DOI:10.16262/j.cnki.1005-0841.2005.01.029

Quantitative Science Policy and Management by Using Scientometrics and Scientometric Indicators

Tibor Braun[1,2 2]

(1 Information Science and Scientometric Research Unit (ISSRU),
2 Institute of Inorganic and Analytical Chemistry, Eötvös University,
Budapest PO Box 123, 1443 Budapest, Hungary)

Introduction

There arc many proofs indicating that economic growth in the modern era has been grounded on the exploitation of scientific knowledge. The sphere of human activities, which can be identified as "The Republic of Science" has grown too important for the rest of society to leave alone. Most of the industrial nations and many among the LDC's acknowledge this today, and virtually all societies in which modern science is practiced pay at least lip service to the belief that it is important to pursue some form of science and technology policy. Many papers are dealing deeply with qualitative features of the abovementioned issue. Although most of the readily observed features are familiar to managers and decision makers, it is symptomatic of the relatively underdeveloped status of science and technology policy that many of its implications remain unexplored and untested against systematic models based on quantitative data. Furthermore, as is generally the case when new theoretical perspectives are gained, new questions and puzzles arise. The agenda for future research in this field, therefore, remains both extensive and varied. Nevertheless, even in their present nascent state, quantitative science studies, including scientometrics, can offer some measure of guidance for science and technology policies and management.

In the present paper I will try to present some examples in this respect.

For starting, I will use a very simple, even primitive, input-output model of the working mechanism of science R and D (Fig. 1). As visible, the mechanism works in time from left to right with some input "ingredients" to be fed into the science R and D "black box" for having some outputs. Here I will deal only with the R in the black box and will concentrate on "knowledge" as an output component. It is well known that knowledge has a real value only as so called "recorded knowledge", because when not recorded, knowledge disappears together with its holder when the holder dies. It is also clearly visible in Fig. 1 that the components of recorded knowledge taken all together represent in fact the formal literature of science.

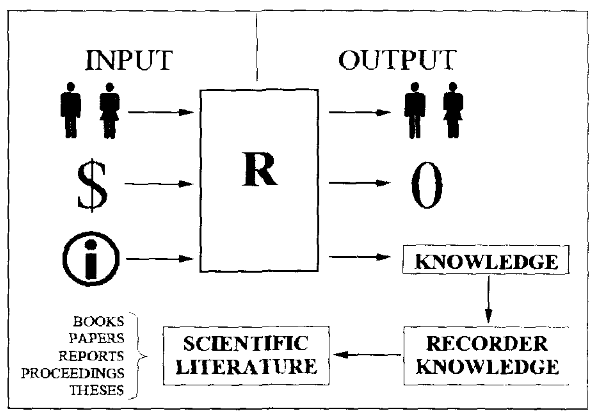


Fig. 1 Flow chart of the model of working mechanism of basic research (simplified version).

The literature of science is the body of knowledge on a subject, the prime means of communication in any subject, the only genuine representation and record of the knowledge, activities and scientific achievement in the subject .

It is an open question on whether, as seen in Fig. 1, the recorded knowledge and its components, i.e., "the literature of science" can be considered as a true mirror of the activities of R in the science R and D black box?

Our postulate is that although not being a Belgian mirror, the literature of science can be considered a fair output reflection of worldwide R activities and its careful statistical processing by scientometric methods can provide meaningful approaches to science policy and decision making.[1-3]

Scientometrics[4]

Roughly we can distinguish two different major orientations of the scientometric field. One which is theoretical and the other a pragmatic one, as shown in Fig. 2. The first approach is self-explaining, the second one is shortly discussed in what follows.

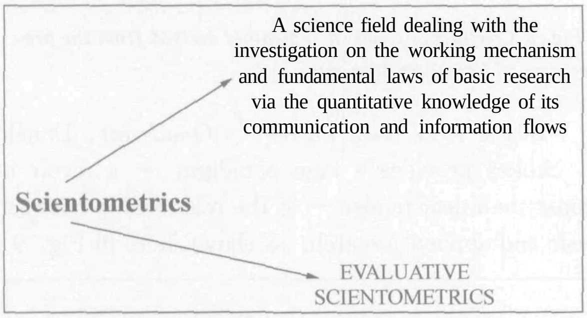


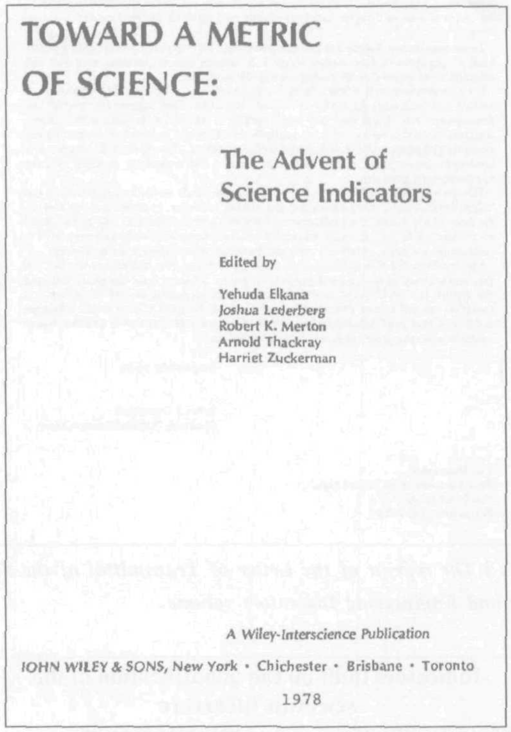
Fig. 2 ***Scientometrics as a research field and evaluative scientometrics.***

Evaluative scientometrics[4,5]

The systematic use of Scientometric Indicators, the basic pilars of evaluative scientometrics, was implemented in the US in the seventies (see Figs 3 and 4). In its Preface, the *Science and Engineering Indicators* volume of the National Science Board of the United States publishes a "Letter of Transmittal" written by the President of the Board, which mentions: "*I and my colleagues on the National Science Board trust that this report will be of value to your Administration, to the Congress, and to those concerned with science and technology policy*". The addressee of the letter is the President of the United States (Fig. 5). This is a quite convincing proof of the importance of scientometric indicators.

Figure 6 presents a schematic view on the various indicators which can be built on the quantification of the scientific literature. The use of some of them will be demonstrated in what follows in this paper.

Another very important topic is the number of the items scientometric evaluations can be deal with. Figure 7 presents a scheme in this respect. As scientometric indicators are based on the statistical analysis of the different populations of journal papers and/or citations, it is a self-explaining fact that the confidence level of the evaluation depends on the size of the evaluated data, as seen in Fig. 8.


TOWARD A METRIC OF SCIENCE:

The Advent of Science Indicators

Edited by

Yehuda Elkana
Joshua Lederberg
Robert K. Merton
Arnold Thackray
Harriet Zuckerman

A Wiley-Interscience Publication

JOHN WILEY & SONS, New York • Chichester • Brisbane • Toronto

1978

Fig. 3 The beginnings of the Science Indicators movement in the US.

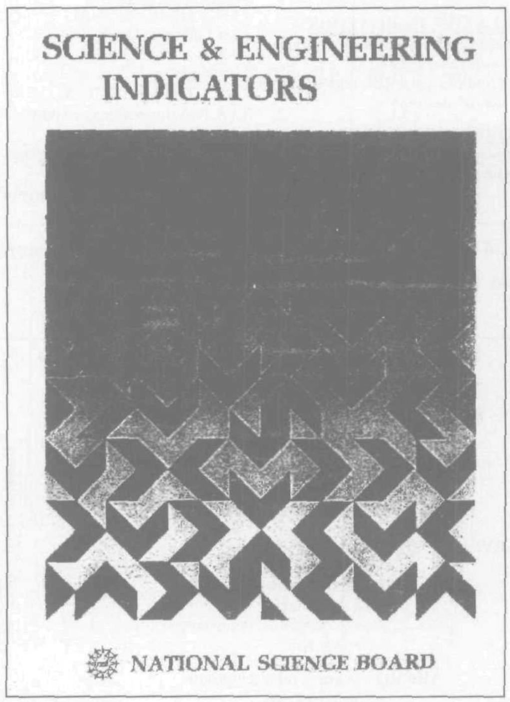
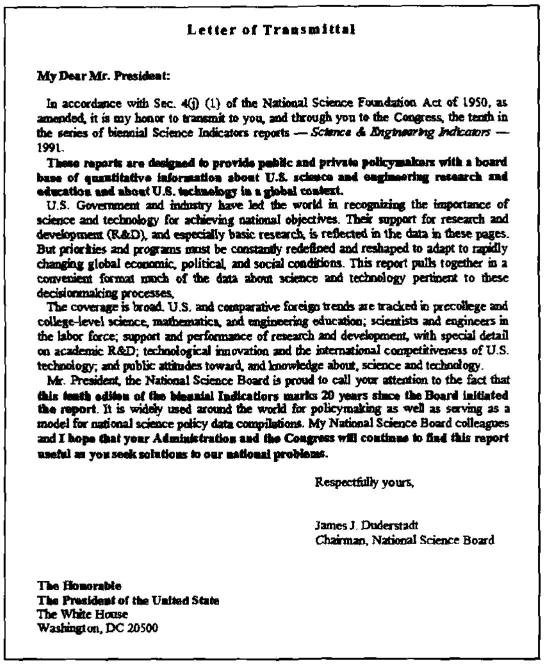


Fig. 4 ***Cover page of the bi-annually published Science and Engineering Indicators volume of the US National Science Board.***



Letter of Transmittal

My Dear Mr. President:

In accordance with Sec. 4(j) (1) of the National Science Foundation Act of 1950, as amended, it is my honor to transmit to you, and through you to the Congress, the tenth in the series of biennial Science Indicators reports — *Science & Engineering Indicators* — 1991.

These reports are designed to provide public and private policymakers with a board base of quantitative information about U.S. science and engineering research and education and about U.S. technology in a global context.

U.S. Government and industry have led the world in recognizing the importance of science and technology for achieving national objectives. Their support for research and development (R&D), and especially basic research, is reflected in the data in these pages. But priorities and programs must be constantly redefined and reshaped to adapt to rapidly changing global economic, political, and social conditions. This report pulls together in a convenient format much of the data about science and technology pertinent to these decisionmaking processes,

The coverage is broad. U.S. and comparative foreign trends are tracked in precollege and college-level science, mathematics, and engineering education; scientists and engineers in the labor force; support and performance of research and development, with special detail on academic R&D; technological innovation and the international competitiveness of U.S. technology; and public attitudes toward, and knowledge about, science and technology.

Mr. President, the National Science Board is proud to call your attention to the fact that **this tenth editon of the biennial Indicatiors marks 20 years since the Board initiated the report**. It is widely used around the world for policymaking as well as serving as a model for national science policy data compilations. My National Science Board colleagues **and I hope that your Administration and the Congress will continue to find this report useful as you seek solutions to our national problems.**

Respectfully yours,

James J. Duderstadt
Chairman, National Science Board

The Honorable
The President of the United State
The White House
Washington, DC 20500

Fig. 5 The reprint of the Letter of Transmittal of the Science and Engineering Indicators volume.

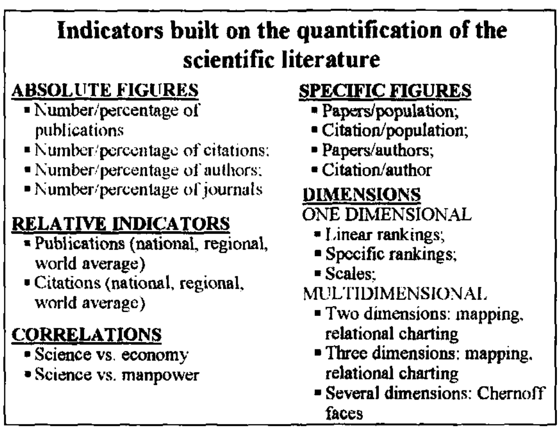


Fig. 6 Some indicators which can be built on the quantification of the scientific literature.

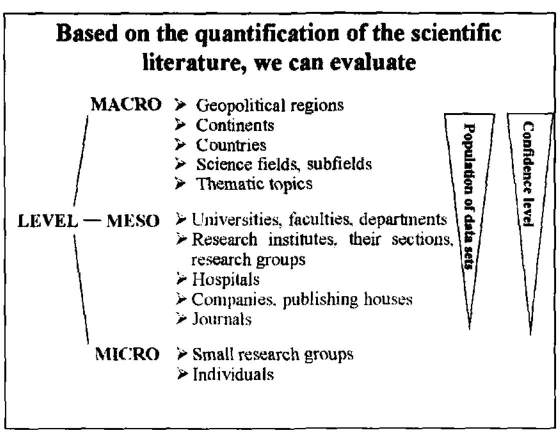


Fig. 7 Levels of quantification and objects which can be evaluated by the quantification of the scientific literature.

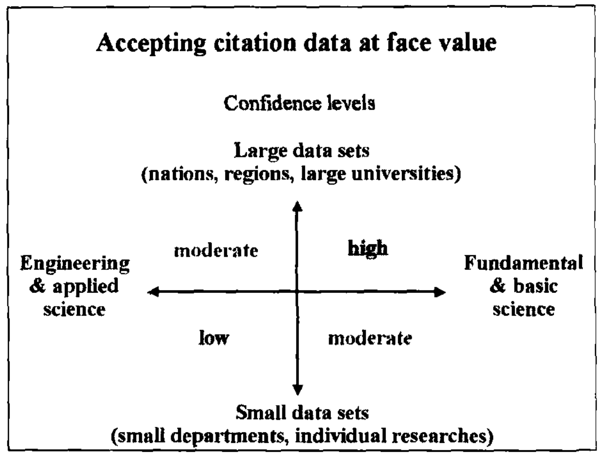


Fig. 8 Confidence levels of indicators derived from the processing of literature data sets.

In the 1997 book *Pasteur' s Quadrant*, Donald E. Stokes provides a new paradigm — a revolt against the linear model — of the relationship between basic and applied research, as shown here in Fig. 9.

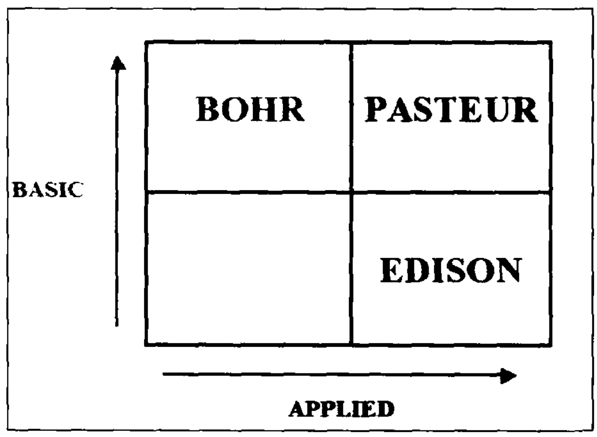


Fig. 9 The Pasteur quadrant.

Stokes contends that a large part of the important research being done today lies in the Pasteur quadrant; that is, it is driven, and justified by, both a desire for knowledge per se and the intention of serving a predetermined, practical end use. The diagram in Fig. 9 nicely shows that the Pasteur quadrant is contiguous with both the Bohr quadrant (essentially pure basic research) and the Edison quadrant (essentially applied or engineering research).

The general topic of this paper spins mainly around the Bohr quadrant, with some overlapping with the Pasteur quadrant.

When used in a correct and well thought way, scientometric indicators have to be implemented hand in hand with "classical" peer review, basic research is already making use for centuries. The aims of evaluations are, of course, to be objective and relevant. Their characteristics in evaluation are revealed in Fig. 10.

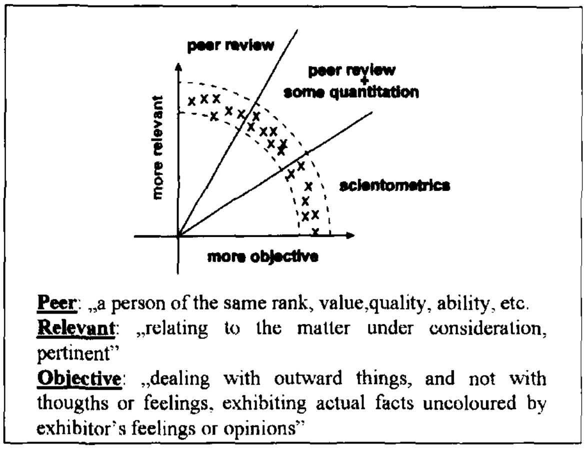


Fig. 10 Objectivity and relevance of evaluation. x: *procedures* .

The methodologies dealt with in this paper are based on ISSRU's Scientometric Indicator Datafiles[5] which are an example of carefully cleaned and reprocessed version derived from the Institute of Scientific Information (ISI, Philadelphia) Science Citation Index (SCI) database, which by its basic function, is a literature retrieval tool.

Quantitative science policy

An important factor refers to the whole sequence of the scientometric evaluation process. A simple model of the sequences as visible in Fig. 11, reveals an analogy between decision making in medical diagnosis and science policy.

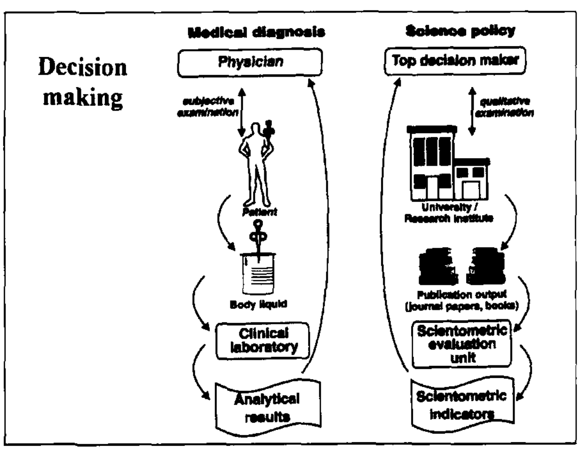


Fig. 11 A model on the analogy of medical diagnosis and scientometrics-based science policy.

In the model, science is compared to a patient in medical examination. The patient, as seen in the figure, is first exposed to a subjective examination by the physician. Paralelly, a body liquid is extracted and sent to the clinical laboratory. The outputs are clinical analytical results which go back to the physician. The diagnosis and therapy are a result of a combination of conclusions from the subjective exam and the results of analytical tests.

In science policy, the decision maker first examines qualitatively the object concerned (a country, an institution, etc.). This is usually done by peer review, expert assessment, etc. Parallely, the publication output of the same object is examined by a scientometric evaluation unit. Scientometric indicators are then sent back to the top decision maker for a combination of qualitative (e. g. , peer review), and quantitative (scientometric) data (as illustrated in Fig. 10).

Some selected scientometric indicators and their use in revealing the worldwide position of Chinese science

Liniar rankings

China's position in the ranking of journal papers productivity of the top countries during the 1990—1998 period is shown in Table 1. The same type of ranking is presented in Table 2 for citations. As visible, the ranking is in general size dependent with the big, developed countries on the top of the list.

Table 1 China's position among the most productive countries. Total number of science papers, 1990—1998

Rank	Country	No. of papers
1	USA	1763421
2	UK	468660
3	Japan	454302
4	Germany	394409
5	France	307733
⋮		⋮
13	India	95885
14	Switzerland	87901
15	P. R. China	80486
16	Israel	59930
17	Belgium	59269

Table 2 China's position among the most productive countries. Total number of citations, 1990—1998

Rank	Country	No. of citations
1	USA	9744445
2	UK	2043618
3	Germany	1719469
4	Japan	1552932
5	France	1246160
⋮		⋮
18	Austria	154051
19	India	140870
20	P.R. China	132140
21	Poland	116297
22	Norway	106047

A somewhat different situation is visible in Tables 3 and 4, where the number of journal papers and citations are appearing specified with the population of the countries in question. This seems to fortify the old saying according to which "small is beautiful".

Table 3 China's position among the most productive countries. Number of papers per million of population, 1990—1998

Rank	Country	Papers per 10^6 population
1	Switzerland	75773.9
2	Sweden	50654
3	Denmark	43632.4
4	Israel	43057.7
5	Netherlands	39822.7
⋮		⋮
47	Thailand	215.5
48	India	151.1
49	P.R. China	109.7
50	Nigeria	39.9

Table 4 China's position among the most productive countries. Number of citations per million of population, 1990—1998

Rank	Country	Citations per 10^6 population
1	Switzerland	12265.6
2	Israel	11296.6
3	Sweden	11209.4
4	Denmark	9537.7
5	Finland	8205.6
⋮		⋮
48	India	102.8
49	Thailand	83
50	P.R. China	66.8

As our main aim in this paper is to give a general picture on the topic of the paper's title, I am using China's situation only as an illustration, and I am leaving the reader to decide whether China's position in world science satisfies his/her expectations.

Time series

Quite frequently it is useful to have a view on the dynamics (i.e., the trend) of the variations in time of a certain indicator. The avowed purpose of time series is to predict what is possible in the coming year (s). Unfortunately, no one ever really foresees the beginning of a slump.

Just for illustration, in Fig. 12 we have illustrated the trend of publication activity of Chinese science as compared to the same data for India and Japan.

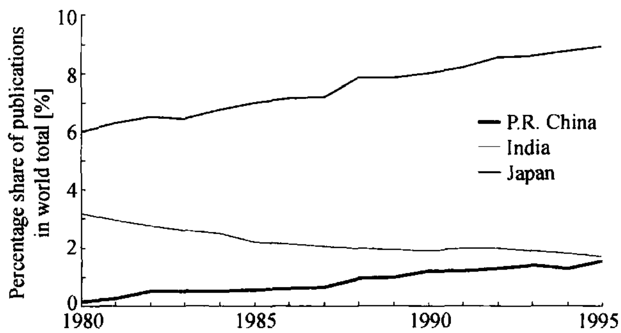


Fig. 12 trend analysis of publication productivity of Chinese, Indian and Japanese research.

Field distribution of national publication output

Pie charting and radar plots

For science policy decisions it is useful to see how much of national publication output is dedicated to each of the main science fields, e.g., to the life sciences, chemistry, physics, engineering sciences, and mathematics.

The simplest way to represent such data is the classical pie chart approach as seen in Fig. 13, for some selected countries. As seen in the figure, the distribution varies considerably from country to country, and differences can be easily distinguished.

The Relative Specialisation Index (RSI) has been conceived for revealing the concentration and neglection tendencies in the considered countries towards eight major science fields. Thus the publication profile of national research in the Asian region will be

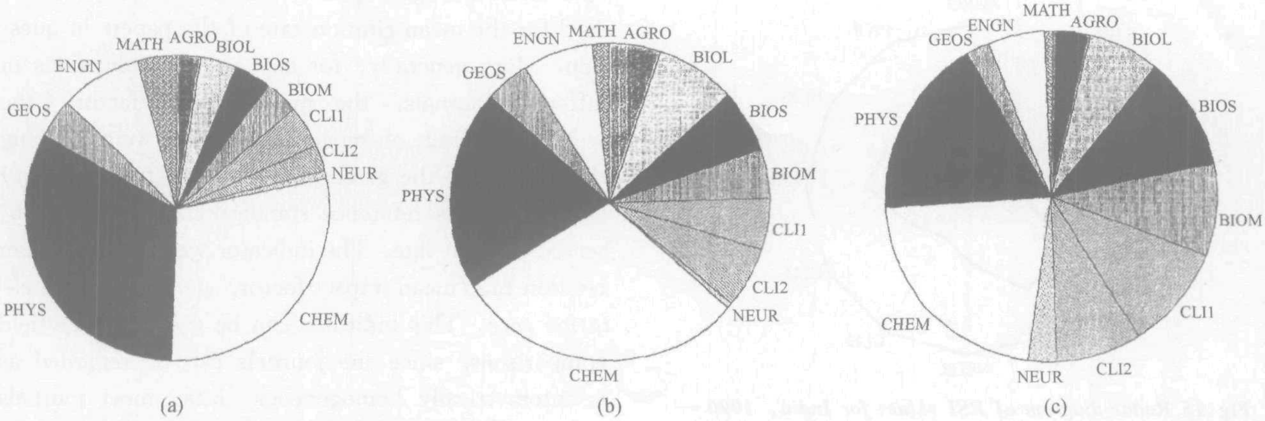


Fig. 13 *Pye chart for the internal distribution by science fields of the publication productivity for China* (*a*), *India* (*b*), *Japan* (c), *1990—1998.*

expressed by this measure which indicates whether a country has a relatively higher or lower share in world publications in a particular field of science than its overall share in world total publications. RSI has been defined as being and is closely related to the Activity Index (AI).

It is important to note that RSI reflects an internal balance among the fields in the given country, that is, positive RSI values must always be balanced by negative ones: RSI values of a country can never be all positive (or negative).

The RSI is obtained from the Activity Index (AI) by an elementary transformation, and can be defined as follows:

$$AI = \frac{\text{the world share of the given country in publications in the given field}}{\text{the overall world share of the given country in publications}}$$

or, equivalently,

$$AI = \frac{\text{the share of the given field in publications of the given country}}{\text{the share of the given field in the world total of publications}}$$

The Relative Specialisation Index is then defined as:

$$RSI=(AI-1)/(AI+1).$$

From its definition follows that RSI can take values in the range [-1,1]. RSI = - 1 indicates a completely idle research field, RSI = 1 if the country is active in no other than the given field. RSI<0 indicates a lower-than-average, RSI>0 a higher-than-average activity; RSI = 0 reflects a completely balanced "average" situation. RSI = 0 for all fields corresponds to the "world standard".

For the analysis, the following eight science fields have been used: Clinical medicine (MED), Biomedical research (BRE), Biology (BIO), Chemistry (CHE), Physics (PHY), Mathematics (MAT), Engineering (ENG) and Earth and sapece sciences (ESS).

In Figs 14—16 RSI values (world averages) are presented as the vertexes of a regular octagon. Values below or above the vertex denote the neglection or concentration of science output effort of certain science fields in the countries in question.

As visible, the tendencies which can be discerned in Figs 14—16 are quite diverse.

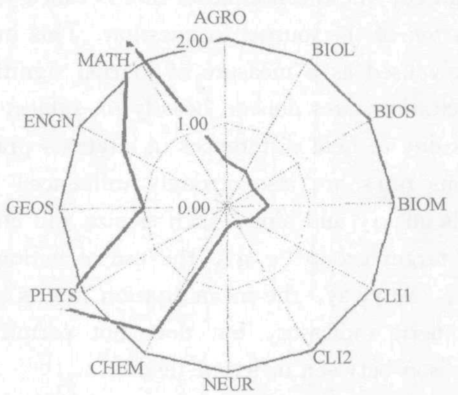


Fig. 14 Radar diagram of RSI values for China, 1990—1998.

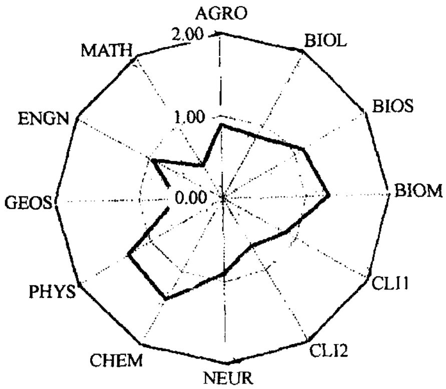


Fig. 15 Radar diagram of RSI values for India, 1990—1998.

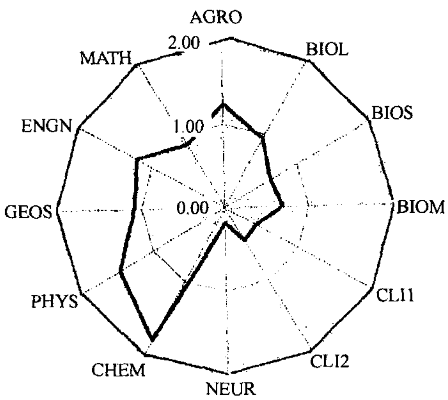


Fig. 16 Radar diagram of RSI values for Japan, 1990—1998.

Relative scientometric indicators[8]

A classical indicator for measuring the impact of a set of publications is their mean citation rate, i.e., the citations/publication ratio. If, particularly, the set represents the two years publication output of a science journal and citations in the subsequent year are counted, the mean citation rate is called the *impact factor* of the journal in question. This indicator is widely used as a measure of journal significance. Mean citation rates depend heavily on subject fields. This is due to field differences in citations practices. Citations rates are also strongly influenced by the journals quality and aims, such as size and character of the target group (e. g., the use of national languages). Anyway, the mean citation rate is a useful scientomeric indicator, but does not permit direct comparison between different fields.

Let us consider a set of papers published in one and the same journal in a given time period. The impact factor of the journal, i.e., the *expected citation rate* of an average paper, seems to be a proper standard for the mean citation rate of the papers in question. More generally, for any set of publications in different journals, the mean impact factor (the weighted average of impact factors, the weights being the number of the given journal's papers in the set) can be used as reference standard for the mean observed citation rate. The indicator we define as mean citation rate/mean impact factor, we call *relative citation rate*. This indicator can be used in cross-field comparisons, since the journals can be regarded as scientometrically homogeneous units: most journals represent a specialized subject field, a certain quality level and are addressed to a particular target group. The relative citation rate relates *observed* to *expected* citation rate. If the relative citation rate is approximately 1.0, the citation rate of the publications in question coincides with the expected one, if they received more or less citations than expected, the relative citation rate is greater or lower 1.0, respectively. *Relative citation rate* actually measures the citation impact of a given set of publications (e.g., the publication output of countries or research institutions) as related to the respective world average.

For being more explicit, it is worth mentioning that the *Mean Expected Citation Rate* (MECR) is the average expected citation rate per publication, i. e., (expected number of citations)/(number of publications), where the expected number of citations is calculated on the basis of the average citation rates of the publishing journal, i.e., each paper is expected to receive the citation rate of an average paper of the same age in the same journal.

The *Mean Observed Citation Rate* (MOCR) represents the average citation rate per publication, i.e., (number of citations)/(number of publications).

Accordingly, the Relative Citation Rate (RCR) is the MECR/MOCR value.

Relational charts and zoning

Fig. 17 is the prototype of a *"relational chart"* (RC) as defined e. g., in References 6 and 7. The *Mean Expected Citation Rate* (MECR) indicator on the horizontal axis characterizes the average citation impact of the journals, in which the researchers of a given country publish their papers. The higher this value, the higher the impact of the journals they used for publication. (Citation impact of journals is calculated in this study as mentioned above: citation received in the year of publication plus in the subsequent two years were considered.) *Mean Observed Citation Rate* (MOCR) on the vertical axis characterizes the actual citation rate per publication (counted according to the same standards as the expected citation rates) for papers of researchers from the countries under study. The diagonal MOCR — MECR has particular significance: it divides objects (countries, in our case) with citation rates above expectation from those below expectation. The *Relative Citation Rate,* RCR = MOCR/MECR indicator numerically characterizes the measure of deviance from expectation. The RCR has a value of 1.00, if the mean citation rate of the paper published by researchers of a given country exactly equals to its expected value. If it is lower or higher than expected, the RCR indicator is lower or higher than 1.00, respectively. Two other auxiliary lines on Fig. 17 are the world average MECR and MOCR values. These values are, by definition, identical.

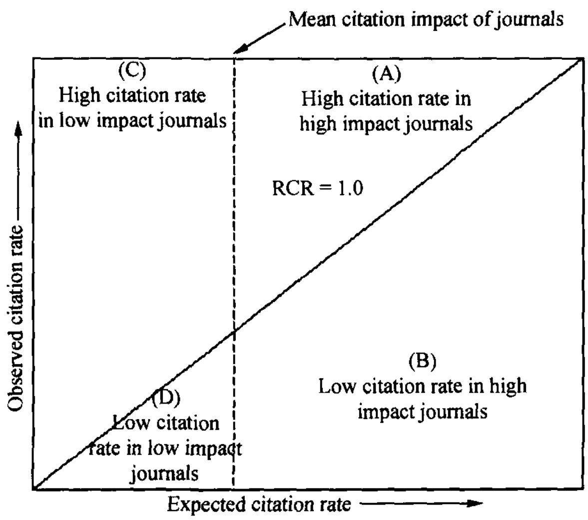


Fig. 17 Prototype of a Relational Chart and its zones.

Three lines divide the RC into six " zones" marked in Fig. 17 by the letters A to F.

Zone A: Scientists of countries in this zone publish in scientific journals having higher citation impact than the world average and receive higher citation rate than expected.

Zone B: Scientists of countries in this zone publish in scientific journals having higher citation impact than the world average and receive higher citation rate than the world average but lower than expected.

Zone C: Scientists of countries in this zone publish in scientific journals having higher citation impact than the world average, but receive lower citation rate than the world average and than expected.

Zone D: Scientists of countries in this zone publish in scientific journals having lower citation impact than the world average and receive lower citation rate than expected.

Zone E: Scientists of countries in this zone publish in scientific journals having lower citation impact than the world average and receive lower citation rate than the world average, but higher than expected.

Zone F: Scientists of countries in this zone publish in scientific journals having lower citation impact than the world average, but receive higher citation rate than the world average and than expected.

Fig. 18 shows the position of some countries on a RC based on total science data.

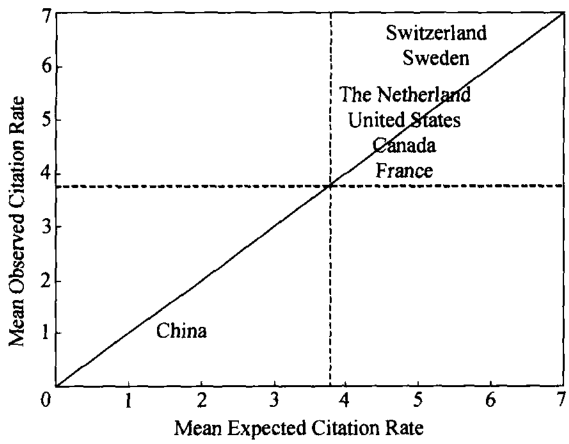


Fig. 18 Relational Chart for 32 countries in all fields of science (combined).

Table 5 gives the relative position (rank) of China by its Relative Citation Rate (RCR) indicator in all science fields combined and in the 12 main science fields separately, as well as their zone code (A to F) in the corresponding RC.

Conclusions

(1) The journal literature can be used for comparison of the publication activity of countries in quantitative form, in any scientific field and subfield, if the used sources and methods are characteristic for all the countries investigated and the number of processed items (journal articles, citations) is statistically significant.

(2) We must, however, accentuate that all results and conclusions of this work are related only to the 1990—1998 period.

Table 5 China's comparative rank in world science based on relative publication, citation Indicators (RCR) and its position in the Relation Chart (RC) 1990—1998

Position of P.R. China in the World according to three main scientometric indicators (Publications: 2000; Citations: 2000-2002)

Field	Number of Publications	World Rank	Mean Observed Citation Rate (MOCR)	World Rank	Mean Expected Citation Rate (MERC)	Relative Citation Rate (RCR)	World Rank
AGRI	958	13	1.744	39	1.516	0.869	48
BIOL	1418	15	2.284	47	1.992	0.872	49
BIOS	1966	13	3.039	57	2.544	0.837	54
BIOM	1004	13	2.989	38	2.789	0.933	44
CLI1	1341	15	3.856	51	3.910	1.014	45
CLI2	1323	21	2.688	47	2.508	0.933	50
NEUR	434	16	3.419	36	3.036	0.888	38
CHEM	11872	4	1.925	48	1.825	0.948	35
PHYS	7621	7	2.066	52	1.653	0.800	47
GEOS	1538	10	1.988	60	1.761	0.886	50
ENGN	3468	5	1.026	51	0.870	0.848	48
MATH	2016	5	0.941	34	0.930	0.988	36

AGRI: Agriculture & Enivronment
BIOL: Biology (Organismic & Supraorganismic Level)
BIOS: Biosciences (General, Cellular & Subcellular Biology; Genetics)
BIOM: Biomedical Research
CLI1: Clinical and Experimental Medicine I (General & Internal Medicine)
CLI2: Clinical and Experimental Medicine II (Non-Internal Medicine Specialties)
NEUR: Neuroscience & Behavior
CHEM: Chemistry
PHYS: Physics
GEOS: Geosciences & Space Sciences
ENGN: Engineering
MATH: Mathematics

(3) An interesting phenomenon, is we call the "banana shape" pattern of the points which appears on all our relational charts. This means that Zones E and F are totally unpopulated, Zones B and C are scarcely populated, Zone A is reasonably populated and the mass is concentrated in Zone D.

(4) Preliminary conclusions would attest that at a statistical significant level of publications and citations, the publication in high impact journals will attract many citations. This causes a certain "magnetic effect" which results in a higher citation rate for author who have published in highly cited journals even in cases they publish in low impact ones. On the other hand it seems that the striving to publish in high or higher impact journals although in some cases not rewarded by high citation rates, has at least a promise for a positive change in the zoning on the relational charts (RCs).

(5) Table 5 shows facts regarding the competitive aspects of Chinese science at the end of the second millennium.

References

[1] Menard H., Science, Growth and Change, Harvard Univ. Press, Cambridge, Mass., 1971.

[2] Schubert A., Glänzel W., Braun T. Against Absolute Methods: Relative Scientometric Indicators and Relational Charts as Evaluation Tools, in A. F. J, van Raan (Ed.): Handbook of Quantitative Studies of Science and Technology, North-Holland, Amsterdam, New York, Oxford, Tokyo, 1988.

[3] Kealey T. The Economic Laws of Scientific Research, MacMillan Press Ltd., London, 1996.

[4] The interested reader is suggested to browse the volumes of the journal *Scientometrics*, Elsevier Science Publisher, Amsterdam.

[5] Braun T., Glänzel W., Schubert A. Scientometric Indicators Datafiles. A Multidimensional, Machine Readable Database for Evaluative Purposes, Scientometrics, 1993, 28: 137—150.

[6] Schubert A., Braun T. Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 1986, 9: 281—291.

[7] Glänzel W., Schubert A., Braun T. On the theory and applications of scientometric indicators, *Science, Technology & Human Values*, 1988, 13(1—2): 125—126.

[8] Schubert A., Glänzel W., Braun T. Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields. *Scientometrics*, 1989, 16: 3—478.