Arch. Immunol. Ther. Exp., 2008, **56**, 373–379
PL ISSN 0004-069X
DOI 10.1007/s00005-008-0040-3

A new generation of relational charts for comparative assessment of citation impact

Wolfgang Glänzel[1, 2], Andreas Schubert[2], Bart Thijs[1] and Koenraad Debackere[1]

[1] Katholieke Universiteit Leuven, Steunpunt O&O Indicatoren and Department of MSI, Leuven, Belgium
[2] Institute for Research Policy Studies, Hungarian Academy of Sciences, Budapest, Hungary

Received: 2008.10.07, **Accepted:** 2008.10.27, **Published online first:** 2008.12.01

Abstract

A common problem in comparative bibliometric studies at the meso and micro level is the differentiation and specialization of research profiles of the objects of analysis at lower levels of aggregation. In this study, institutional profile clusters are used to examine which level of the hierarchical subject classification should preferably be used to build subject-normalized citation indicators. It is shown that a set of properly normalized indicators can serve as a basis of comparative assessment within and even among different clusters, provided that their profiles still overlap and such comparison is thus meaningful. Using the example of 24 European universities, a new version of relational charts is presented for the comparative assessment of citation impact.

Key words: citation analysis, scientometric indicators, subject normalization, relational charts.

Corresponding author: Prof. Wolfgang Glänzel, Katholieke Universiteit Leuven, Belgium, Institute for Research Policy Studies, Hungarian Academy of Sciences, Budapest, Hungary; e-mail: Wolfgang.Glanzel@econ.kuleuven.ac.be

INTRODUCTION

The evaluation of the research performance of institutions and research teams has become one of the principal tasks of present-day bibliometrics. The bibliometric toolbox comprising appropriate standard indicators was developed very early in our field, but the correct application of these tools to different levels of aggregation is still a challenge. Thus bibliometric meso- and micro-level studies need different standards from, for instance, those necessary for comparative macro studies (cf. Glänzel [3]). These requirements include such issues as statistical reliability based on the lower publication counts and advanced techniques of data-cleaning, address and author identification, and reliable subject delineation. Another issue arising from institute- or group-specific specialization or diversification is the necessity of classification according to institutional research profiles (Thijs and Glänzel [16]). While medium-sized and large countries are active in practically all fields of the sciences, institutions and research groups usually have more specific publication profiles.

The relative indicators developed at ISSRU (Budapest) in the 1980s and preferably presented in relational charts (e.g. Schubert and Braun [14]) proved to be useful instruments in cross-national comparisons of national research performance. In cross-institutional comparisons, where even multidisciplinary profiles do not necessarily cover all areas of science or may cover most areas but not to the same extent, these relational charts might not reflect citation impact in an adequate manner and might thus distort the interpretation. In the present study we therefore propose the application of the following three solutions.

The first solution refers to clustering research institutes by their research profiles in order to compare like institutes with each other (cf. Thijs and Glänzel [16]). Since some of the clusters represent specialized, while others rather multi-disciplinary institutions with medical, technical, or general profiles, a second solution is proposed. While in relational charts the mean observed citation rate (MOCR) is plotted against the journal-based mean expected citation rate (MECR), a new version of relational charts uses subject-normalized observed and expected citation rates to avoid possible biases caused by different activity profiles. More than European 2000 universities and other research institutes are used to determine at which level (major fields, subfields, or ISI subject categories) normalization should be done. The second question concerns the underlying publication period and the citation window. Two different windows, a 3-year and a 5-year citation window, are applied here to the above “sample” in order to analyze the robustness and reliability of the indicators at this level of aggregation.

The citation impact-related charts are supplemented by relational charts reflecting the relative share of highly cited papers published by institutes and research teams as well as the share of citations with respect to the international reference standard. This forms the third of the above-mentioned approaches. The method of characteristic scores and scales, providing subject-sensitive and self-adjusting thresholds, is finally used to determine appropriate thresholds for the selection of highly cited papers. The methodological part is supplemented by a comparative analysis of the citation impact of selected European research universities.

DATA SOURCES AND DATA PROCESSING

All data used for this study were extracted from the yearly updates of the Web of Science database of Thomson Scientific (Philadelphia, PA, USA). Only papers of the document types article, letter, note, and review indexed in the 1999–2001 volumes were selected. Publications were assigned to countries and institutions according to the address in the by-line of the paper. National thesauri produced at the Steunpunt O&O Indicatoren (SOOI) at KU Leuven on the basis of corporate address data were used to identify and assign publications to European institutions.

METHODS AND RESULTS

Subject diversification and profile analysis of research institutions

Unlike in the macro case, where the national research systems of medium-sized and large countries are principally expected to reflect multidisciplinarity, the question of diversification and specialization becomes an important issue in the evaluation of institutional research. The breakdown by research fields for comparative analysis alone does not grasp and reflect the complexity and subject diversification of the institutions’ research activities in an adequate manner. To overcome this deficiency we searched for a solution for identifying institutions with similar research profiles. In earlier papers (e.g. Thijs and Glänzel [16]) we developed a method to classify and map the European and national institutional landscapes on the basis of the individual institutes’ research profiles. This method, which consists of three steps, is briefly summarized in the following. The first step in our procedure was to break down the publication output of each institute into research fields and to construct their individual publication profiles. For this exercise we used the above-mentioned 16 major fields of the sciences, social sciences, and humanities developed at KU Leuven and ISSRU-Budapest (Glänzel et al. [10]). To avoid distortions, small institutes with publication output below a given threshold were removed. In the second step a cluster analysis using the Ward algorithm with squared Euclidean distances was conducted on these research profiles. The Je(2)/Je(1) index introduced by Duda and Hart [2] was used to find the optimum solutions. Beyond the almost trivial case of two clusters (medical and non-medical institutes), we found eight clusters as the second optimum solution. This classification is presented in Table 1. We would like to stress that all clusters still have a considerable overlap in their research profiles. This is a precondition for building relative citation indicators for possible application to both intra- and interclass comparison. Research performance of institutions with (almost) completely different profiles should, however, not be subjected to direct comparisons since, for instance, comparing a medical university with a business school is an exercise of “comparing apples and oranges” and would therefore not make sense.

Within the framework of this study we use the profile clusters only to answer the question of which field-depth should preferably be used for subject-based normalization of citation indicators to meet the requirements of diversification and specialization at the institutional level.

Possible levels of subject assignment for normalization of citation indicators

It is a commonly known fact that the subject-specific peculiarities in the publication and citation behavior of scientists result in considerable differences in the standards of publication activity and citation-impact of the different subject areas. Thus we have to expect different citation standards in the individual profile clusters as well. A further question arises from the classification shown in Table 1, in particular whether research profiles of the individual fields possibly differ among clusters. In other words, does chemistry or physics research have the same profile in the chemistry cluster 6 as in the multidisciplinary cluster 3 or the technical and natural sciences cluster 4? A similar question can be addressed regarding medical research in clusters 3, 7, and 8. Although this question seems to be rather rhetorical, in practice it might become quite significant if the publication output of different clusters is broken down by fields and these fields serve as the basis for inter-cluster comparisons. The reason therefore lies in the heterogeneity of major fields in terms of scientists’ communication behavior. According to the above-mentioned Leuven/Budapest subject-classification scheme, which uses the ISI Subject Categories as the lowest hierarchic level, the highest level consists of 12 major fields in the sciences and applied sciences and the medium level comprises 60 subfields (see Glänzel and Schubert [10]).

In their recent paper, Thijs and Glänzel [17] illustrated, using the example of chemistry, that the subfield profiles of chemistry research do indeed vary consider-

Table 1. The eight clusters resulting from the second optimum solution

Cluster	Code
Cluster 1 (Biology)	BIO
Cluster 2 (Agriculture)	AGR
Cluster 3 (Multidisciplinary)	MDS
Cluster 4 (Geo & Space Science)	GSS
Cluster 5 (Technical & Natural Sciences)	TNS
Cluster 6 (Chemistry)	CHE
Cluster 7 (General & Research Medicine)	GRM
Cluster 8 (Specialized Medicine)	SPM

ably among the individual clusters. Thus physical chemistry was, for instance, almost absent in the medical clusters, whereas in the Technical and Natural Sciences and Chemistry clusters, organic and medicinal chemistry was less predominant than in the medical groups. The analysis of institutional research performance by subfields or even by ISI Subject Categories is because the often small publication sets resulting from the breakdown are less practicable at this aggregation level. However, this does by far not suffice to eliminate all profile-specific biases. The question therefore arises of how and at which level citation data are to be normalized in order to "equalize" the above-mentioned citation biases. One normalized indicator that has been successfully used is beyond any doubt the subject-based normalized mean citation rate (NMCR, see, for example, Braun and Glänzel [1]). A similar measure (CPP/FCSm) is used at CWTS (cf. Moed et al. [13]). These indicators have in common that the MOCR is gauged against its individual expectation on the basis of the subject to which it belongs. In the light of the above brief discourse on subject-related issues, the question arises which hierarchic subject level should preferably be used as the reference standard for the definition of such an indicator for application of institutional citation impact.

The above examples suggest that the highest level, i.e. the major fields, might be too coarse for this exercise. Moreover, one intuitively expects that using the lowest level, providing a fine-grained subject assignment, might be the most accurate approach. However, taking into account that many journals covered by the Web of Science are assigned to 4, 5, and even more ISI Categories each and the assignment of publications into subfields is by far less fuzzy, it becomes conceivable that subfields could serve as the favored reference level. A consequence of multiple assignments is the necessity of fractionating subject assignment and thus of calculating weighted averages for the corresponding individual field-expected citation rates. Such averaging of subject standards still occurs at the level of subfields and major fields as well, but to a much lesser extent as in the case of the ISI Categories.

Now we shall summarize the outcomes of a systematic comparison of the subject-normalized citation impact of European research institutions on the basis of the three different hierarchic levels of subject assignment. For this exercise we selected those 676 of the about 2000 European institutions which have published at least 50 papers in the period of 1999–2001. We applied 3-year citation windows beginning with the publication year. The pair-wise linear regression between the NMCR of the 676 institutions calculated on the bases of the three different subject levels showed the strongest correlation between the subfields and the ISI Categories. The correlation between the lowest and the highest hierarchic level provided the poorest results (cf. Glänzel [7]). These results become even more distinct if the eight profile clusters are studied separately. While the linear regression model provides almost the same patterns for the multidisciplinary cluster, we found Geo and Space Science the most problematic one. We present the scatter plots and the linear regression for three clusters in Fig. 1. The fact that the slope of 1.4 distinctly exceeds the value of 1.0 in the case of cluster 4 clearly demonstrates the inconsistency of the major-field model (Fig. 1, bottom right).

Summarizing, we can conclude that we obtain consistent results for the calculation of subject-normalized citation indicators for both the ISI Subject Categories and the subfields. However, to avoid exaggerated fractionation, weighting, and averaging for multiple subject assignment, we prefer the use of subfield-based standards. All further analyses will be based on this approach.

Choice of appropriate citation windows for the analysis of institutional research performance

Before we proceed further with the necessary theoretical and methodological considerations, we have to clarify an important validity question. In particular, the question of which citation window should underlie the citation analysis becomes more significant, and also more critical, at the lower levels of aggregation. No doubt a larger citation window results in a higher reliability of indicators than a shorter one. Also, the often-heard argument that the usual windows of three to five years are not sufficient in specific fields of the applied sciences and in mathematics suggests the application of citations windows as large as possible. On the other hand, science policy and research management is interested in the evaluation of the most recent results. The standard solution can therefore only be a compromise (Glänzel [4]). Thus the three-year citation window has, as larger windows, long been used successfully at different levels of aggregation (see, for instance, Moed [12], Glänzel [5], Schubert and Glänzel [15], van Raan [18, 19]).

Nevertheless, we will have a closer look at how sensitive subject-normalized citation indicators based on the subfield level are to different initial citation periods. We used a 3- and 5-year window for this exercise. Figure 2 presents the scatter plot of the NMCR based on a 5-year citation window vs. a 3-year window for the 676 European research institutions with at least 50 papers

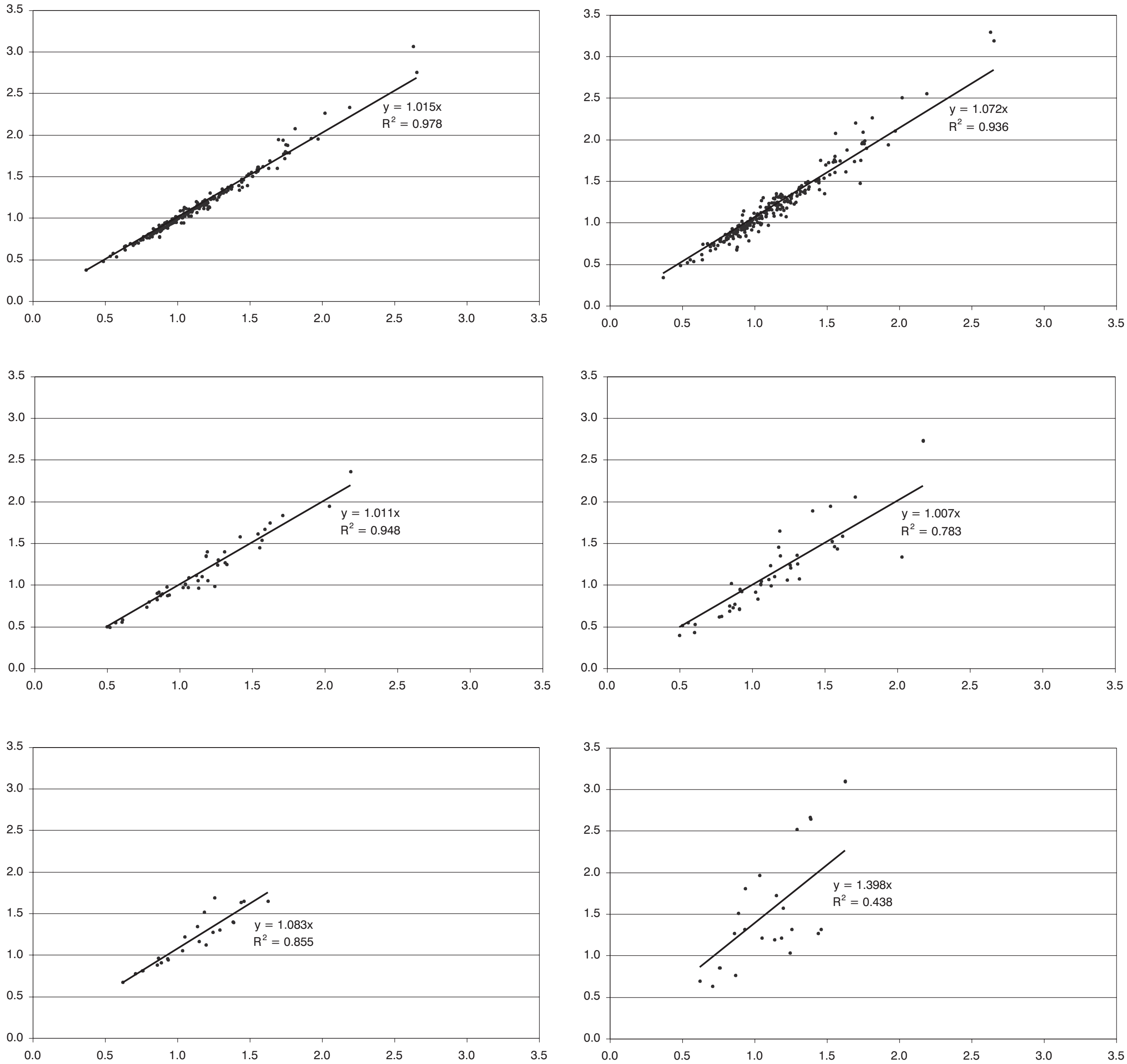

Fig. 1. Plot of NMCR based on subfields vs. ISI Subject Categories (left) and major fields vs. ISI Subject Categories (right) for institutions of Cluster 3 (top), Cluster 1 (center), and Cluster 4 (bottom).

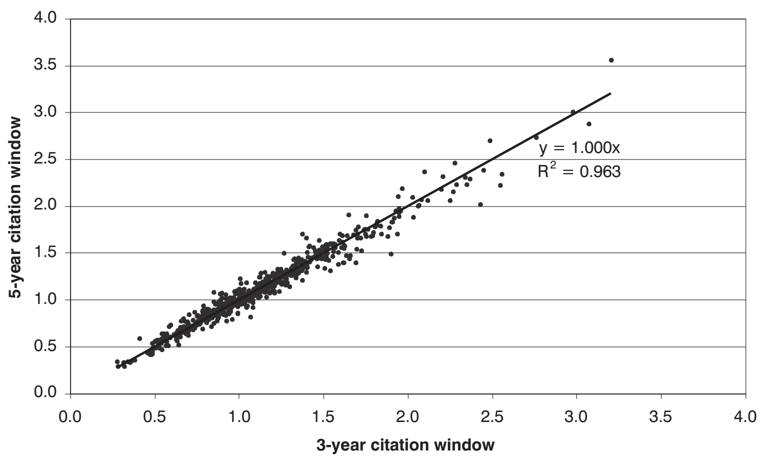

Fig. 2. Plot of NMCR based on 5-year vs. 3-year window for 676 European institutions.

each in the period of 1999–2001. The practically perfect fit of the linear regression model with very strong correlation shows that at this level, too, a 3-year window suffices to measure citation impact in an adequate manner. In verbal terms, the growth of the observed citation rates on average parallels that of their subfield-based expectations. We will therefore use a 3-year citation window in the following sections of this study.

Characteristic scores and scales for highly cited publications

We will use the power-law property of the tail of the citation distributions to derive a straightforward rule for the determination of highly cited articles from the method of characteristic score and scales (Glänzel and Schubert [8], Glänzel [6]). This method can be briefly summarized as iteratively truncating samples at their mean value, b_1, and recalculating the mean ($k>1$) of the truncated sample until the procedure is stopped or the sample is empty. In the following we outline the procedure using the example of citation rates of a given paper set preferably published in the same topic or subfield. The procedure results in subdividing the originally distribution into the following different classes (zones): [0, b_1) "poorly cited", [b_1, b_2) "fairly cited", [b_2, b_3) "remarkably cited", and [b_3, ∞) "outstandingly cited" papers.

As a consequence of a characterization theorem for Paretian distributions by Glänzel et al. [11], the conditional expectations b_k can be expressed by $b_k=(a^0+a^1+\ldots+a^{k-1})\cdot b_1$, where the variable a depends only on the parameter of the power-law distribution. Empirical studies have shown that $a\approx 2$ for citation windows of 3–5 years (Schubert and Glänzel [15], Glänzel [7]). In this case, we obtain $b_2=3b_1$ and $b_3=7b_1$, and so on. For the institutional level we suggest the use of the two groups of "remarkably cited" and "outstandingly cited" papers as highly cited with the corresponding thresholds $k=2$ or $k=3$, respectively. In an earlier study (Glänzel [6]) it was shown that the scores b_k are time dependent and, of course, strongly sensitive to the subject matter as well, but the share of papers in the individual zones are relatively stable over the citation window and also do not noticeably vary among different subfields. Thus indicators on highly cited papers defined on the bases of characteristic scores and scales can *per se* be considered subject normalized.

The reference standard of the share of outstandingly cited papers usually ranges between 1% and 2% and that of remarkably cited papers amounts to about 5% (e.g. Glänzel [6]). As a consequence of this rule, we have to restrict studies of highly cited papers to institutions with at least 1000 paper each for outstandingly cited papers or to those with at least 200 papers in the case of remarkably cited papers. Otherwise the institutional sets of highly cited papers are not expected to have more than 10 publications each. Paper sets of such small size are, however, no longer appropriate for the statistical citation analysis.

Subfield-normalized relational charts for comparative assessment of citation impact

In this section we finally apply the above methodological results to a concrete selection of research universities. For this exercise we selected two large universities each from twelve medium-sized European countries. The countries chosen were Austria, Belgium, Denmark, Finland, Hungary, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden, and Switzerland. Since the largest universities (in terms of publication output) usually have multidisciplinary research profiles, we replaced some of the originally selected universities by medical or technical universities with similarly large publication output. Thus we have two medical universities and three technical universities in the selection.

The publication output of the selected institution ranges between about 1200 and 8500 papers in the period of 1999–2001. Since the selection due to the above-mentioned "correction" is somewhat arbitrary and the intention of this study is merely to present the methodology using the institutions as samples, the universities will be represented anonymously by letters. The two medical universities are U and X and the technical universities are G, T, and Y. A comparative study of a more exhausting and non-anonymous selection of universities is in preparation. Before we present the citation data, we briefly summarize all indicators used within this model. All the indicators are based on three-year citation windows.

1. The mean observed citation rate (MOCR) is defined as the ratio of citation count to publication count.
2. Mean expected citation rate (MECR). The expected citation rate of a single paper is defined as the average citation rate of all papers published in the same journal.
3. Field expected citation rate (FECR). Analogously to the previous indicator, the FECR of a single paper is defined as the average citation rate of all papers published in the same subject in the same year.
4. The ratio of the two previous indicators (MECR/FECR) expresses whether the unit under study publishes on average in higher (lower) impact journals than expected on the basis of the subfields where the unit was active.
5. Normalized mean citation rate (NMCR). This indicator is defined as the ratio of the observed and field-based expected citation impact. NMCR gauges citation rates of the papers against the standards set by the specific subfields.
6. Relative citation rate (RCR). RCR is defined as the ratio of the observed and journal-based expected citation impact. This indicator measures whether the publications of the unit under study attract more or fewer citations than expected on the basis of the journal impact measures of the journals in which they appeared.
7. High-impact activity is the ratio of the unit's share of highly cited papers in all papers and the correspond-

ing world standard. This measure was introduced by Glänzel and Schubert [9], but applied to a different definition of high impact. Here we base this measure on the definition given in the previous section and use the threshold of $k=3$ to select outstandingly cited papers for high-impact analysis.

8. High-impact attractivity is defined analogously to the previous indicator, in particular as the ratio of the unit's share of citations attracted by its highly cited papers in all citations received by the unit under study and the corresponding world standard. The same conditions and threshold as in the definition of the previous indicator have to be used.

Relative indicators are preferably presented in a relational chart (see Schubert and Braun [14]). Subfield-normalized relational charts are subdivided into six sections by the main diagonal starting from the origin and by adding the two straight lines indicating the conformity with the reference standards for observations and expectations (Braun and Glänzel [1]). First the scatter plot of (MECR, MOCR) values is presented in a traditional relational chart (Fig. 3). Both expectation and observation cover a quite large range of citation impact. The fact that U is a medical university but Y is a technical university certainly contributes to the huge deviations of their citation impact from each other. We also mention that both impacts (MECR and MOCR) of U considerably exceed those of the technical university G, and we can observe a similar situation for X and Y, albeit on a much lower level. Also, T, as a technical university, appears in the low-end group of this diagram.

We obtain a completely different situation if subfield-based normalization is applied. The plot of subfield-normalized observation against subfield-normalized expectation is presented in Fig. 4. The positions of universities G and U have interchanged; the same applies to X and Y. The effect of the field-specific lower impact of technical universities and the usual high impact of medical of medical universities has thus been eliminated. X and Y take almost the same position in the lower-left section of the chart. The position of university T has changed as well; in the new chart it has joined the group in the most advantageous section. Some changes effect also universities P and F, both, however, with outstanding citation impact.

To deepen these results we present the plot of high-impact activity vs. high-impact attractivity for these universities (see Fig. 5). This chart mirrors the previous diagram apart from the fact that the attractivity of the high-end group does not quite meet their expectation set by their high-impact activity. Both the less favorable situa-

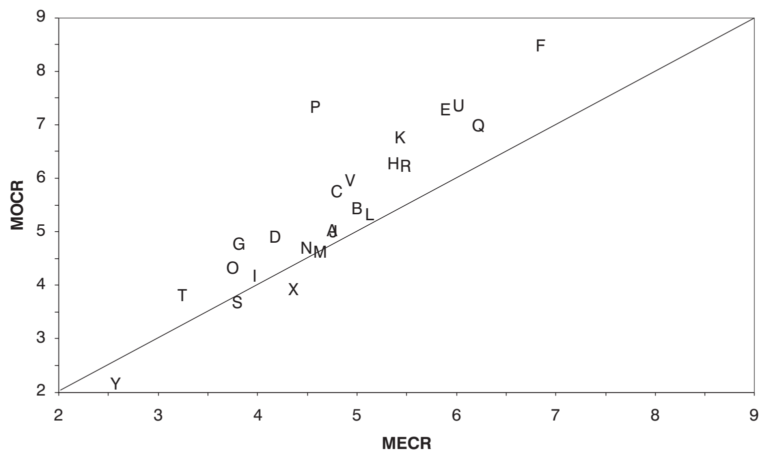

Fig. 3. Relational chart of expected and observed citation rate for 24 universities.

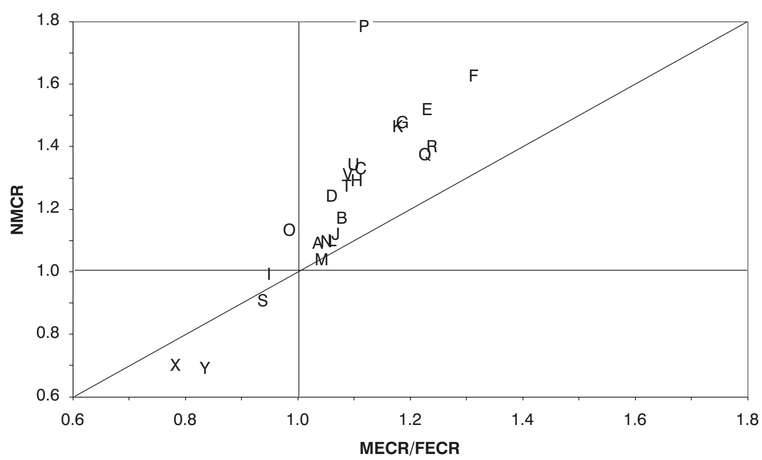

Fig. 4. Subfield-normalized relational chart for 24 universities.

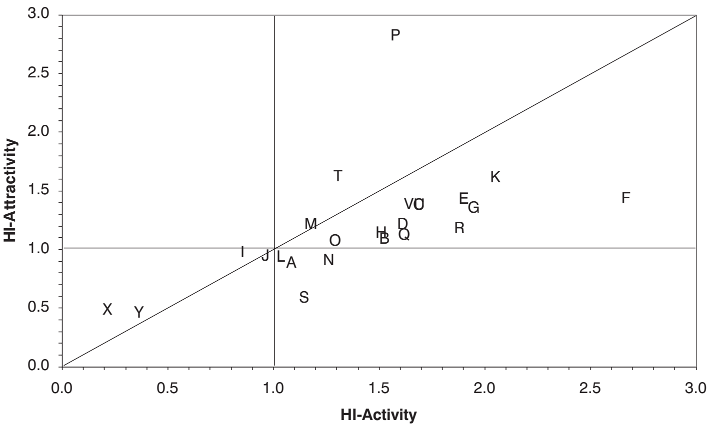

Fig. 5. Relational chart of high-impact activity and attractivity for 24 universities.

tion of universities X and Y as well as the outstanding citation impact of F and especially that of P are confirmed by this diagram as well.

CONCLUSIONS

Both the three-level hierarchical subject-classification scheme based on the ISI Subject Category system and the institutional profile clusters developed at SOOI/KULeuven were used to determine the appropriate field-depth for the calculation of subject-normalized citation indicators. The level of 60 subfields proved to be an acceptable choice and to provide stable and consistent results. The normalized indicators can be used for both intra- and inter-cluster comparative analysis as well as for domain studies at this level of aggregation. In addition to the relative citation indicators, the high-end of institutional research is best reflected by highly cited papers. Two thresholds are currently suggested, depending on the underlying publication output. The study has also shown that the use of a 3-year citation window suffices for building relative and high-impact citation indicators.

Acknowledgment: An earlier version of the paper was presented by the first author at the "Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología" (Havana, Cuba, 23–25 April 2008). An extended version appears in Scientometrics in 2009 (DOI 10.1007/ /s11192-008-2109-5).

REFERENCES

1. Braun T. and Glänzel W. (1990): United Germany: The new scientific superpower? Scientometrics, **19**, 513–521.
2. Duda R. O. and Hart P. E. (1973): Pattern Classification and Scene Analysis. Wiley, New York.
3. Glänzel W. (1996): The need for standards in bibliometric research and technology. Scientometrics, **35**, 167–176.
4. Glänzel W. (1997): On the reliability of predictions based on stochastic citation processes. Scientometrics, **40**, 481–492.
5. Glänzel W. (2000): Science in Scandinavia: A bibliometric approach. Scientometrics, **48**, 121–150. (Correction: Scientometrics, **49**, 2000, 357)
6. Glänzel W. (2007): Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. J. Informetrics, **1**, 92–102.
7. Glänzel W. (2008): On some new bibliometric applications of statistics related to the h-index. Scientometrics, **77**, 187–196.
8. Glänzel W. and Schubert A. (1988): Characteristic scores and scales in assessing citation impact. J. Information Sci., **14**, 123–127.
9. Glänzel W. and Schubert A. (1992): Some facts and figures on highly cited papers in the sciences, 1981–1985. Scientometrics, **25**, 373–380.
10. Glänzel W. and Schubert A. (2003): A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. Scientometrics, **56**, 357–367.
11. Glänzel W., Telcs A. and Schubert A. (1984): Characterization by Truncated moments and its application to Pearson-type distributions. Z. Wahrscheinlichkeitstheorie verwandte Gebiete, **66**, 173–183 (correction: Probability Theory and Related Fields, **74**, 1987, 317.)
12. Moed H. F. (1996): Differences in the construction of SCI based bibliometric indicators among various producers: A first over view. Scientometrics, **35**, 177–191.
13. Moed H. F., De Bruin R. E. and Van Leeuwen T. N. (1995): New bibliometric tools for the assessment of national research performance: database description, overview of indicators and first applications. Scientometrics, **33**, 381–422.
14. Schubert A. and Braun T. (1986): Relative indicators and relational charts for comparative-assessment of publication output and citation impact. Scientometrics, **9**, 281–291.
15. Schubert A. and Glänzel W. (2007): A systematic analysis of Hirsch-type indices for journals. J. Informetrics, **1**, 179–184.
16. Thijs B. and W. Glänzel W. (2008): A structural analysis of publication profiles for the classification of European research institutes. Scientometrics, **74**, 223–236.
17. Thijs B. and Glänzel W. (2009): A structural analysis of benchmarks on different bibliometric indicators for European research institutes based on their research profile, Scientometrics (in press).
18. Van Raan A. F. J. (2006): Statistical properties of Bibliometric indicators: Research group indicator distributions and correlations. J. Am. Soc. Information Sci. Technol., **57**, 408–430.
19. Van Raan A. F. J. (2006): Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics, **67**, 491–502.