Impact Factor, Citation Index, H-Index: are researchers still free to choose where and how to publish their results?

Stefano Solarino

Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Sede di Genova, Italy

ABSTRACT

Over the last decade, the demand to evaluate the impact of any given research study, the credentials of a researcher, and the influence that any single research unit or agency has on the world of research has constantly grown. Many tools have been developed and applied to evaluate the level of innovation, originality and continuity of a single researcher in an objective way. As a consequence, there are comparisons of the performances of different research agencies. Some of these tools, which often provide the result as an 'index', are briefly described in this study. However, it is clearly evident that the evaluations provided by these instruments do not always correspond to the real impact of the research, nor are they unique. Indeed, the same index computed using similar criteria on different databases gives different scores, which can lead to confusion and contradictions. In this contribution, the principal anomalies, problems and failures of these evaluation schemes are described. The most evident of these arise from the nature of the evaluation, which being automated, cannot establish the role of any single researcher in papers of 'pooled' research, and cannot recognize similar or duplicated papers by the same researcher(s) in more than one journal. The 'selecting' effects that these evaluation indices can have on the research are then discussed. Indeed, in an attempt to obtain the highest possible scores in terms of citations, there is a tendency of the single scholar to avoid studies that deal with small areas, or with scientific problems that do not have a broad interest or provide applicative results. In all of these cases, an article describing such studies will in all likelihood appear in a 'minor journal' (one with a low impact factor). As a consequence, this will provide a low citation index, will not significantly contribute to the authors' H-index, and/or will only be published as a report. Moreover, a discussion on the role that these evaluation indices can have in the world of research is presented. Particular attention is paid to the consequences in the field of the geoethics, where scientific, technological, methodological and socio-cultural aspects need to be considered in a different order to that expected in a pure meritocracy.

1. Introduction

Over the last decade, it has become necessary to objectively evaluate the impact of any research study, and of the 'credentials' of a researcher or the influence that any single research unit or agency has on the world of research. To be effective, this evaluation must be rapid, easy to update, based on sound elements, widely available, easy to use, and of course, as objective as possible. Once a methodology to evaluate a research study is designed, the applications of such an assay can be very large. Indeed, they span the mere impact of the single academic, to the allocation of funds, career advancement, job/position assignment, and awards, even if the very different natures of these applications require additional factors to be taken into account. For example, the amount and quality of the data from which a scientific result is obtained often depend on the availability of instruments, computers and other facilities, rather than simply on the skills and experience of the research unit or the single researcher.

Given this situation, I discuss here the problems that can arise and are often not taken into account when using these tools, and especially the consequences that the using these indices can have on the research. As known, most of these evaluations are compacted into a number or an index. I will focus on the research field that I know more about, which is geophysics, but I am almost convinced that similar problems affect most research fields, if not all.

The debate that I will summarize applies with different relevance across various countries. To cite just a few of the 'filters', English-speaking countries have a natural tendency to write papers in the English language (which is one of the requirements for many of the evaluation indices). However, some countries host the most important press companies in any particular science, which are therefore a favourite target for the scientists living in that country. Some other countries, and especially those in the Far East, have historic roots and origins in science, due to which they prefer to use their own language, and indeed, were not allowed to use other languages in the past, or have advised researchers against publishing in foreign journals.

Italy is slowly approaching the need for an objective way to evaluate its research and scholars. There are a few elements that render this approach more clumsy: the cultural importance of the Italian language; the existence of many locally distributed journals, some of which are very slowly becoming international by opening the Editorial Board to foreign scientists and by hosting articles in English; and the pioneering nature of geophysics in Italy, which was the reason for establishing the XX century relationships, which included bilingual journals, with advanced countries like Germany, more than with English-speaking countries. Just to mention an example, one of the most important journals in geophysics, the *Rivista di Geofisica Pura ed Applicata* [Bossolasco 1939], turned to the current structure and name of Pure and Applied Geophysics in the late 1960's, although it was still publishing papers in German, Italian and English at that time.

The first part of this report sketches out the main evaluation tools that are used to score scholars, and describes some of their incongruities. The aim of the discussion propounded here is not to criticize any single index or to label it as good or bad. I do not have enough skill and experience to even dare to attempt this. Furthermore, I am firmly convinced that even with the failures that I will briefly describe here, the indices in current use are sensitive and objective enough to grossly describe the credentials of scientists in Earth sciences. However, a discussion of the roles that these evaluation indices have in the world of research is well supported. This is especially relevant in the light of the geoethics, where scientific, technological, methodological and socio-cultural aspects have to be considered in an order that is different to what is expected in a pure meritocracy. This aspect will be the argument of the second part of this article, together with a few suggestions on how to render the use of evaluation indices if not more effective, at least more adherent to the real importance of the research.

2. Citation Metrics

Over the last few years, many tools to quantify the scientific importance of any research have been proposed, developed and adopted. Although some of these are relatively new (e.g. the H-index dates back to 2005 only), they have been rapidly included in the computation schemes and are currently provided by several indexing agencies. These evaluations score the publications or compute the score of the single scholar on the basis of his/her publications. A detailed description of the evaluation criteria or the single number/ index goes far beyond the scope of this article; however, a short summary of the main characteristics might help the reader to get to the point, and so these are described in the next lines.

In particular, the discussion here is based on the Impact Factor (IF), the Citation Index, and the H-index. Generally speaking, these numbers are often grouped under the name of 'citation metrics', or bibliometrics.

The IF is a measure that reflects the average number of citations to articles published in science and social science journals. In a given year, the IF of a journal is the average number of citations received per paper published in that journal during the two preceding years. For example, if a journal has an impact factor of 5 in 2010, then the papers it published in 2008 and 2009 received 5 citations each on average in 2010. The impact factor of a journal is calculated as follows:

A = the number of times articles published in 2008 and 2009 were cited by indexed journals during 2010.

B = the total number of 'citable items' published by that journal in 2008 and 2009. These citable items are usually articles, reviews, proceedings, or notes, but do not include editorials or letters to the Editor.

The IF for the year 2010 is then **A/B**, and it is computed only in 2011 because it cannot be calculated until all of the 2010 publications have been processed by the indexing agency.

This IF was devised by Eugene Garfield [1998], and it is provided to subscribers by the agency founded by Garfield himself, the Institute for Scientific Information (ISI; http://apps.webofknowledge.com/), which is now part of Thomson Reuters. Although the IF by the ISI accounts for more than 11,000 science and social sciences publications, these represent only a part of all of the present-day journals.

There are alternative ways to rank the influence of any journals and to compute the IF. These include:

– the PageRank [Brin and Page 1998] algorithm of the Google search engine;

– the Eingenfactor (http://www.eigenfactor.org/), in which journals are rated according to the number of incoming citations, with citations from highly ranked journals weighted to provide a greater contribution to the eigenfactor than those from poorly ranked journals;

– the SCImago Journal Rank (http://www.scimago jr.com/), an open-access journal metric that is based on Scopus data and that uses an algorithm similar to PageRank.

The IF is a quick tool to evaluate the breadth of distribution of a journal, but it is not problem free. Its value is not merely dependent on the citations (A, in the formula). Indeed, if a journal publishes only a few, longer articles, or does not publish any item in a given year, this will increase the IF the following year (B will be small); conversely if the number of articles per year is high, the IF will be small if the number of citations is not also significantly higher.

Moreover, when journals are included in the ISI list, it takes three years before they are assigned an IF, which means that issues published before or during the first two

years of computation are not assigned an IF. Finally, the IF can change abruptly from year to year, but the authors know the IF relative to their publication only at the end of the successive year, making it very hard to forecast the popularity of a journal, and in turn, the benefit of their choice to publish in that journal.

On the other hand, it is very clear that the IF is a qualitative more than a quantitative tool. In practice, any article published in journals for which an IF is computed concur to the credential of a scholar as a single unit, no matter how high the rank of the journal is. A scientist is attributed the number of published articles, and not a sum weighted on the IF of the journal where they appear. An exception is with the very high IF journals, which are sometimes mentioned separately in a scientist's resume, because, being top publications, they carry concrete credibility to the authors. Nevertheless, there is indeed a sort of IF 'filter' on the research. Indeed, in principle, the higher the IF of a journal is, the better known, distributed and read it is. For this reason, the top-ranked journals will not be likely to host studies related to minor topics or small geographical areas, and the Editors will strive to publish arguments regarding national to planetary matters, and novel arguments. I will recall this point in the discussion on the roles of these evaluation tools.

A second, common way to evaluate research is the Citation Index, which sums up the number of citations of all articles for a given author. There are several indexing agencies and bibliographic databases that formulate this evaluation.

The oldest of these citation indices, the Science Citation Index, was established by Garfield in 1960 (http://thomsonreuters.com/products_services/science/free/essays/history_of_citation_indexing/), and it is currently accessible to subscribers under the ISI Web of Science webpage or on a CD. Then there is the bibliographic database SCOPUS (http://www.scopus.com/ home.url), which is owned and run by Elsevier and offers a similar service to subscribers. Finally, the Google Scholar database (http:// scholar.google.com/) can be queried via a program (Publish or Perish [PoP]) [Harzing 2010] that is becoming very common among scientists, because it is very easy to use (it self connects to the web and searches for the most updated database), and it is free of charge and open to any user.

Unfortunately, these three bibliographic databases are very different, and the choice of a reference dataset turns into divergent results with respect to using one over the others. Table 1 shows a comparison of a query to evaluate my own research activities using these three databases. The results are very different, with variations of the order of 50% in these citation indices and the H-index, the description of which is below.

The most evident problems with the citation indices are:

– they do not take into account the number of authors on each paper, so that if an article is written by more scientists, each will get an identical score, no matter how big their contribution;

– they do not take into account the IF of the journal, meaning that an article published in a very low ranked journal will be rated as being as important as a very highly ranked paper;

– they are not 'time weighted' or 'dependent'. The citation indices give the sum of the citations over the career period (or at least over the last 20 years) without any differences between long or short activity periods. Very prolific young scholars will have similar performances as older, less productive scientists, and vice versa.

Finally, the H-index [Hirsch 2005] is based on the set of a scientist's most cited papers and the number of citations that they have received in other publications. The definition of the H index is given as: "A scientist has an index H if h of his/her Np papers have at least h citations each, and the other Np-h papers have no more than h citations each". In other words, a scholar with an index of H has published h papers, each of which has been cited in other papers at least h times. Thus, the H-index reflects both the number of publications and the number of citations per publication. The H-index is designed to improve upon simpler measures, such as the total number of citations or publications. The H-index works correctly only for comparing scientists working in the same field, as citation conventions differ widely among different fields.

The H-index can be manually determined using citation databases or can be determined using automatic tools. Subscription-based databases such as Scopus and the Web of Knowledge provide automated calculators. Harzing's PoP program (http://www.harzing.com/ pop.htm) calculates the H-index based on the Google Scholar entries. Each database will produce a different H for the same scholar because of the different cover of the scholar's publications (see Table 1): Google Scholar has more citations than Scopus and the Web of Science, although the smaller citation collections tend to be more accurate

The H-index should cover up the main disadvantages of these other bibliometric indicators, such as the total number of papers or the total number of citations. The total number of papers does not account for the quality of the scientific publications, while the total number of citations can be disproportionately affected by participation in a single publication of major influence (for instance, methodological papers proposing successful new techniques, methods or approximations, which can generate a large number of citations), or having many publications with a few citations each. The H-index is intended to simultane-

Source	Number of publications	Number of citations	H index
Web of Science	19	199	6
Cited reference search	29	235	
Scopus	20	212	7
Google Scholar (via PoP)	29	247	8

Table 1. Computation of the author's personal score as an example of the use of the different databases. Comparison of the number of published papers, the citation index and the H-index according to ISI (two computations), SCOPUS and Google Scholar. The scores vary a lot across the different sources, and show variations of up to 50%. The choice of a specific database biases the (absolute) rate of the single scholar, although it probably affects comparisons between scientists only if their scientific production is very dissimilar (many books or conference papers *versus* journal articles).

ously measure the quality and quantity of scientific output.

There are a number of situations in which the H-index can provide misleading information about a scientist's output. Although most of these are not exclusive to the H-index, they include:

– The H-index does not account for the number of authors on a paper. In the original paper, Hirsch suggested partitioning citations among co-authors. Due to this, the H-index and similar indices tend to favor fields with larger groups, e.g. experimental over theoretical.

– The H-index does not account for the typical number of citations in different fields. Different fields, or journals, traditionally use different numbers of citations.

– The H-index discards the information contained in the author placement in the author list, which in some scientific fields is of significant importance.

– The H-index is bound by the total number of publications. This means that scientists with a short career are at an inherent disadvantage, regardless of the importance of their discoveries. This is also a problem for any measure that relies on the number of publications.

– The H-index does not consider the context of the citations. For example, citations in a paper are often made simply to 'flesh out' an Introduction, and can otherwise have no other significance to the article. Also, the H-index does not resolve other contextual instances, such as citations made in a negative context, and citations made to fraudulent or retracted work. This is also a problem for regular citation counting.

It is of note that various proposals to modify the H-index to emphasize different features have been made, and some of these modifications are currently computed, including the G, Hc and Hm indices.

3. Discussion and conclusions

As already stated, in principle it is reasonable to evaluate the impact of a scholar provided that the rating tools apply to the personal merits of that scientist. However,

from what I have described above, it is clear that in most cases, a score can be influenced by the environment where the scientist works, including the 'pool' he / she belongs to, the availability of money and laboratories, the facilities he / she can profit from, and the importance of the agency where he / she works. This 'starting' condition reflects on all of the rest of a career, because the rating of a scholar is often used as a discriminating element for funding, job advancement, and personnel selection. This thus provides academic impact and makes the scholar suitable for more credits in the future, with respect to other colleagues.

In addition, within any similar 'environmental' situations, a scholar can perform better than others if he / she makes particular choices. In the introduction to PoP, it says: "If an academic shows good citation metrics, it is very likely that he or she has made a significant impact on the field. However, the reverse is not necessarily true. If an academic shows weak citation metrics, this may be caused by a lack of impact on the field, but also by one or more of the following:

– working in a small field (therefore generating fewer citations in total);

– publishing in a language other than English (effectively also restricting the citation field);

– publishing mainly (in) books."

It is then clear that the topics, the conditions, and the means of research of a scholar looking for impact are not completely free. Indeed, to get a higher score, the results should not be published in a language other than English, should not be printed on books (this last point does not apply to some indexing agencies, like ISI, which recently included citations from books in the computation), and should not be devoted to 'small fields' of research. However, I would also add that they should not be published as a 'User's manual', nor as a United States Geological Survey Open Report File, just to mention two examples. Indeed, in the latter case, as these publications are not subjected to peer review, they will not be included in the computation of citations by most of the indexing agencies, although it is well known that anybody using a program code will certainly use and then cite the accompanying manual or report describing its features and use.

And what about publication costs? Sometimes to publish on a high-impact journal the authors have to cover printing expenses. The amount of money available in research projects depends on many factors, but seldom it is proportional to the target of the research. As an example, theoretical studies are often not adequately funded, as they do not need field or laboratory activities, although these studies can produce results suitable for a widely distributed journal that can only be accessed if the costs for publication can be covered.

There are significant implications from all of the ar-

guments discussed here. I believe that the need to obtain the highest possible score can make a scholar avoid studies that deal with small areas, to give low priority to scientific studies of narrower interest, and to refrain from research activities in applicative fields. In all of these cases, converse to what is required to increase the impact, an article describing the study will probably appear in a 'minor journal' (with low impact factor). It will thus have a low citation index, will not significantly contribute to the H-index, and/or will only be published in reports.

The search for fame and glory of a scholar and the decisions and choices linked to this can reflect on the society and consequences in the field of geoethics, where scientific, technological, methodological and socio-cultural aspects have to be considered in an order different to what would be expected in a pure meritocracy. The most important scientific discoveries, including those in the Earth sciences, have been achieved for ethical reasons, such as to help people, to enlarge perspectives, to investigate causes, and to be prepared to react. Sometimes these studies have been rewarded later, although the converse has never happened. No scientist has up to now estimated his/her work in terms of 'popularity', but only in terms of benefits to the scientific community, with the related consequences for society. Avoiding 'small' arguments or less-widely distributed journals can, in the long term, interfere with the freedom of the single researcher. Indeed, even now the 'recommendation' to publish in highly cited journals is already a limitation to our freedom.

In my opinion, a good compromise between the current rating scheme and a more ethically correct one would be to assign a fraction of the score to the scholar that is proportional to his/her contribution to a study. This can be carried out by simply taking into account the number of authors or by establishing thresholds, like a full score for a single author, a half score for up to three authors, and so on. This feature of dividing the merit according to the number of authors is already included in the computation of some indices, although it has not been steadily accepted. By undertaking this improved assignment, we would avoid the paradox of those very basic studies that sometimes appear in widely distributed journals and are authored by the many scholars , each of which gets a score, participating in a field experiment. Although the article is often dealing with the simple description of the field activity and then carries no significant or novel results, it is cited by the hundreds of colleagues using the data collected in the experiment (increasing as a consequence the citation index of each author) and, even worse, with citations that persist for very long times.

References

Bossolasco, M. (1939). Geofisica pura ed applicata, In: Pure and Applied Geophysics, 1, pp. 1-2.

Brin, S., and L. Page (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine, in: Seventh International World-Wide Web Conference (WWW 1998), April 14-18, 1998, Brisbane, Australia.

Garfield, G. (1998). The Impact Factor and using it correctly, Der Unfallchirurg 101, 413-414.

Harzing, A.-W. (2010). The Publish or Perish Book. Your guide to effective and responsible citation analysis, Tarma Software Research Pty Ltd, Melbourne, Australia, 250 pp.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. USA, 102 (46), 16569-16572; doi: 10.1073/pnas.0507655102.

Corresponding author: Stefano Solarino,
Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Sede di Genova, Italy; e-mail: stefano.solarino@ingv.it