

Improved author profiling through the use of citation classes¹

Bart Thijs*, Koenraad Debackere**, Wolfgang Glänzel***

* *Bart.Thijs@kuleuven.be*

Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven, Leuven (Belgium)

** *Koenraad.Debackere@kuleuven.be*

Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven, Leuven (Belgium)

*** *Wolfgang.Glänzel@kuleuven.be*

Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven, Leuven (Belgium)

Department of Science Policy & Scientometrics, Library of the Hungarian Academy of Sciences, Budapest (Hungary)

Introduction

At the level of research teams and individual scientists bibliometric standard indicators are easily distorted by the citation impact of papers that considerably differ from their expected value. In previous studies by the authors (Glänzel, 2013; Glänzel et al., 2014), a parameter-free solution providing four performance classes has been proposed to replace mean-value based indicators by performance classes. This has been found useful at the level of national and institutional assessment of citation impact. The approach, which is based on the method of Characteristic Scores and Scales (CSS) that has originally been introduced by Glänzel & Schubert (1988) in the context of journal analysis, has several striking advantages over traditional methods. The application of these performance classes completely eliminates the effect of the heavy tails of citation distributions including their typical outlier-based biases at lower levels of aggregation. Unlike in methods based on percentiles (e.g., Leydesdorff et al. (2011), this approach is not sensitive to ties and ensures seamless integration of measures of outstanding and even extreme performance into the standard tools of scientometric performance assessment. The method can be interpreted as a reduction of the original citation distribution to a distribution over a given number of performance classes with self-adjusting thresholds without requiring arbitrarily pre-set values. The method proved thereby to be insensitive to both subject-specific peculiarities and the particular choice of publication years and citation windows. The application to the micro-level, that is, to the level of individual scientists and research teams, however, still remained a challenge. Career evolution of individual scientists and changing team composition along with the typically small paper sets underlying the citation distribution at this level require extremely stable and robust solutions. In the present study we will analyse in how far the CSS-based method meets these requirements, and will give examples for its application to the level of individual scientists. We will also show that the method can be applied using reference standards defined by any appropriate base-line distribution forming a superset of the publication set under study.

An additional advantage is that the classes can, because of the high robustness of the distribution of papers across fields (see Glänzel et al., 2014), be calculated at any level of field aggregation and that multiple field assignments do not hinder the calculation of the specific thresholds and the performance scores for individual authors. Furthermore, only the calculated

¹ All data presented in the tables and figures of this study are based on data sourced from Thomson Reuters Web of Science.

threshold values are needed in a real world application of the method in an evaluative exercise without the underlying citation distributions.

Data

The data set was built on downloaded Researcher-ID data of 4.271 registered researchers from eight selected countries. For reasons of statistical reliability only authors with at least 20 publications were taken into consideration for the retrieval. This selection of data creates a build-in bias towards more prolific and excellent authors. Heeffter et al., (2013) showed that authors registered for a Researcher ID are indeed more productive than others. Also the applied threshold of 20 publications implies the exclusion of many occasional authors whose impact is rather limited. However, as the goal of this paper is not to define any particular reference standard but to investigate the applicability of the CSS-method, we can proceed from these data sets without loss of generality. Moreover, it can be expected that in a real-world evaluative exercise or selection procedure, the studied authors are indeed the more active ones.

Publication data were matched with data retrieved from Thomson Reuters Web of Science (WoS). Only journal publications indexed as article, letter, note or review between 1991 and 2010 are taken into consideration. Papers are assigned on the basis of the journals in which they appeared to subfields according to the Leuven-Budapest scheme (Glänzel & Schubert, 2003). Also papers only assigned to Social Sciences or Humanities have been excluded in the analysis. Moving three-year citation windows were used throughout the analysis.

Methods and results

First we briefly recall the outlines of the model (cf. Glänzel, 2013). Characteristic scores are obtained by iteratively calculating the mean value of a citation distribution and subsequently truncating this distribution by removing all papers with less citations than the conditional mean. As described earlier, the process is stopped after three iterations. This results in three scores b_k with ($k=1, 2, 3$). By adding $b_0=0$ and $b_4=\infty$, four distinct performance classes can be created each defined by a pair of threshold values $[b_{k-1}, b_k)$ with ($k=1, 2, 3, 4$). This definition solves the problem of ties that otherwise might occur in ranking approaches as each paper can be uniquely assign into one of these four half-open intervals.

These scores are now calculated at the field level and for each publication year. As papers can be assigned to multiple fields, this requires some special attention both in the calculation of the scores as with the attribution to the performance classes. The applied methodology is similar as the calculation of the subfield-expected citation rate. The contribution of a paper assigned to multiple subfields is fractionated based on the number of assigned fields. This means, for instance, that a paper classified in two subfields counts only as a half in the numerator of the mean calculation. And only the half of its citations contributes to the denominator. A detailed description of this method can be found in a recent study by Glänzel et al., (2014).

The results of the CSS-based methods are also gauged against traditional measures. For this purpose we use normalized citation rates, particularly Relative Citation Rate (RCR) and the Mean Normalized Citation Rate (NMCR) (Glänzel et al., 2009). The first indicator compares the observed citation rate to a journal based expected citation rate while the second one uses a subfield expected citation rate. It is important to mention here that score b_1 is in fact an expected citation rate that depends on the given reference distribution.

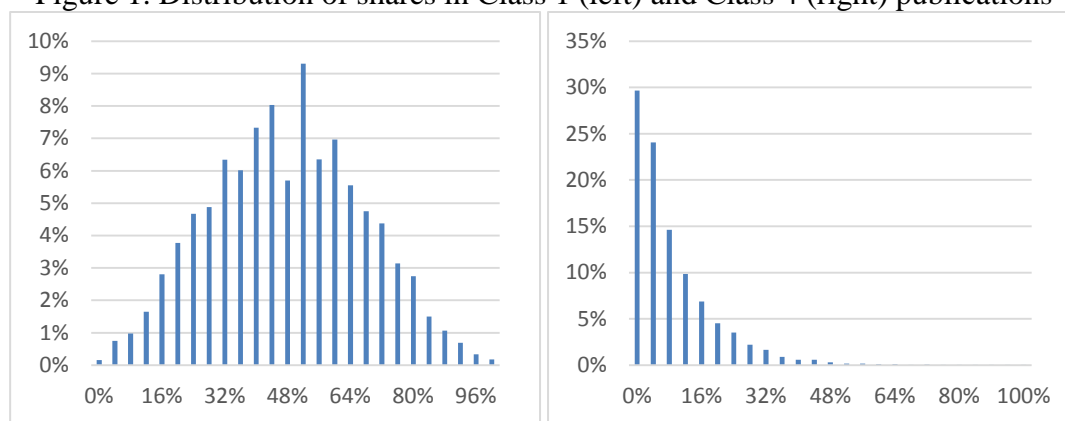
Performance Classes

The first result we present is the share of publications in each performance class in Table 1. The first row contains the average over all the distinct authors of the distribution of papers across classes. And the second row presents the share of the distinct classes within the total set of 116.467 unique papers. This data is extended in this table with the results for the distribution over classes of all the publication data indexed in 2010 in the WoS database. Because of the robustness of the method this could be considered a suitable reference standard (cf. Glänzel, 2007 and 2013). The shares in the last row are in line with earlier reported shares (see Glänzel et al., 2014). Also Albarran & Ruiz-Castillo (2010) found the same 70-21-9 rule where they merged the two upper classes for publications indexed between 1998 and 2002 with a 5-year citation window and classified into 22 fields according to Thomson Reuters *Essential Science Indicators* (ESI). As mentioned above, only prolific authors are included in the analysis as the calculation of shares within small publication sets is otherwise subject to possible fluctuations, which might distort the resulting statistics and thus reduce their statistical reliability. As a consequence of this reliability-related selection criterion shares are shifted towards the upper classes as compared with the reference standard of the complete population. This “bias” can also be observed in the citation indicators for the set of papers of the selected authors (RCR=1.42 and NMCR=2.10). The implications of this discrepancy will be discussed later with respect to the choice of benchmarks for comparison in particular applications.

Table 1. Distribution of papers over performance classes.

Data Set	Class 1	Class 2	Class 3	Class 4
Average over all authors with R-ID	42.8%	33.1%	15.0%	9.1%
All papers of authors with R-ID	44.3%	32.9%	14.5%	8.4%
All papers indexed in 2010	69.8%	21.5%	6.2%	2.5%

Figure 1. Distribution of shares in Class 1 (left) and Class 4 (right) publications



The distribution of shares in the first two classes proved to be normal. The third class deviates slightly but the last class deviates strongly from a normal distribution. In order to illustrate this effect we show the patterns for Class 1 and 4 in Figure 1. The strong deviation from normality in Class 4 reflects once again the problematic behaviour of extreme values and their particular distribution as already reported by Glänzel (2013).

Citation Indicators

In a second analysis we compare the distribution across classes with the traditional citation indicators. A Spearman rank correlation is calculated among the classes. The results are presented in Table 2. The first class is negatively correlated with the three other classes. The negative correlation of the first class with the other ones *and* the traditional relative indicators is in line with our expectations since the first class relates to poorly cited papers. It is striking that Class 2 is not correlated at all with the upper two classes and with the RCR and that it has a moderate correlation with the other citation indicators. Both classes 3 and 4 have higher correlations with the citation indicators but have an inter-correlation of 0.48. These observations substantiates that citation behaviour cannot sufficiently be represented by one class or any individual indicator alone. This is a strong argument for the choice of this method with four performance classes at this aggregation level too.

Table 2. Rank correlation between performance classes and citation indicators for data set 1.
(values marked with * do not statistically deviate from 0)

	Class 1	Class 2	Class 3	Class 4	RCR	NMCR
Class 2	-0.459					
Class 3	-0.743	0.058*				
Class 4	-0.685	-0.055*	0.485			
RCR	-0.630	-0.001*	0.547	0.734		
NMCR	-0.839	0.125	0.693	0.863	0.786	
NMCR/RCR	-0.632	0.202	0.482	0.547	0.163	0.701

Author profiling

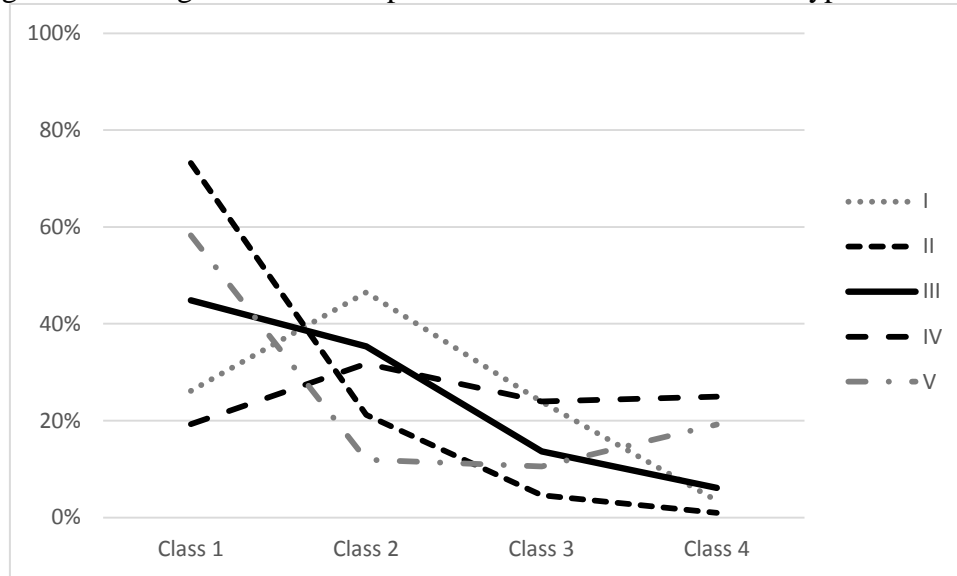
As already described in an early study (Glänzel et al., 2014), different profiles according to the deviation from the base-line distribution in each class are possible. Unlike in the case of traditional indicators, where just higher/lower than the expectation can occur, here a variety of deviations are possible. Each author can thus be characterized by an individual profile indicating the distribution of his papers over the four performance classes. The advantage of these profiles is that they enable a direct comparison between distinct authors but it may also be applied for comparison with an appropriate reference distribution.

First, a χ^2 -test indicates whether or not an individual profile deviates from the reference (at a confidence level of $p = 0.05$). The null hypothesis assumes that the distribution of an author's papers across these four classes is consistent with the chosen reference. In our case the number of publications instead of the shares is used to take the size of the publication set into account. Only if H_0 is rejected, the deviation in each of the four classes from the reference is used to identify performance types to be distinguished.

- I. The χ^2 -test indicates that the profile deviates from the reference and the shares of both classes 1 and 4 are lower than the reference.
- II. The χ^2 -test indicates that the profile deviates from the reference and the share of class 1 is higher and class 4 is lower than the reference.
- III. The χ^2 -test indicates that the profile does not deviate from the reference.
- IV. The χ^2 -test indicates that the profile deviates from the reference and the share of class 1 is lower and class 4 is higher than the reference.
- V. The χ^2 -test indicates that the profile deviates from the reference and the shares of both classes 1 and 4 are higher than the reference.

We present the classification of each of the 4271 authors according to these profile types as an example. The choice of an appropriate reference is crucial here. It is clear that the total set of publications as presented in the last row of table 1 is not suitable as our sample consists of authors with a more prolific and excellent profile. Therefore, the distribution of the total set of unique publications published by at least one of the authors was taken as reference standard here (see second row in Table 1). The χ^2 -test indicates whether or not the individual author profile deviates from this reference standard. Six out of ten authors have a profile that does not significantly deviate from the reference and are classified as type III. Type II and IV contain respectively 13.6% and 19.8% of all authors. The two remaining types I and V comprise a much smaller number of authors. Figure 2 presents the average share within each class for the five types. The solid line of type III shows a neat decline.

Figure 2. Average share of four performance classes across five types of authors



These types do not only have different shares among the performance classes nor on other indicators – as Table 3 shows. Type III authors have, on an average, less publications than the other types. This is reasonable as the sensitivity of the χ^2 -test increases with higher number of publications. This effect thus confirms the reliability of the applied classification rule. Another striking observation is the high values for the citation indicators for type I. The lack of highly cited publications in the fourth class is here compensated by the low share in Class 1 publications and high share in Class 2. These publications in Class 2 have already received citation rates higher than the expected citation impact used for the calculation of the citation indicators in Table 3.

Table 3. Standard indicators for the five profile types

	I	II	III	IV	V
Number of authors	211	579	2603	845	33
Average Publication	37.6	38.3	30.7	37.8	31.6
Average RCR	1.33	0.88	1.28	2.22	3.39
Average NMCR	2.06	0.85	1.80	4.36	5.87
Average NMCR/RCR	1.62	0.99	1.44	2.01	1.49

To conclude the section on author profiling, Table 4 presents a sample of 20 authors taken from data set 1. The total set of unique publications published by at least one of the authors from data set 1 are taken as benchmarks and its distribution over the four classes is taken as reference point.

Table 4. Average distribution of papers over performance classes.

Author	Class 1	Class 2	Class 3	Class 4	RCR	NMCR	NMCR/RCR	Type
1	31.8%	27.3%	36.4%	4.5%	1.32	2.01	1.53	I
2	39.3%	60.7%	0.0%	0.0%	1.13	1.24	1.09	I
3	32.5%	45.0%	22.5%	0.0%	0.98	1.59	1.61	I
4	40.0%	24.4%	33.3%	2.2%	1.36	1.65	1.21	I
5	67.5%	28.6%	1.3%	2.6%	1.30	1.01	0.77	II
6	58.3%	32.5%	7.3%	2.0%	1.00	1.18	1.19	II
7	74.2%	16.1%	9.7%	0.0%	1.21	0.75	0.62	II
8	72.0%	20.0%	8.0%	0.0%	0.65	0.82	1.27	II
9	68.2%	27.3%	0.0%	4.5%	0.82	0.79	0.96	III
10	33.3%	52.4%	9.5%	4.8%	1.56	1.82	1.17	III
11	61.9%	38.1%	0.0%	0.0%	0.70	0.80	1.15	III
12	33.3%	28.9%	20.0%	17.8%	1.89	3.35	1.77	III
13	21.4%	21.4%	14.3%	42.9%	2.65	5.85	2.21	IV
14	9.4%	50.0%	21.9%	18.8%	1.68	2.72	1.62	IV
15	17.9%	14.3%	28.6%	39.3%	3.21	6.34	1.97	IV
16	30.0%	20.0%	40.0%	10.0%	1.80	2.29	1.27	IV
17	56.8%	22.4%	11.2%	9.6%	1.47	1.83	1.24	V
18	62.5%	16.1%	12.5%	8.9%	1.38	1.92	1.39	V
19	47.6%	9.5%	9.5%	33.3%	14.58	35.46	2.43	V
20	50.0%	10.7%	14.3%	25.0%	3.75	3.69	0.98	V
Reference	44.3%	32.9%	14.5%	8.4%				

Conclusions

In the present study we showed the general applicability of the CSS-method to the individual level and to author profiling of candidates with scientific excellence, in particular. Publications from authors with a Researcher ID and at least 20 registered publications were classified according to the field and year specific thresholds. The distribution of shares in the first two classes came close to normal and in the third class is deviated only slightly. But once again, it is the tail of citation distributions represented by Class 4 that showed strong deviation. When correlated with traditional citation indicators it became clear that the distribution over classes is only partially correlated with these, especially with the journal based citation indicator (RCR).

We also compared the individual authors with a chosen reference standard and could define five different profile types. In our example the reference was calculated based on all the publications of the selected authors. In this case, the set is a result from within selection procedure but the reference set could also be defined prior to the start of this procedure, e.g., publications from a certain country or institute. Two types, I and V have shares in the two outer classes that are both above or below the reference share. In both cases the presences/absence of publication in both classes compensates for the citation scores. Finally

we would like to stress that the reduction of this method to two instead of four performance classes would bring us back to system of traditional indicators. This and the above observations confirm the seamless integration of the CSS method into the standard toolset of scientometric research evaluation.

References

- Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *JASIST*, 62(1), 40–49.
- Glänzel, W., & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, 14(2), 123–127.
- Glänzel, W., Schubert, A. (2003), A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56 (3), 357-367.
- Glänzel, W. (2007), Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, 1(1), 92-102.
- Glänzel, W., Schubert, A., Thijs, B., & Debackere, K. (2009). Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of Institutional Research Performance. *Scientometrics*, 78(1), 165–188.
- Glänzel, W., Thijs, B., & Debackere, K. (2014). The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and Institutional Research Assessment. *Scientometrics*, DOI: 10.1007/s11192-014-1247-1, in press.
- Heffer, S., Thijs, B., & Glänzel, W. (2013). Are registered authors more productive? *ISSI Newsletter*, 9(2), 29-32.
- Leydesdorff, L., Bornmann, L., Mutz, R., & Opthof, T. (2011). Turning the tables on citation analysis one more time: principles for comparing sets of documents. *JASIST*, 62(7), 1370-1381.