THE IMPACT FACTOR AS A MEASURING TOOL OF THE PRESTIGE OF THE JOURNALS IN RESEARCH ASSESSMENT IN MATHEMATICS

Abstract. The (2-year) Impact Factor of Thomson-Reuters (IF) has become the fundamental tool for analysing the scientific production of academic researchers in a lot of countries. In this paper we show that this index and the ordering criterion obtained by using it is highly unstable in the case of mathematics, to the extent that sometimes no reliability can be assigned to its use. We explain the reasons of this behaviour by the specific properties of the mathematical journals and publications, attending mainly the point of view of the researchers in pure mathematics. Using the Journal Citation Report list of journals as a source of information, we analyse the stability in the position of the mathematical journals -the so called rank-normalized impact factor-, compared with journals in applied physics and microbiology during the period 2002-2012. Due to the lack of stability of the position of the journals of mathematics in these lists, we propose a "cumulative index" that fits better the characteristics of mathematical journals. The computation of this index uses the values of the IF of the journals in previous years, providing in this way a more stable indicator.

1. Background

As a consequence of a well stablished policy of the institutions in charge of the scientific research in most of the European countries that started at the end of the last century, the so called (2-year) Impact Factor (IF) is the most popular tool for measuring scientific activity (Shapiro, 1999; Bordons et al. 2002; Archambault and Larivière, 2009). Recall that the IF of a given journal is computed as the ratio among the number of citations to papers published in the journal in the last two years and the total number of citable papers published in the journal in these two years, and it is a widely used metric of journal prestige. As all researchers know, papers published in scientific journals are not considered any more of equivalent value when their scientific production is measured. Remarkably, revision by peers of the contents of the papers is not always considered as a suitable tool in research assessment, probably due to its high cost.

The reason for IF-based evaluation is clearly the need of the governments and the funding institutions in general of having an automatic criterion to assign a mark to each item, in order to avoid arbitrary conclusions and to make easy the evaluation process (Adam, 2002). From this point of view, the research assessment must be as simple and objective as possible. The Impact Factor satisfies this purpose if it is used to define an ordered list of good level journals. However, it must be taken into account that the IF was invented as a tool for editors, librarians and publishing houses, and not for evaluating the quality of particular articles (Garfield, 2006).

Although IF journal rankings can be an acceptable tool in some disciplines, its use in all the fields of pure mathematics (algebra, geometry, mathematical analysis, topology,...) is inadequate due mainly to the specific characteristics of the mathematical journals and in general to the essence of the mathematical activity. For instance, evaluation based in the impact scores of the journals with respect to a given indicator assumes the fundamental hypothesis of the universality of citation distributions. This idea is not always expressed in a clear way among the assumptions of the researchers and evaluators using impact tools, and may be established as follows: the distribution of the normalized citation scores of all publications in a certain field and a certain year is the same for all fields and all years. This central idea was explicitly analyzed in Radicchi et al. (2008). However, it does not seem to be clear in disciplines with a relatively low average number of citations per publication, -for example, mathematics-, as has been recently proved in Waltman et al. (2012). The universality of citation distributions is sometimes translated in the following practical assumption for the research assessment. Journal rankings reflect how relevant a given journal is in an accurate way, independently of the specific scientific area that is considered. However, if the citation distribution is not the same in a given scientific field than in the rest of the disciplines considered in an evaluation process, the value of a given journal in the corresponding non-standard list should be changed. For example it may happen that in a "normal" discipline the quartile of the list in which a given journal is provides a reasonable way for giving a mark to it -First=4, Second=3, Third=2, Fourth=1-. However, summing up all the contributions of all the published papers in an area with an anomalous citation distribution, it may happen that the relative differences of the values of the IF of journals in the first and second quartiles are very small. In this case, it would be better to use a different rule, for example: First=4, Second=4, Third=2, Fourth=1.

One way to measure the validity of citation analysis is to observe the stability of journal rankings over time (Lercher, 2013). Due to the particular properties of the mathematical journals, the purpose of this paper is: 1) to analyze the properties of the mathematical publications and the consequences in research assessment; 2) to analyze the stability rate of the rank-normalized impact factor of mathematical journals, comparing it with the stability rate obtained for other lists of pure sciences: Applied Physics and Microbiology; 3) finally, we propose an alternative index that we call "mean cumulative impact factor" that better fits the special features of Mathematics.

2. Properties of the mathematical journals

In order to explain the properties of the mathematical publications and the consequences in the research assessment, two elements must be taken into account: the intrinsic characteristics of the journals, and the use that the researchers make of them.

2.1. Journals and editors

2.2.1. A great difference among the journals that publish pure mathematics and the specialized journals in other sciences is that there are very prestigious mathematical journals that publish a little number of papers per year. In other sciences, a typical leading journal publishes hundreds of papers per year. In general, there are a lot of good publications in pure mathematics that publish just a small number of articles. This of course contributes to a big degree of instability, since small changes in the number of citations produce big changes in the value of the IF, and so in the position in the IF list (Pudovkin and Garfield, 2004t).

2.1.2. The index of obsolescence of the papers in mathematics is very long, and the articles start their influence in the mathematical research later than in other sciences (Bensman et al. 2010). A ten/twenty years old paper may be completely in order for a research that is starting now. On the other hand, the papers "start to live" later than in other sciences. In the first two years the articles may have no citations at all. Together with the long waiting list for publication of accepted papers in the great majority of mathematical journals, this may contribute to the increase of the arbitrary behaviour of the IF (IMU 2008).

2.1.3. The absolute value of the IF of the mathematical journals is very small in comparison with other sciences. A good journal with a long academic tradition may have a mean value IF=0.5. This increases the arbitrariness of its value, since a difference of one or two citations affects a lot to the value of the IF.

2.1.4. There are a lot of journals with the almost chaotic behaviour explained in the previous paragraphs in the IF list of mathematics. Together with the previous properties, this increases instability of the list.

2.2. Mathematicians and articles

Mathematicians do not believe that only recent papers can be useful for their research interest. In fact, old papers and books appear in the bibliography of almost all papers in pure mathematics. Consequently, even the 5-year Impact Factor involves a too small evaluation period (IMU, 2008). This means that in practice, researchers take into account all the mathematical literature for their research, without considering that new references are in any sense better than old ones. Moreover, sometimes an old paper of a prestigious author is preferred to a new paper with similar results. Classical books in well-established mathematical disciplines are considered as primary sources and appear in the list of references of almost every paper. Also, bibliographic material that is never considered as primary sources in other disciplines —specialized books, doctoral thesis, (even unpublished) lecture notes— are usual references in mathematical papers. These sources of information are not considered for the computation of the IF. Summing up all these aspects, it can be said that mathematics are in a sense more

similar to classical humanistic disciplines than to the scientific ones. In any case, the 2-year impact factor is considered in general as inadequate for researchers in mathematics (IMU, 2008).

2.3. Mathematicians and citations

It is well known that citation rates and counts vary greatly by field or discipline. In particular, the distribution of citations among mathematics journals differs from that for science journals.

As we said before, the research field that is usually denoted with the label "MATHEMATICS" is composed by a number of separated disciplines with a low level of interaction and with a lot of small specialized journals. There are just a limited number of high level journals covering all disciplines, as Proceedings of the American Mathematical Society, Journal of the London Mathematical Society or Houston Journal of Mathematics. Also, papers have a long cited half-life, placing most citations beyond the time limits of the measuring intervals of the usual impact factors (2- year, 5-year,...).

This results in outliers that distort the measures in many journals (Bensman, 2008; Bensman et al. 2010). The Joint Committee on Quantitative Assessment Research suggests that given the long cited half-life of mathematics, citation measures should be based on back files longer than the time limits of the measures of mean citation rate per article (IMU, 2008). Smolinsky and Lercher (2012) compared citation counts for award-winning mathematicians in different subdisciplines of mathematics. They concluded that the variation of citation rates might be due to variation of internal citing and publication practices in subdisciplines. Some subfields of mathematics may gain citations from their relationship to a higher citation field engaging in more generous citation practices. Citation counting in mathematics can pick out only certain subdisciplines and neglect others in a way that does not match any standard that mathematicians would recognize (Behrens and Luksch, 2011).

3. The instability of mathematical journals versus journals in other areas

We use the Journal Citation Reports (JCR) lists of Thomson-Reuters as a source of information to analyze the stability of the mathematical journal ranking through the years. We also compare it with other fields: applied physics and microbiology. A time period of eleven years was considered.

We have selected ten publications of each of the three areas. The selection was based on the position that journals occupied in the 2002 Science edition of JCR. We took ten journals equidistributed in each list, starting with the top journal of the list of this year. Since the number of journals listed varies each year, a normalization procedure was needed. We decided to use the relative index

$$RNIFn(j)=(Jn+1-POn(j))/Jn,$$

where POn(j) is the position of the journal j in the JCR list and Jn is the total number of journals in the list for a given year n. This index was defined and studied by Pudovkin and Garfield (2004), who called it the rank-normalized impact factor. Note that it is more convenient for evaluation purposes, since it does not depend neither on the scientific field nor in the numerical value of the IF, but in the position in the corresponding list that is often the relevant information for comparing papers and journals.

Table 1. Evolution of the rank-normalized impact factor of the journals of mathematics.

YEAR	B BELG MATH SOC-SIM	GLASGOW MATH J	MANUSCRIPTA MATH	P EDINBURGH MATH SOC	J COMB THEORY A	PUBL RES I MATH SCI	DISCRETE COMPUT GEOM	J ANAL MATH	MATH RES LETT	J AM MATH SOC
2002	0,141	0,229	0,329	0,429	0,529	0,629	0,682	0,806	0,906	1
2003	0,529	0,265	0,592	0,356	0,621	0,747	0,816	0,690	0,810	0,994
2004	0,265	0,144	0,315	0,337	0,613	0,171	0,768	0,773	0,807	0,994
2005	0,099	0,138	0,376	0,425	0,674	0,403	0,779	0,608	0,712	1
2006	0,059	0,241	0,316	0,299	0,701	0,385	0,481	0,380	0,658	1
2007	0,048	0,111	0,174	0,512	0,773	0,560	0,633	0,343	0,739	0,981
2008	0,051	0,065	0,4	0,549	0,814	0,414	0,674	0,609	0,433	0,986
2009	0,423	0,486	0,553	0,647	0,686	0,478	0,792	0,502	0,573	0,996
2010	0,133	0,237	0,462	0,645	0,627	0,480	0,760	0,828	0,702	0,996
2011	0,253	0,519	0,298	0,280	0,775	0,339	0,841	0,754	0,716	1

| 2012 | 0,118 | 0,297 | 0,223 | 0,497 | 0,743 | 0,777 | 0,615 | 0,787 | 0,557 | 1 |

Table 2. Evolution of the rank-normalized impact factor of the journals of applied physics.

YEAR	HIGH TEMP+	SOLID STATE TECHNOL	EUR PHYS J-APPL PHYS	CRYOGENICS	PHYSICA C	PHILOS MAG B	J ELECTRON MATER	J VAC SCI TECHNOL B	EUR PHYS J E	MAT SCI ENG R
2002	0,099	0,197	0,296	0,394	0,493	0,606	0,704	0,803	0,915	1
2003	0,210	0,197	0,342	0,303	0,605	0,750	0,763	0,789	0,921	1
2004	0,076	0,152	0,304	0,468	0,481	0,620	0,671	0,747	0,797	1
2005	0,060	0,108	0,193	0,325	0,410	0,651	0,626	0,699	0,879	0,988
2006	0,024	0,059	0,429	0,405	0,309	0,571	0,658	0,702	0,869	0,988
2007	0,074	0,096	0,277	0,404	0,447	0,596	0,521	0,585	0,777	0,989
2008	0,053	0,042	0,210	0,263	0,179	0,516	0,474	0,547	0,726	0,979
2009	0,046	0,0463	0,296	0,389	0,278	0,500	0,546	0,565	0,768	0,981
2010	0,186	0,076	0,297	0,424	0,525	0,466	0,534	0,449	0,729	0,983
2011	0,112	0,04	0,288	0,232	0,368	0,56	0,552	0,512	0,696	0,984
2012	0,133	0,008	0,227	0,422	0,242	0,594	0,617	0,453	0,672	0,977

Table 3. Evolution of the rank-normalized impact factor of the journals of microbiology.

YEAR	MICRO-BIOLOGY+	J GEN APPL MICROBIOL	CAN J MICROBIOL	ORAL MICROBIOL IMMUN	VET MICROBIOL	ARCH MICROBIOL	FEMS MICROBIOL ECOL	CURR TOP MICROBIOL	ADV MICROB PHYSIOL	MICROBIOL MOL BIOL R
2002	0,134	0,232	0,329	0,429	0,524	0,744	0,719	0,817	0,915	1
2003	0,119	0,167	0,286	0,369	0,429	0,714	0,75	0,762	0,952	1
2004	0,107	0,143	0,274	0,405	0,476	0,762	0,69	0,845	0,94	1

2005	0,081	0,151	0,233	0,512	0,477	0,581	0,721	0,86	0,942	1
2006	0,067	0,09	0,213	0,449	0,438	0,404	0,73	0,854	0,933	1
2007	0,075	0,14	0,247	0,387	0,419	0,484	0,742	0,86	0,871	0,978
2008	0,099	0,121	0,176	0,418	0,571	0,494	0,736	0,857	0,934	1
2009	0,084	0,137	0,105	0,495	0,621	0,453	0,737	0,81	0,884	0,958
2010	0,168	0,206	0,243	0,57	0,701	0,458	0,757	0,813	0,935	0,963
2011	0,132	0,175	0,246	0,596	0,693	0,368	0,719	0,833	0,912	0,965
2012	0,103	0,121	0,172	0,56	0,664	0,422	0,741	0,836	0,905	0,983

Tables 1 to 3 show the evolution of rank-normalized impact factor in 10 journals of mathematics, applied physics and microbiology from 2002 to 2012.

The behavior of the journals in the impact factor list of mathematics (figure 1) shows a clear difference depending on the situation in the list. It can be seen that the journals that have the biggest IF are stable regarding their positions in the list. This is the case of the top journal considered in our study. However, the rest of the journals have a rather erratic behavior. For example, all of them jump from one year to the next one the 25 per cent of the positions in the list at least one time in the last 10 years.

Figure 1. Journals of mathematics
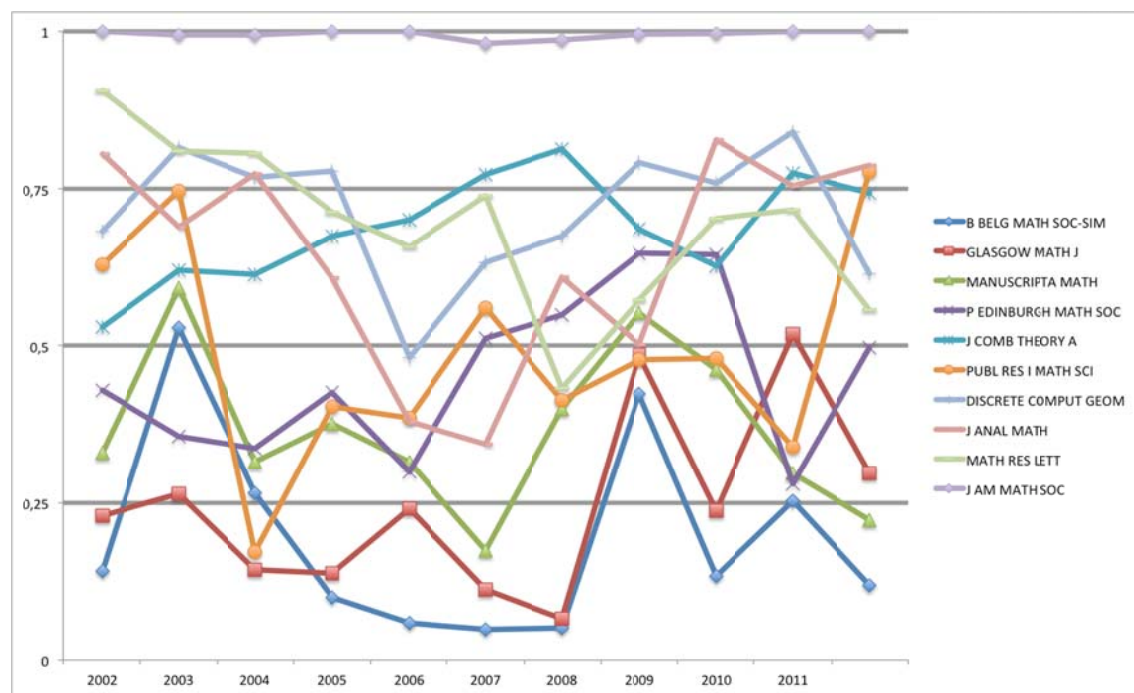
This extreme manner cannot be observed in the other two lists that we have analyzed (figures 2 and 3). It must be noted also that these jumps between consecutive years are localized in a different year depending on the journal, so it cannot be associated to any global change in the editorial policy affecting all the journals or the way Thomson-Reuters computes the IF.

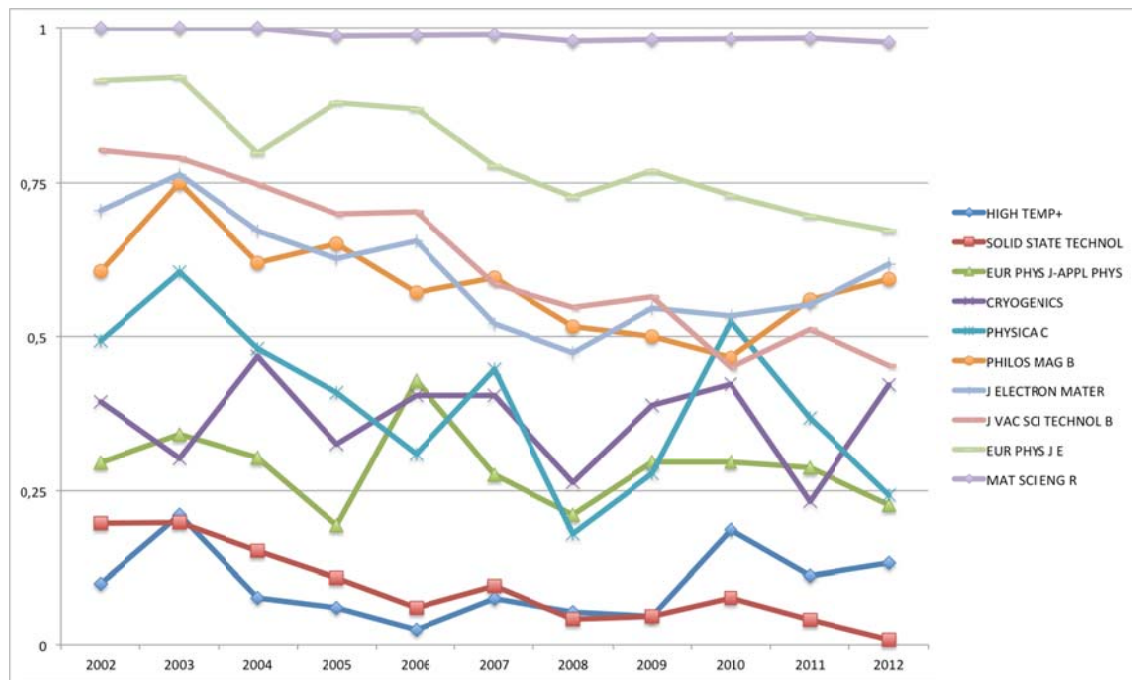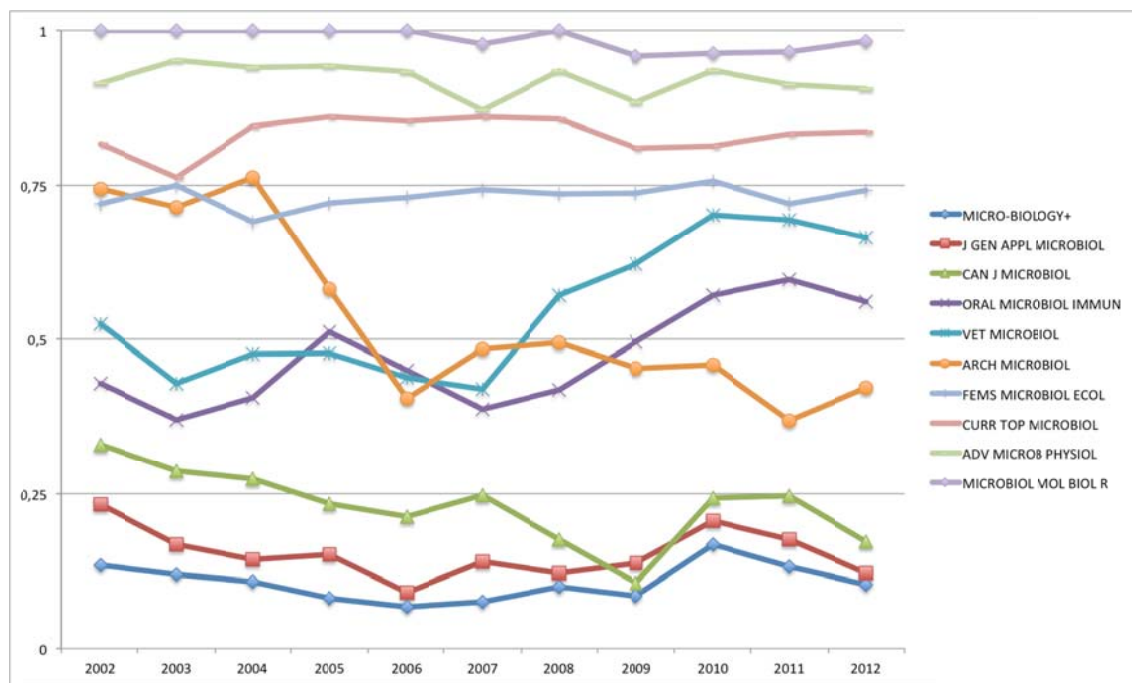Figure 2. Journals of applied physics

Figure 3. Journals of microbiology

4. Proposed alternative: mean cumulative impact factor

An alternative procedure for the measuring of the mathematical production would be given by the replacement of the rank-normalized impact factor by an index -that we call mean cumulative impact factor-, which takes into account the values of the rank-normalized impact factor in the previous years. We will denote by MCIF to our new index.

4.1. Computation of the mean cumulative impact factor.

Consider a natural number m and a fixed increasing sequence of positive weights a(1),...,a(m) such that a(1)+...+a(m)=1. Take a year n and write RNIF(n+1-m)(j),...,RNIFn(j) for the values of the rank normalized impact factor of a given journal j for the years (n+1-m) to n. We define MCIFn(j) -the index MCIF for j in the year n- as

$$MCIFn(j)=a(1)RNIF(n+1-m)(j)+ a(2)RNIF(n+2-m)(j)...+a(m)RNIFn(j).$$

This gives a new index that preserves some information of the values of RNIF of the journal in the m years previous to n, smoothing the behavior of the index. For instance, m=4 and the weights a(1)=1/10, a(2)=1/5, a(3)=3/10 and a(4)=2/5 would be adequate. We will use these values in the next computations.

In what follows we present the tables and figures that prove that the behavior of this new index is much more stable than the rank-normalized impact factor.

Table 4. Evolution of the mean cumulative impact factor of the journals of mathematics

YEAR	B BELG MATH SOC-SIM	GLASGOW MATH J	MANUSCRIPTA MATH	P EDINBURGH MATH SOC	J COMB THEORY A	PUBL RES I MATH SCI	DISCRETE COMPUT GEOM	J ANAL MATH	MATH RES LETT	J AM MATH SOC
2005	0,239	0,174	0,396	0,385	0,631	0,425	0,773	0,694	0,780	0,997
2006	0,159	0,193	0,361	0,350	0,667	0,384	0,661	0,558	0,719	0,998

2007	0,083	0,159	0,271	0,413	0,716	0,437	0,630	0,450	0,716	0,992
2008	0,057	0,121	0,313	0,476	0,765	0,451	0,634	0,483	0,598	0,989
2009	0,200	0,260	0,408	0,556	0,743	0,466	0,694	0,490	0,573	0,990
2010	0,195	0,265	0,448	0,613	0,697	0,474	0,740	0,638	0,613	0,993
2011	0,231	0,382	0,408	0,490	0,717	0,417	0,790	0,711	0,655	0,997
2012	0,192	0,371	0,326	0,477	0,724	0,556	0,730	0,757	0,635	0,999

Table 5. Evolution of the mean cumulative impact factor of the journals of applied physics

YEAR	HIGH TEMP+	SOLID STATE TECHNOL	EUR PHYS J-APPL PHYS	CRYOGENICS	PHYSICA C	PHILOS MAG B	J ELECTRON MATER	J VAC SCI TECHNOL B	EUR PHYS J E	MAT SCI ENG R
2005	0,099	0,148	0,266	0,370	0,479	0,657	0,675	0,742	0,866	0,995
2006	0,064	0,106	0,325	0,383	0,403	0,623	0,662	0,719	0,863	0,992
2007	0,056	0,093	0,309	0,395	0,402	0,602	0,598	0,659	0,827	0,990
2008	0,054	0,068	0,272	0,340	0,309	0,565	0,540	0,605	0,785	0,985
2009	0,052	0,056	0,280	0,356	0,285	0,531	0,531	0,577	0,767	0,983
2010	0,106	0,062	0,277	0,379	0,374	0,499	0,524	0,517	0,745	0,982
2011	0,115	0,052	0,285	0,324	0,378	0,515	0,538	0,507	0,723	0,983
2012	0,129	0,035	0,266	0,362	0,340	0,549	0,574	0,481	0,700	0,981

Table 6. Evolution of the mean cumulative impact factor of the journals of microbiology

YEAR	MICRO-BIOLOGY +	J GEN APPL MICROBIOL	CAN J MICROBIOL	ORAL MICROBIOL IMMUN	VET MICROBIOL	ARCH MICROBIOL	FEMS MICROBIOL ECOL	CURR TOP MICROBIOL	ADV MICROB PHYSIOL	MICROBIOL MOL BIOL R
2005	0,102	0,160	0,266	0,443	0,472	0,678	0,717	0,832	0,941	1,000
2006	0,084	0,127	0,239	0,451	0,456	0,560	0,721	0,845	0,939	1,000
2007	0,077	0,128	0,237	0,432	0,442	0,507	0,729	0,857	0,911	0,991
2008	0,084	0,124	0,210	0,424	0,489	0,482	0,735	0,858	0,916	0,993
2009	0,085	0,128	0,166	0,446	0,547	0,467	0,737	0,839	0,901	0,979

2010	0,120	0,162	0,189	0,499	0,623	0,466	0,745	0,826	0,913	0,970
2011	0,130	0,171	0,210	0,550	0,669	0,425	0,736	0,825	0,916	0,967
2012	0,123	0,156	0,202	0,566	0,676	0,416	0,737	0,828	0,911	0,971

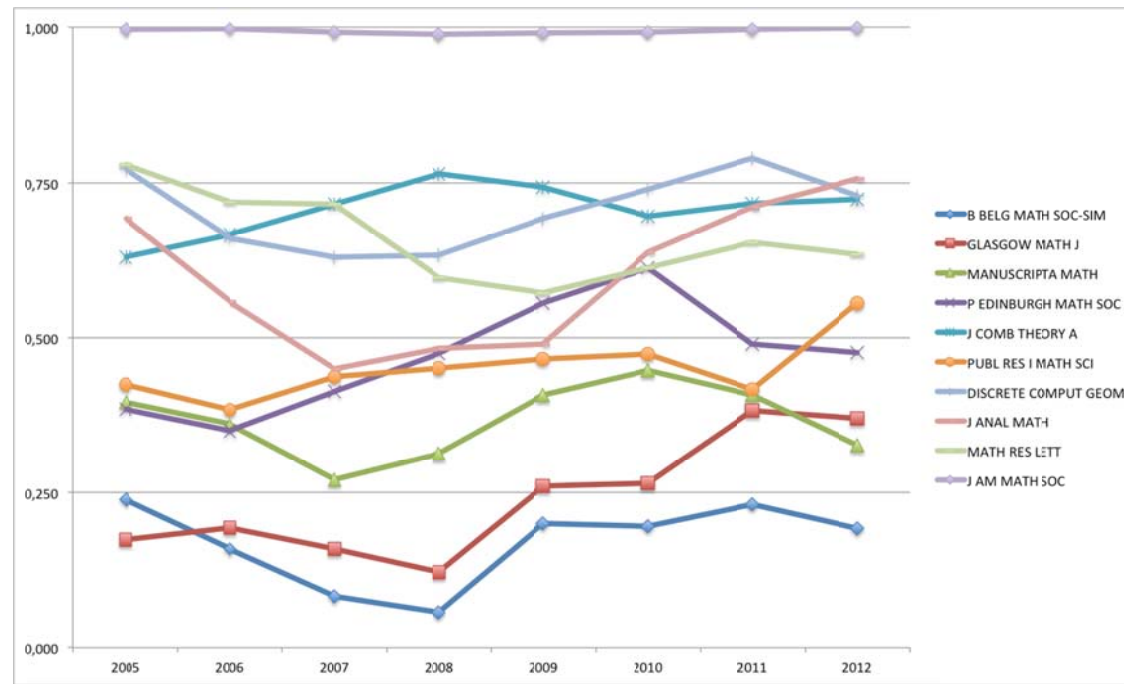Figure 4. Mean cumulative impact factor: journals of mathematics

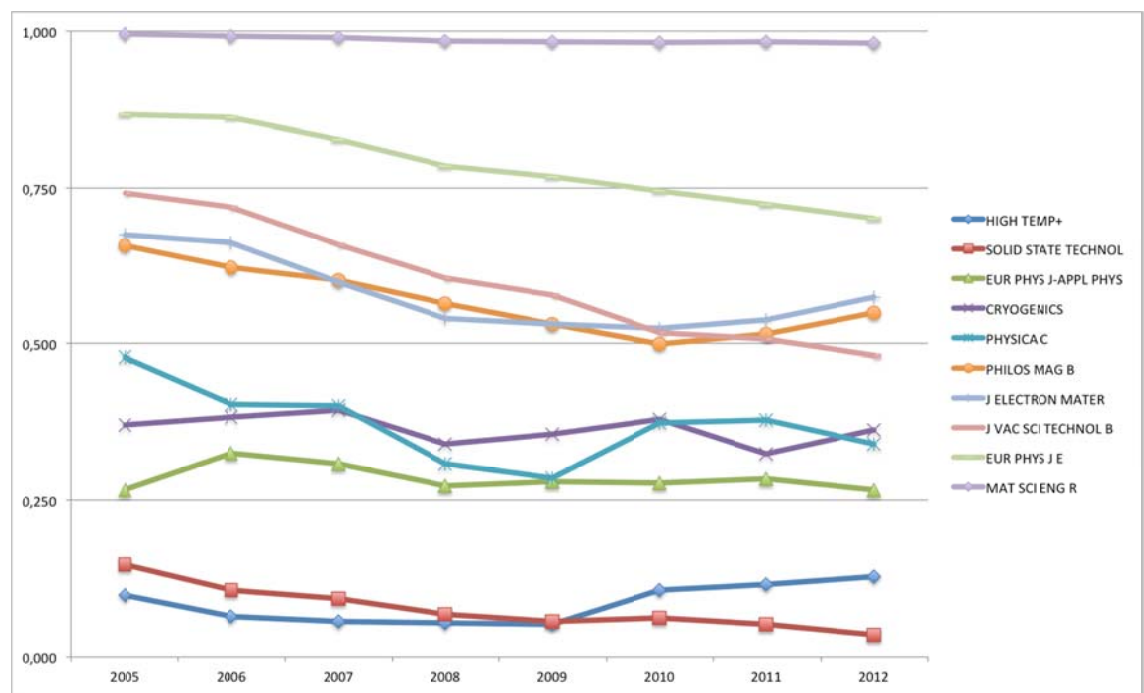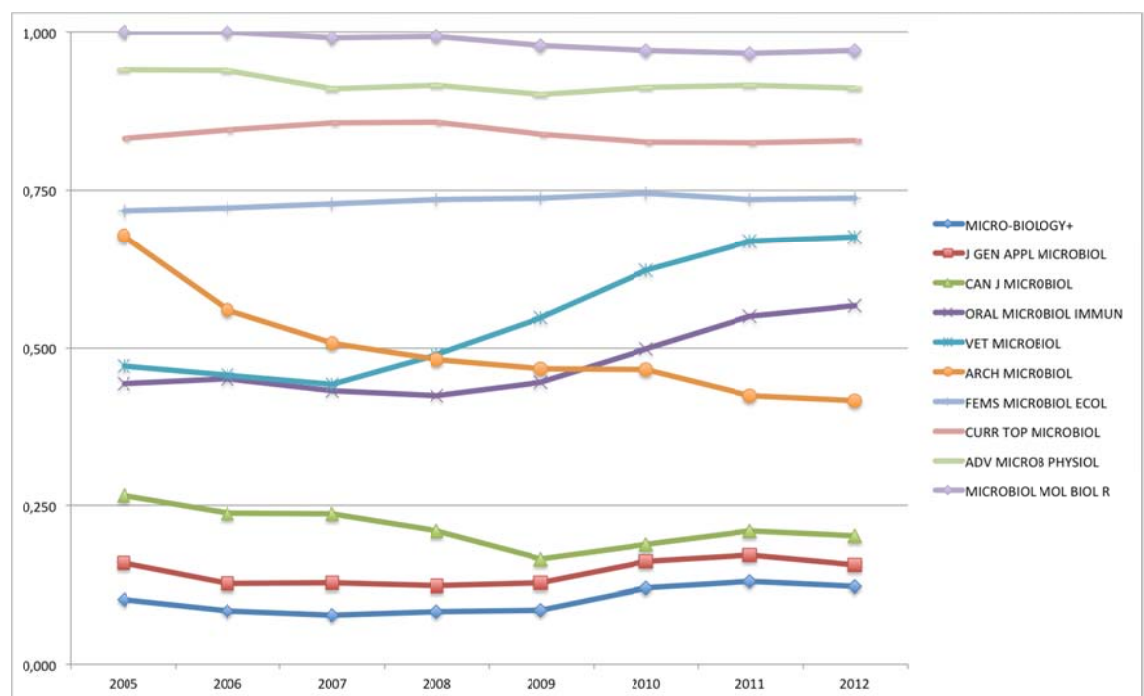Figure 5. Mean cumulative impact factor: journals of applied physics

Figure 6. Mean cumulative impact factor: journals of microbiology

The reader can notice that the mean cumulative impact factor is much more stable than the rank-normalized impact factor in the three cases that we have considered. This is not surprising, since its mathematical structure follows a standard regularization method used often in signal processing for example. The first year appearing in the time series is 2005, since the information of the previous three years is necessary for defining the index. Due to the more stable behavior of the mean cumulative impact factor the individual researcher can choose its own publication strategy in a better way even in mathematics, since it can be expected that the journal ranking will be similar in the next few years. This makes this index more convenient for research assessment.

4.2. Further improvements: a confidence interval associated to each journal.

Another tool can be implemented for improving the research assessment based on the IF. The idea is to use a confidence interval to exclude particular articles of the automatic assessment for evaluating it following a different rule. In case the RNIF of the journal in which the paper is published in a given year n is out of the "security interval" defined by the value of the RNIF of the previous year, no automatic conclusion on the paper can be given, and an ad hoc procedure must be provided. For example, such a procedure would be that the mark given to such a paper is the one that is given to papers belonging to the next better category.

The radius r of such an interval for a given paper that is published in a journal j may be given for example as the maximum of the absolute value of the difference RNIFm(j) – RNIF(m-1)(j), where m refers to any year in a given known time series of values of the RNIF (for example, from 2002 to 2011). The security interval that must be assigned to a paper published in a journal j in the year n (in the example, 2012) is then given by

[RNIF(n-1)(j)-r, RNIF(n-1)(j)+r].

That is, in the example, the automatic evaluation of the paper published in the journal j in 2012 is allowed if RNIF2012(j) belongs to the interval [RNIF2011(j)-r, RNIF2011(j)+r]. Otherwise, a different criterion must be used for the paper, for instance to give the mark of the previous quartile to which the journal belongs this year n.

5. Discussion and conclusions

The adequacy of citation counts, impact factor and other bibliometric indicators in the field of mathematics has already been treated in previous works. Korevaar and Moed (1996) were the first authors in publishing that many mathematicians are not convinced about the fact that citation counts provide useful information in the field of mathematics. The reason is that citation and publication habits in mathematics are different than the ones in other scholar fields. Moed et al. (1998) found that the 2-years impact factor is more strongly biased toward rapidly maturing journals than it is toward rapidly declining ones, noting that the older the documents, the less relevant the statistic is as a measure of the current impact of a journal. In the same line, Bensman et al. (2010) found that a time span of 5 years instead of 2 for the impact factor makes it more suitable for mathematics titles, which tend to have more cited half-life. This is a well-known property of the dynamics of citations in mathematics: Bouche et al. (2012) have proved that time lag in mathematical references is meaningfully bigger than in other sciences, and it even tends to increase in recent years. Meanwhile, the Joint Committee on Quantitative Assessment of Research of the International Mathematical Union, International Council of Industrial and Applied Mathematics, and the Institute of Mathematical Statistics pointed out that in disciplines like mathematics, where many citations occur outside the 2-year window, journal impact factor will be far lower (IMU, 2008). The Committee also noted that the impact factor can vary considerably from year to year and the variation tends to be larger for the smaller journals due to their small simple size (Bensman et al. (2010)). However, although a complete study has not been done, the authors have observed that the use of the 5-year impact factor does not improve the stability of the corresponding rank-normalized impact factor. The reason is that instability is mainly due to the small relative values of the original impact factor, that make the rank-normalized impact factor highly sensitive to small variations.

One impediment that restricts the applicability of citation counts and impact factor in the evaluation of scientific activity is its variability according to scientific fields, especially the mean number of references and the article age in the field. Variations of mean citations per article by discipline and by subdiscipline have been measured by Radicchi et al. (2008) and Bornmann and Daniel (2006) in a set of disciplines. They found differences between disciplines, with noticeable fluctuations in the average number of citations (from 25.26 in Microbiology to 5.97 in Mathematics) and in the cited half-life indicator. In other study, Waltman et al. (2012) conclude that claiming citation distributions to be universal for all fields of science is not warranted, especially in fields with a relatively small number of citations per publication, as can be found in the engineering sciences, material sciences, and social sciences. The results of the present paper suggest that the indiscriminate use of impact based ordered lists for the aim of scientific assessment may be not adequate for disciplines with anomalous citation behavior, that is, for example for fields that do not fit with the assumption of the existence of a universal citation distribution.

The fact that the IF defines a non-stable ordering in the case of mathematics can be explained by several reasons. Although it is a verifiable fact that the ratio of citations of a given paper of mathematics in the first two years after publication with respect to the total amount of final citations is small, it does not seem to be the main reason for the failing in the stability of the ordering through the time. Fix a year and consider the list of all the journals in the category of mathematics for this year. Recent investigations have shown that the correlation coefficient among the sequence of the number of citations per journal in this list in n years and the same sequence of number of citations in n+1 years tends to 1 very fast. In fact, it has been proved by van Leeuwen (2012), that except for the year-1/year-2 comparison, the correlation coefficient is almost one after the year-2/year-3 comparison. Although this tendency is weaker when the class of the journals containing the lowest quantity of publications per year is considered, the general behavior is that the number of citations to the journals of a given list increase uniformly with the number of years considered. The ordering provided by the number of citations among the journals in the list is so almost preserved after three years.

Another possible reason for the unstable behavior of the rank normalized impact factor through the years in the case of mathematics is the big amount of journals with a small number of papers published. This has also been analyzed in the paper by van Leeuwen (2012), who proves that in the particular case of mathematics there is a meaningful number of journals with a small number of papers published in comparison to other disciplines. However, this is not an exclusive feature of the journals of mathematics, and so, this does not justify by itself the anomalous behavior of the ordering induced by the IF in the list.

Unstable ordering may be also caused by the wrong definition of the sets of documents that must be considered in the computation of the 2-years IF. As was studied in van Leeuwen and Moed (2002), this may produce a big distortion of the expected tendency of the impact-based ordering, since there are documents that must be taken into account in the numerator and the denominator defining the IF that are not considered. In our opinion, the main reason can be of statistical nature, due to the small values of the 2-year IF in the list of mathematics that is produced by the small rate of citations in this discipline. For example, in the Thomson-Reuters JCR list of 2013 for this subject, there are 43 journals -of a list of 302 journals- with a score among 0.5 and 0.6. This may produce the big changes in the induced ordering that we have observed.

Another factor that compromises the utility of citation counts and impact factor is due to manipulations. Arnold and Fowler (2011) evidenced three types of impact factor manipulations: (1) abuse of self citations (citing themselves); (2) suggestions or requirements by editors to authors with a paper under review to cite other papers from the journal, a practice that some author consider border on the extortion; (3) and publishing review items with large numbers of citations to the journal.

They demonstrate impact factor distortions in three journals in the field of applied mathematics using these manipulations. They calculated concentrations of 71.5% of the citations fell within the two-year window, and a subsequent decrease of 6.64 points in the impact factor of a specific journal using a 5-years window instead of the usual two-preceding-years window. They also found 29% account of citations from members of the editorial board and high percentages of citations from proceedings of conferences issues edited by the editorial board of the same journal.

In conclusion, although the introduction of objective criteria for research assessment of scientists and research teams have had a lot of positive consequences in the scientific policy in some countries in Europe, in the particular case of mathematics the usual 2-year impact factor shows to be inadequate. This fact is well known for mathematicians, which observe how a wrong criterion is used for their evaluation year after year. However, the proved objectivity of the evaluation based on the IF provides a feeling of accuracy of the obtained result. Is an extended opinion among researchers in pure mathematics that this procedure has produced some benefits for the general evaluation of the scientific production, including mathematics, although the results obtained are close to be in a sense arbitrary. However, some general elements of the system are reliable, as the fact that a journal appears in the list as a proof of a certain quality level. On the other hand, it is clear that tools coming from the information science must be at the basis of the scientific assessment. In fact, there are available indexes that can be used in substitution of the IF, as the Eigenfactor, h-index, Article Influence score, SNIP or SJR, but any of them would give only a partial answer to the question about the relevance of a journal. Also, the use of the IF is so widespread that it would be difficult to change. Consequently, some modification of the IF in order to make reliable the conclusions of the evaluation are necessary. For instance, the proposed mean cumulative impact factor clearly shows a more stable behavior and could help researchers to perform a better publication strategy, if they know that this index is going to be used for the research assessment by the funding institutions. Another purpose that could improve the research assessment is to include a security interval in the classification of the journals computed taking into account the year-to-year maximal variations.

In any case, the IF as a unique tool does not seem to be a satisfactory solution for the problem of the research assessment; the use of complementary indicators in the evaluations measuring different aspects of the scientific production would reduce the degree of arbitrariness of the process, improving in this way its final results.

References

Adam, D. (2002). "Citation analysis: The counting house". *Nature*, 415: 726-729.

Archambault, E., Larivière, V. (2009). "History of the journal impact factor: Contingencies and consequences". *Scientometrics,* 79: 635-649.

Arnold D.N., Fowler K.K. (2011). "Nefarious Numbers". *Notices of the AMS,* 58: 434-437.

Behrens H, Luksch P. (2011) "Mathematics 1868–2008: a bibliometric analysis". *Scientometrics,* 86:179–194.

Bensman, S. J. (2008). "Distributional differences of the impact factor in the sciences versus the social sciences: An analysis of the probabilistic structure of the 2005 journal citation reports". *Journal of the American Society for Information science and technology*, 59: 1366-1382.

Bensman, S. J., Smolinsky, L. J., Pudovkin, A. I. (2010). "Mean citation rate per article in mathematics journals: Differences from the scientific model". *Journal of the American Society for Information science and technology*, 61:1440–1463.

Bordons, M., Fernández, M. T. (2002). "Gómez, I. Advantages and limitations in the use of impact factor measures for the assessment of research performance". *Scientometrics*, 53: 195-206.

Bornmann, L., Daniel, H. D. (2006). "What do citation counts measure? A review of studies on citing behavior". *Journal of Documentation*, 64: 45–80.

Bouche, Th., Teschke, O., Wojciechowski, K. (2012). "Time lag in mathematical references", *Newsletter of the European Mathematical Society*, 86: 54-55.

Garfield, E. (2006). "The History and Meaning of the Journal Impact Factor". *Journal of the American Medical Association*, 293: 90-93.

IMU-Joint Committee on Quantitative Assessment of Research. (2008). Citation statistics: A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Berlin, Germany: International Mathematical Union.

Korevaar J.C.; Moed, H.F. (1996). "Validation of bibliometric indicators in the field of mathematics". *Scientometrics*, 37: 117-130.

Lercher A. (2013). "Correlation over time for citations to Mathematics articles". *Journal of the American Society for Information science and technology*, 64: 455-463,.

Moed, H.F., Van Leewen, T.N., Reeduk, J. (1998). "A new classification system to describe the ageing of scientific journals and their impact factors". *Journal of Documentation*, 54: 387-419.

Pudovkin,A.I., Garfield, E. (2004). "Rank-normalized impact factor: A way to compare journal performance across subject categories". *Proceedings of the 67th Annual Meeting of the American Society for Information science and technology*, 41: 507-515.

Radicchi, F; Fortunato, S.; Castellano, C. (2008). "Universality of citation distributions: toward an objective measure of scientific impact". *PNAS*, 105: 17268-17272.

Shapiro, F.R. (1999). "Origins of bibliometrics, citation indexing and citation analysis. The

neglected legal literature". *Journal of the American Society for Information science and technology*, 43: 337-339.

Smolinsky L, Lercher A. (2012). "Citation rates in mathematics: a study of variation by subdiscipline". *Scientometrics* 91: 911-924

van Leeuwen, T. (2012). "Discussing some basic critique on Journal Impact Factors: revision of earlier comment". *Scientometrics*, 92: 443-455.

van Leeuwen T.N., Moed, H.F. (2002). "Development and application of journal impact measures in the Dutch science system". *Scientometrics*, 53: 249-266.

Waltman L, van Eck NJ, van Raan AFJ. (2012) "Universality of citation distributions revisited". *Journal of the American Society for Information science and technology*, 63: 72-77.