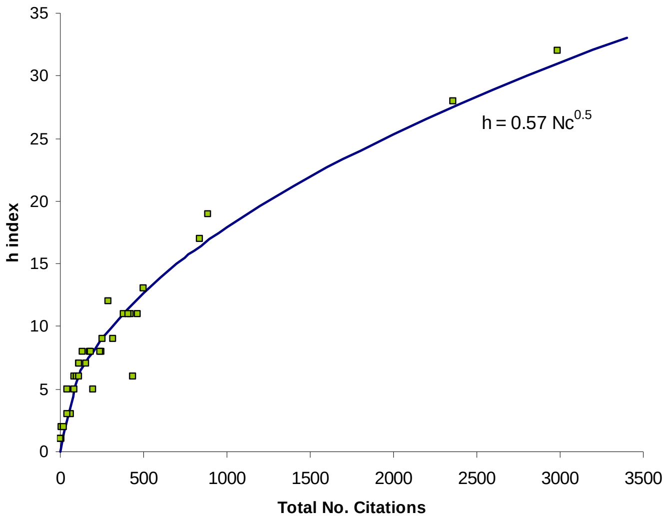

Figure 2. Relationship between total no. citations and *h* index for pedometricians.

We can see that *h* is closely related to the total number of citations. Thus *h* depends on how many citations one can earn in the pedometrics subject. This in part depends on the number of pedometricians.

Another factor that controls *h* is obviously the age of the researcher. Typically *h* increases linearly with time, assuming that a researcher has a constant output of papers and the papers are cited. A linear relation with age is proposed by Hirsch (2005):

$$h = m\, t \qquad (4)$$

where *t* is the "scientific age" of the researcher, and *m* is the impact or productivity of the researcher. As a measure of *t*, it can be approximated by the number of years after the first published paper until the present. The year when the first paper is published usually occurs at the end of the PhD degree (approximately 25-30 years or age). For physics, Hirsch found that $m = 1$ characterised a successful scientist (meaning that after ten years the top 10 paper will be cited more than 10 times), and $m = 2$ is outstanding. The relationship assumes that the researcher has a constant output of *p* papers per year and each paper gets cited *c* times per year.

This "standard" in physics may not be applicable in pedometrics. Fig. 3 shows the relationship between *h* and scientific age *t* for the 35 pedometricians. We found the "average" productivity and impact curve for pedometrics:

$$h = 0.7\, t \qquad (5)$$

This means that on average a pedometrician should get an annual increase of 0.7 in the *h* index. It will take one and a half years for an increase of one *h* unit, assuming a constant output of papers and annual citations.

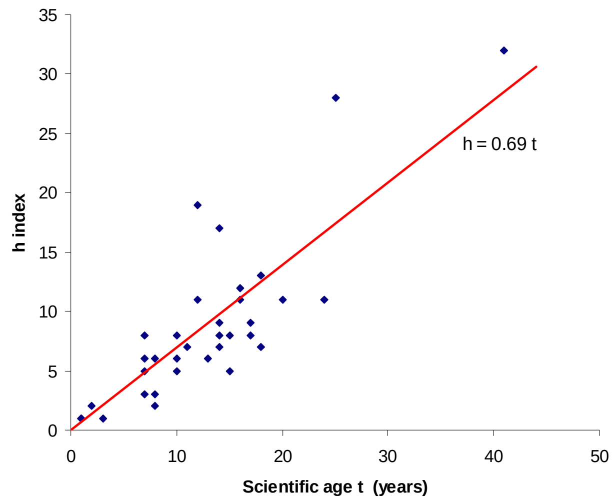

Figure 3. Relationship between scientific age and *h* index. The line represents the average impact curve for pedometrics.

We performed a regression between "scientific age" and the number of papers and average citations per paper. From Figs. 4 & 5 we can deduce that pedometricians publish on average 2 papers per year and each paper is citied 0.75 times per year.

$$N_p = 2.3\, t \qquad (6)$$

$$N_c/N_p = 0.75\, t \qquad (7)$$

Although some authors deemed the average no. of citations is better than other indices (Lehmann et al., 2005, 2006), it could be a bit deceptive. Most of the time, the top papers will be cited more frequently and the rest may not be cited at all. Most people have a highly skewed citation pattern.

Clearly, the average *h* index in pedometrics and soil science is lower than major science disciplines like physics or chemistry. The highest *h* index in physics about 110, and in biology is an unimaginable 190 (See: http://en.wikipedia.org/wiki/Hirsch_number). The *h* index is strongly related to the square root of the total number of citations. Pedometrics is a young and expanding area of research, but the number of researchers interested is still small. Compared with a larger dataset that we analysed (about 200 soil scientists) we found that the average relationship in pedometrics (Eq. 5) is the same as the average in soil science.

Based on Eq. (5) you can set your benchmark of the *h* index. You can use this as your argument for promotion, that is if your *h* index is similar to or larger than the average.

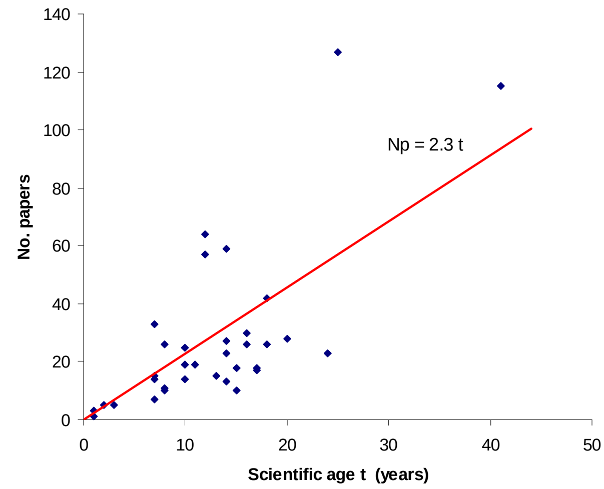

Figure 4. Number of papers as a function of time. Pedometricians on average publish 2 papers a year.

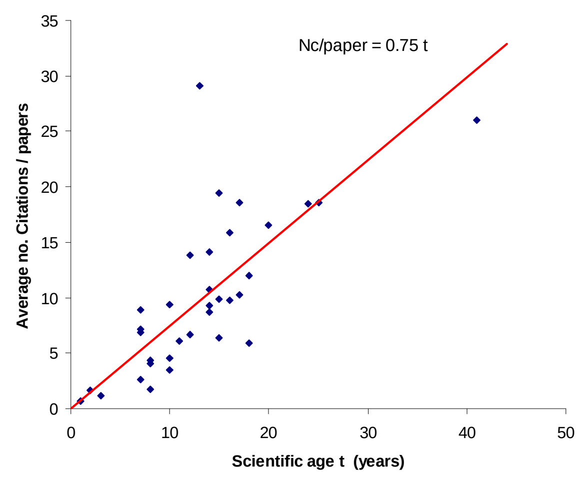

Figure 5. Average no. citations per paper as a function of time. Pedometrics paper on average get cited 0.75 time per year.

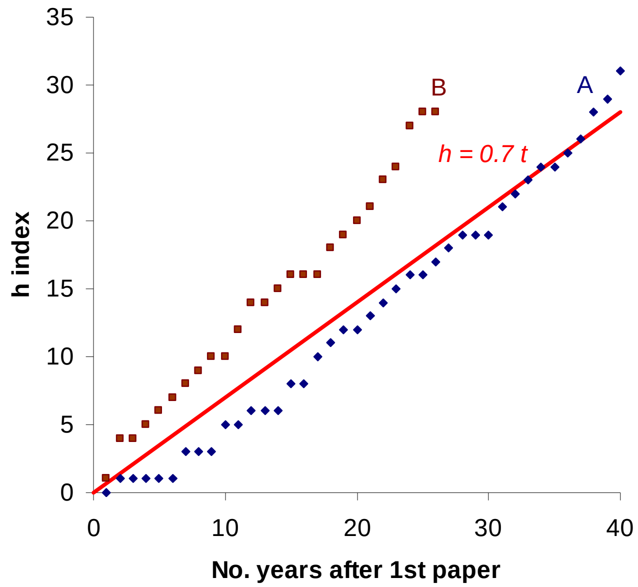

Figure 6. *h* index as a function of time for two pedometricians.

To see the trend of *h* index with time, we calculated the *h* index for two pedometricians: A and B (Fig. 6). A is a senior pedometrician and his first paper was published in the mid 1960's. The *h* index for A appears to be below the average line, this is because the *h* index is not linear with time, rather there is an initial lag of take-up of the subject (about 10 years). The *h* index seems to be increasing more recently. Pedometrics is still new at that time and requires some time for the topic to be accepted. Meanwhile the *h* index for B is increasing linearly with time, and above the average line. B started publishing in early 1980s and it appears now that pedometrics are well received.

Realistically, there is no single index that can capture everything, echoes of Philip (1974, p.268). We think that combinations of no. papers, average no. citations, and *h* index can give a good indication of your performance. Equations (5), (6), and (7) should give you a standard to compare.

If you don't have access to ISI, you can use Google scholar as a database. The webpage from University of Århus Denmark calculates the *h* and *m* indices from Google scholar:
http://www.brics.dk/~mis/hnumber.html
The software from Harzing, "Publish or Perish" also does the same thing:
http://www.harzing.com/resources.htm.

References

Ball, P., 2005. Index aims for fair ranking of scientists. Nature 436, 900.

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, 16569–16572. DOI: 10.1073/pnas.0507655102

Laherrere, J., Sornette, D., 1998. Stretched exponential distributions in Nature and Economy: "Fat tails" with characteristic scales, European Physical Journal B 2, 525-539. (http://xxx.lanl.gov/abs/cond-mat/9801293)

Lehmann, S., Jackson, A.D., Lautrup, B.E., 2005. Measures and Mismeasures of Scientific Quality http://arxiv.org/PS_cache/physics/pdf/0512/0512238.pdf

Lehmann, S., Jackson, A.D., Lautrup, B.E., 2006. Measures for measures. Nature 444, 1003-1004 http://www.nature.com/nature/journal/v444/n7122/full/4441003a.html

Philip, J.R., 1974. Fifty years of progress in soil physics. Geoderma 12, 265-280.