INFORMETRICS 87/88
L. Egghe and R. Rousseau (Editors)

MONITORING ADVANCES IN CHEMICAL ENGINEERING

H.P.F. PETERS, D. HARTMANN, and A.F.J. VAN RAAN

Science Studies Unit, Lisbon-Institute, University of Leiden, Stationsplein 242, 2312 AR Leiden, The Netherlands

Abstract

This paper describes an approach to monitoring scientific progress in chemical engineering in order to operationalize concepts such as 'research performance' which can be used in the retrospective evaluation and the future anticipation of scientific research activities. We focus on various quantitative methods. Bibliometric methods form an important, but not the only, part of the work. The use of bibliometric approaches and measures is plagued by many problems. This is especially true for the more applied sciences, like chemical engineering. Nevertheless, there is very few research on the problem of the use of bibliometric indicators in applied sciences. The differences between the many disciplines can be very large, and therefore it is important to investigate this point thoroughly, for each major applied discipline. Only then can other possibilities to construct performance indicators (e.g., on a non-bibliometric quantitative base) be studied systematically.
We proceeded along two main lines. First, a number of prestigeous journals is used as a data source. Secondly, outstanding scientists in chemical engineering have been identified, and their bibliographies have been analyzed. Using these different data sources, a variety of quantitative techniques is applied to elaborate the data and to construct (possible) indicators: publication analysis, (co-)word analysis, citation analysis (including journal-to-journal citation data). Furthermore, recently developed advanced data-analytical techniques are used to display complicated data-relations.
We thus explore the role of quantitative, and in particular bibliometric techniques, in monitoring advances in chemical engineering. The usefulness of bibliometric indicators will be discussed in detail, additional methods and techniques will be presented.

1. INTRODUCTION

Identification of scientific progress by careful operationalization of concepts such as 'performance' and 'knowledge growth' is of major importance in retrospective evaluation of (current) research and for anticipation of future research activities.
In recent years, we explored quantitative (particularly bibliometrical) methods and techniques to monitor research performance in basic sciences. Considerable difficulties, however, can arise for applied and technology-oriented research.
A major assumption in bibliometric research performance analysis is that scientific journals are the main - and in fact eventually the only important - carriers of knowledge. For applied sciences this assumption does not hold. An important difference with basic science is that journals are not the only knowledge carriers. Reports (for governmental organizations, for international agencies, for industry), patents, and conference papers play, in many applied fields, an important role. Still these non-journal data are bibliometric in character. An important difference with journals however is that citation analysis

becomes virtually impossible.
We studied chemical engineering as a field of applied science with, at the one side, strong connections with basic science, and, at the other side, strong connections with technology. It provides an interesting possibility to assess the applicability of quantitative, and in particular bibliometric methods for research performance measurement, and to find out what other, in particular non-bibliometric approaches will appear to be worthwile, or even necessary.
Different approaches will be discussed. First, premier ('top-') journals in chemical engineering have been identified and analyzed in order to picture, what we call, the 'high quality mainstream' of the field. Several quantitative techniques are applied: journal-to-journal citation analysis, journal-to-field analysis, co-word analysis. Secondly, we identified high reputation ('top-') scientists.
Characteristics of their publicatons were analyzed, with an emphasis on citation-analysis. Some preliminary results on co-word analysis will be discussed. We assume that an analysis of top-scientist activities will reveal the 'leading research fronts' of a field. Results of these approaches will be presented together with indications for future research.

2. BASIC DATA

2.1 Top Journal Data

The choice of 'most important' or 'top' journals in a particular field of science is a difficult one. Lists of such journals given by experts in the field will always somewhat be slanted by their own interests. We created a list of all journals used by scientists in Dutch chemical engineering departments over a period of five years (1980-1984) and asked several scientists to indicate the most important journals. Also, with the help of a postsurvey we recieved from foreign chemical engineering scientists their lists of most important journals. In addition, we checked the ISI impact factors for a large group of chemical engineering or chemical engineering-related journals. Finally, and of major importance, were our findings from the oeuvres of top scientists, especially the journals used by these scientists over a period of ten years.
We had to make a selection. Nine important journals were defined as a 'set of top journals'. Additional journals were used to compose a broader set of 21 journals for analysing a journal-network structure, involving more general parts of chemistry. In table I we present both our set of top-journals as well as the broader set of journals. Taking these nine chemical engineering top journals as one set, we performed a two-step frequency analysis of specific words in the titles of publications within these nine journals, of keywords and of database ('controlled') terms. This frequency analysis (using online techniques with Chemical Abstracts) has been performed for three periods of time: 1978-1980, 1981-1983, and 1984-1986. The idea behind this analysis is that it supplies a method for a time-dependent identification of words, or combinations of words, that might indicate research areas within the high quality mainstream of chemical engineering. The total number of publications involved in this word analysis was 11461 (3475 in 1978-1980, 3578 in 1981-1983, and 4408 in 1984-1986).
Completely different from the above type of journal data are the journal-to-journal citation data. These data describe the mutually citing process between journals. From the Science Citation Index (SCI), more specifically from the Journal Citation Reports (JCR), we collected citation data for the broader set of 21 journals, for the periods 1976-1980 and 1981-1985. Here we aim at structuring journal networks and determining the change in such a network over time. These journal networks can reveal characteristics in a specific field, in relation to neighbouring fields.
A third type of data is the relation between journals and specific subfields. Some journals are quite general, other journals focus on rather specialized, older journals may shift towards new (sub)fields. We collected data for the nine top journals on the number of publications classified by Chemical Abstracts (CA) in

different subfields. With these data a journal-to-field analysis was performed for the periods 1978-1980,1981-1983, and1984-1986.

Table 1: Top-journals active in the field of Chemical Engineering and related journals.

Journal Name	**Abbreviation**	**Number in j-j display**
top-journals:		
Aiche Journal American Institute of Chemical Engineers	AIChE J	1
Canadian Journal of Chemical Engineering	Can J C Eng	4
Chemical Engineering Journal and the Biochemical Engineering Journal	Chem Eng J	5
Chemical Engineering Science	Chem Eng Sci	6
Chemie Ingenieur Technik	C I T	7
Industrial & Engineering Chemistry Fundamentals	Ind Eng C Fund	9
Industrial & Engineering Process Design and Development	Ind Eng C PDD	10
Industrial & Engineering Product Research and Development	Ind Eng C PRD	11
Powder Technology	Powder Technol	18
related journals:		
Applied Catalysis	Appl Catal	2
Biotechnology and Bioengineering	Biotech Bioeng	3
Journal of the American Chemical Society	J Am Chem Soc	12
Journal of Catalysis	J Catalysis	13
Journal of Chromatography	J Chromat	14
Journal of Molecular Catalysis	J Mol Catal	15
Journal of Organic Chemistry	J Org Chem	16
Antonie van Leeuwenhoek Journal of Microbiology	A van Leeuwen	17
Recueil des Travaux Chimiques des Pays Bas	Recl Tr Chim PB	19
Tetrahedron Letters	Tetrahedr L	20
Transactions of the Institution of Chemical Engineers	Tr Inst C Eng	21
Journal of Chemical Engineering of Japan	J Chem Eng Japan	22

2.2 Data on Top-Scientists and Their Research Programmes

In order to identify areas of excellence in chemical engineering we adopted the following basic assumption: research of high quality is defined by (at least) the work done by highly recognized scientists. Thus, the identification of areas of research excellence shifts toward the identification of highly recognized scientists. Here we state a second basic assumption: highly recognized scientists are often very productive researchers, and therefore, in a first approximation, we can apply the reverse: the search for very productive researchers offers a good possibility to identify highly recognized scientists, and, therefore, areas of excellence. We

emphasize however that this is only a first approximation, and by no means we assume that less productive researchers perform lower quality research.
Under the above assumptions, we searched the files of Chemical Engineering Abstracts (CEA). For a number of western countries (U.S., Great Britain, West Germany, France, Belgium, Switzerland and the Netherlands) we identified most publishing chemical engineering scientists for the years 1982-1985. These lists were discussed with an internationally recognized researcher in the field in order to account for database artefacts and errors and to choose twenty scientists from the original CEA data.
Although we carefully regarded in this choice the distribution of these top-scientists over countries and subfields, undoubtedly this choice is only a sample and we do not claim to fully cover the whole field of chemical engineering.
As a first approach, we have gathered bibliographic, i.e. publication and citation data on these twenty scientists. Using online data retrieval in Chemical Abstracts (CA), data on 1135 publications of these chemical engineering scientists in the period 1977-1986 were collected. Furthermore, for all publications of twelve top scientists from 1977 up to 1986 a citation analysis was performed.
After the above identification of the top chemical engineering scientists and the collection of data on their scientific oeuvres, we performed for each of the scientists a similar word analysis as we did for the top journals: a frequency analysis of specific words in the titles of publications, of keywords and of controlled terms. This frequency analysis was performed for the same periods: 1978-1980, 1981-1983, 1984-1986. In this way trends over time in the research agendas of top scientists can be assessed (by means of recently 'emerging' words or co-occurrences of words), thus identifying leading research fronts within chemical engineering (with the 'high quality mainstream' revealed by the top journal word analysis as a reference frame).

3. ANALYSIS OF RECENT DEVELOPMENTS IN CHEMICAL ENGINEERING

3.1 Trends in Chemical Engineering Revealed by Top Journal Data

3.1.1 Journal-to-Journal Citation Data

The basic assumption made here, is that relations between scientific journals could provide a 'worldwide framework' to position research fields in mutual connection. To do this, journals or groups of journals must be rather field-specific, but the model should allow for (the reality) that some journals take a very central position in a whole discipline. Such disciplinary core journals in a journal network structure are of particular importance, since they make it possible to assess the mutual relations of various (sub)fields with respect to the central cognitive developments of a major discipline such as chemistry.
So far we did not specify what precisely is meant by journal relations. There are a number of possibilities. Two of them can be operationalized in a straightforward quantitative way: journal-to-journal (j-j) citation data and journal-to-(sub)field (j-f) data. In the Journal Citation Reports (JCR) of the Science Citation Index, lists of all ISI-covered journals are given both in the citing mode, as well as in the cited mode. Having established a specific set of journals in some field of science, a citation matrix can be composed with the help of the above mentioned JCR data. Most citations within a journal are given to the same journal: 'journal self-citation', which results in strong diagonal elements of the matrix. Such a strong diagonal dramatically influences most of the well-known multi-variate analysis techniques and therefore the (relatively) weak connections between different journals will come out much less clear or even disappear.
For this type of citation-data matrices we applied a recently developed data-analytical technique, quasi-correspondence analysis. An extended description of this technique is given by Tijssen, De Leeuw and Van Raan (1987). We here present the results of a quasi-correspondence analysis on the j-j citation data for

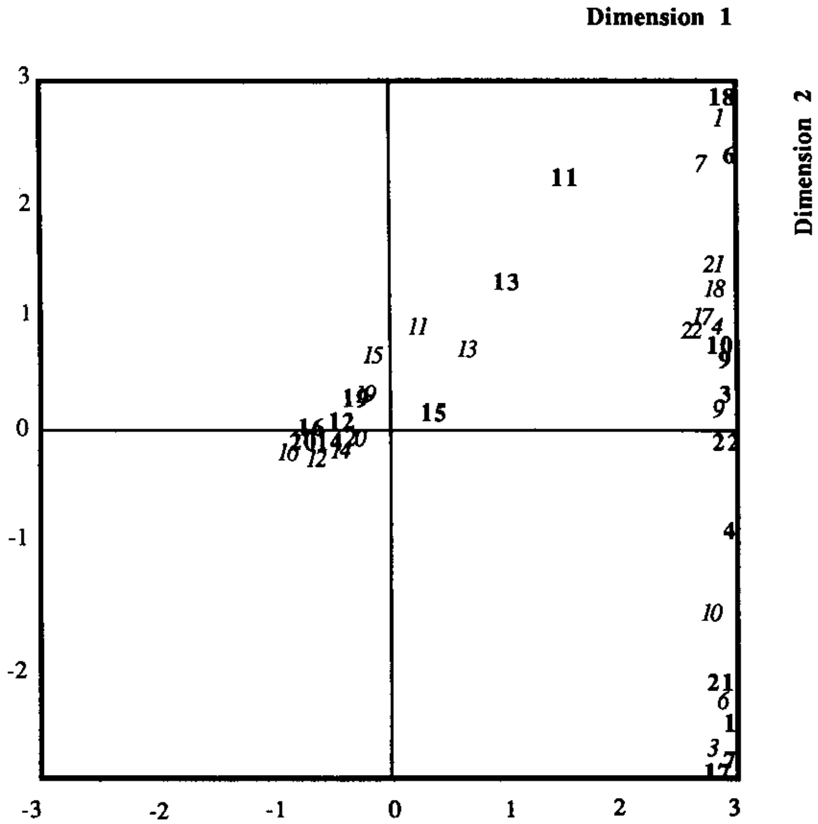

Figure 1: Quasi-correspondence analysis of a Journal-to-Journal citation matrix for 22 journals in Chemical Engineering and neighbouring fields for the period 1976-1980.

italics label: citing journal; bold label: cited journal
Accounted variance in dimension 1: 56%
2: 11%.

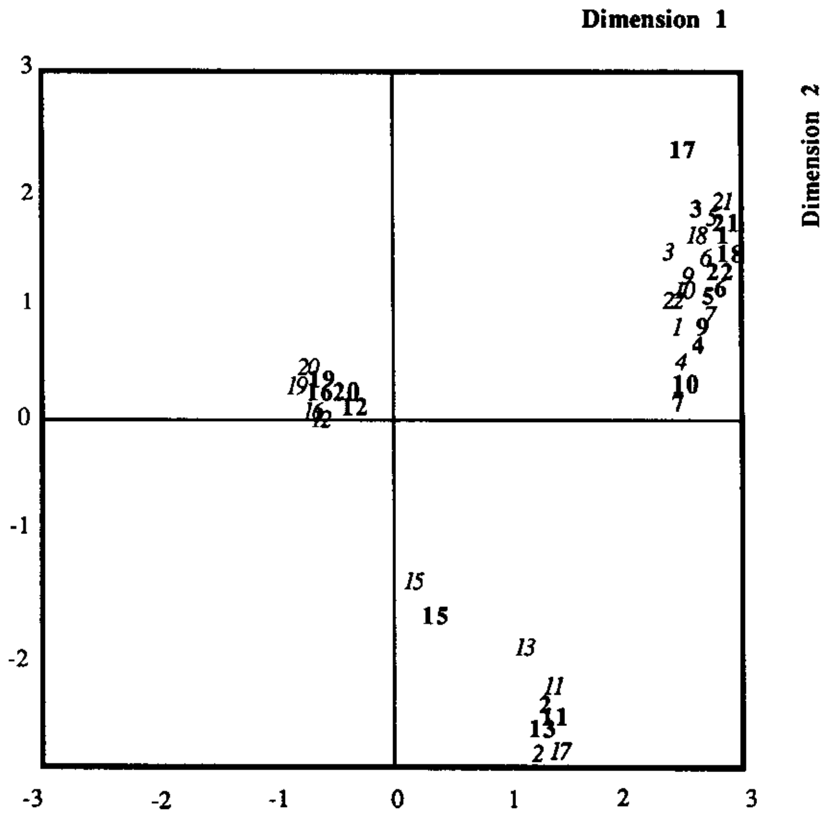

Figure 2: Quasi-correspondence analysis of a Journal-to-Journal citation matrix for 22 journals in Chemical Engineering and neighbouring fields for the period 1981-1985.

italics label: citing journal; bold label: cited journal
Accounted variance in dimension 1: 42%
2: 23%.

the 21 journals. In Figure 1 results are shown for 1980 (based on citations in the period 1976-1980), and in Figure 2 the journal relations are displayed for 1985 based on citations in the period 1981-1985). Comparison of both pictures reveals the dynamics of the journal network. We emphasize that quasi-correspondence analysis is not a multi-dimensional scaling technique, and therefore these figures can not be explained directly in terms of distances. Quasi-correspondence analysis indicates similarity-relations in both the citing and the cited mode as well. Thus, in Figure 1, the journals 12, 14, 16, 19, and 20 are very similar as a (bold label) cited journal, but they are also similar as a (italics label) citing journal. Furthermore, for these journals the cited and citing mode are also mutually similar. On the other hand, at the right side (upper quadrant) of Figure 1, we see that, for example, the journals 17 and 22 are very similar in the citing mode (italics), but they are much less similar in the cited mode (22 in bold label now is in the lower right quadrant near the horizontal axis, and 17 is at the bottom of this quadrant). It is also clear that the citing mode of 17 (italics) is not very similar to its cited mode (bold label).

We remember that the strenght of this type of analysis is twofold: it suppresses the extreme dominance of self-citation data, in order to focus on relations between different journals, and, secondly, it allows for displaying the journal network in the citing and in the cited mode, instead of taking the average of both modes.

The interesting point now, is to compare the 1980 display (Figure 1) with 1985 (Figure 2). We here focus on main impressions, since for a more detailed analysis one must allow for the accounted variance in the different dimensions and consider additional dimensions. First, we see a clustering of 'central', i.e. more general chemistry journals: 12, 16, 19, and 20. This cluster retains its position in the period 1980-1985. A second cluster comes about in the 1985 display, consisting of journals which were five years earlier rather dispersed: 2, 11, and 13, with 15 in an intermediary role with respect to the central cluster. It is a group of catalysis journals, and we conclude that the journal Ind Eng Chem PRD, although having not 'catalysis' in its name, must be strongly catalysis-oriented, much more than its closely related 'sister'-journals Ind Eng Chem Fund (9) and Ind Eng Chem PDD (10). The position of 17 is rather arbitrary, because of its very low or non-existent citation relations with other journals. The remaining journals form a third cluster, which was again more dispersed in 1980. This clearly is the chemical engineering group.

3.1.2 Positions of Journals in Chemical Engineering Subfields

In this section we present preliminary results of a new type of analysis developed in our group (Tijssen, De Leeuw, and Van Raan, 1988): the journal-to-field (j-f) analysis. In this analysis we established which CA-sections (chemistry subfields) are assigned to publications in particular journal, for 1977-1979, 1980-1982, and 1983-1985. There is always only one CA-section (subfield code) attached to each publication. These assigned field codes per journal can be ranked in a list (e.g., for the journal Chemical Engineering Science, 1983-1985, section 48 was found 506 times, section 67 122 times, etc.). Field codes representing less than 2% of a journal's publications were omitted.

In this way we constructed for all three periods a data matrix with nine important chemical engineering journals as items in the matrix columns, and 26 subfield codes as items in the matrix rows. Because no dominating diagonal is present in the matrix (like in the case of j-j citation) we used correspondence analysis instead of quasi-correspondence analysis (Tijssen et al, 1987,1988).

Let us discuss the main features of the three j-f displays presented in Figures 3, 4, and 5 for the periods *1977-1979, 1980-1982,* and *1983-1985, respectively. First,* the coupling with the j-j structure. In Tijssen et al (1988) we found for the case of materials science that the dimension with the highest accounted variance (i.e., the first dimension, often 40-60%) in both analysis (j-f and j-j) reveal a remarkably similar structure. The second dimension, however, clearly deals with different factors for the different types of analysis. And this is an intresting *finding: a further structuring of the journal-network* (beyond the main features

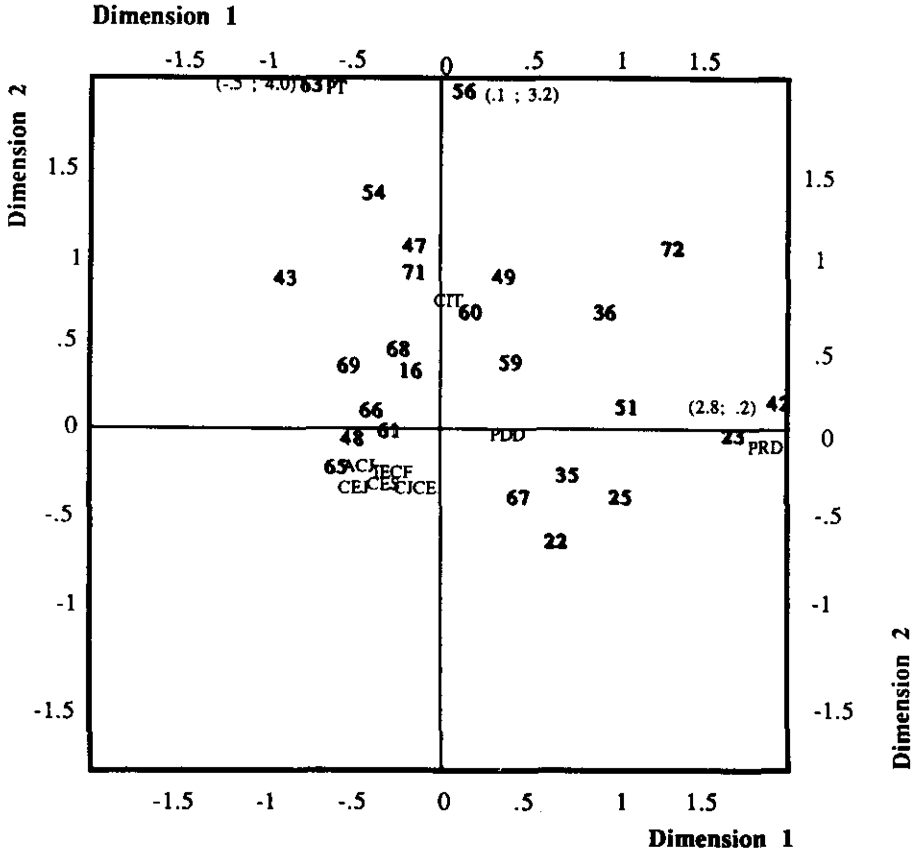

Figure 3: Correspondence analysis of a Journal-to-field matrix for the set of top-journals in the period 1977-1979.

Accounted variance in dimension 1: 40%
2: 25%.

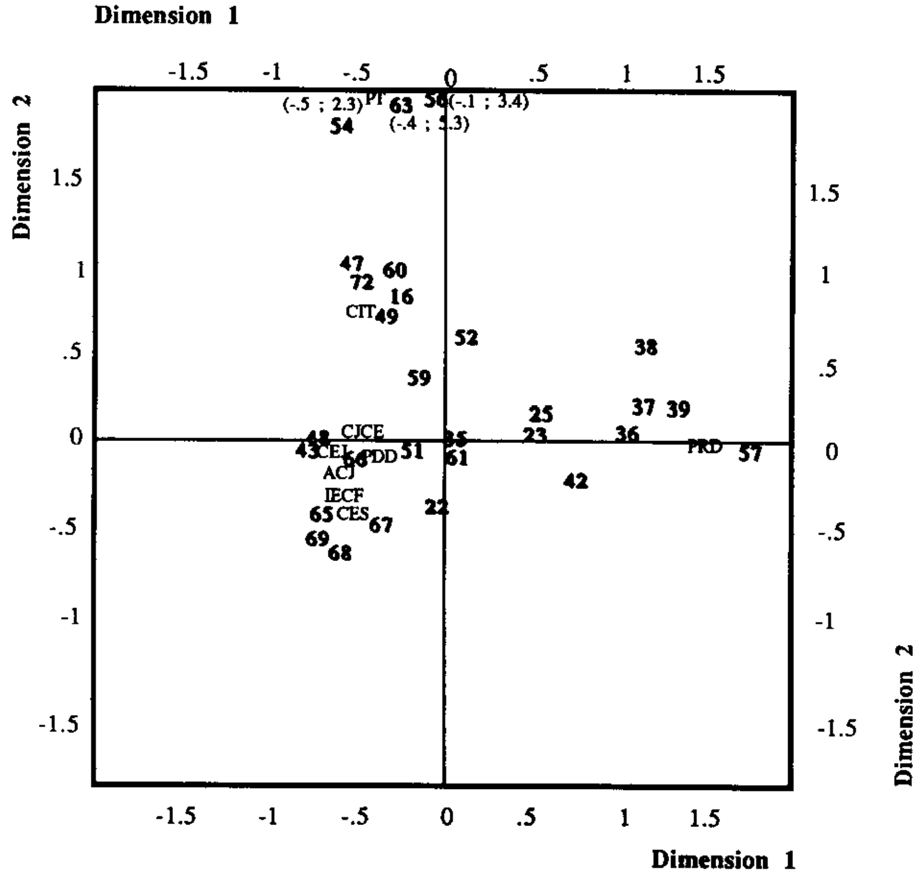

Figure 4: Correspondence analysis of a Journal-to-field matrix for the set of top-journals in the period 1980-1982.

Accounted variance in dimension 1: 71%
2: 10%.

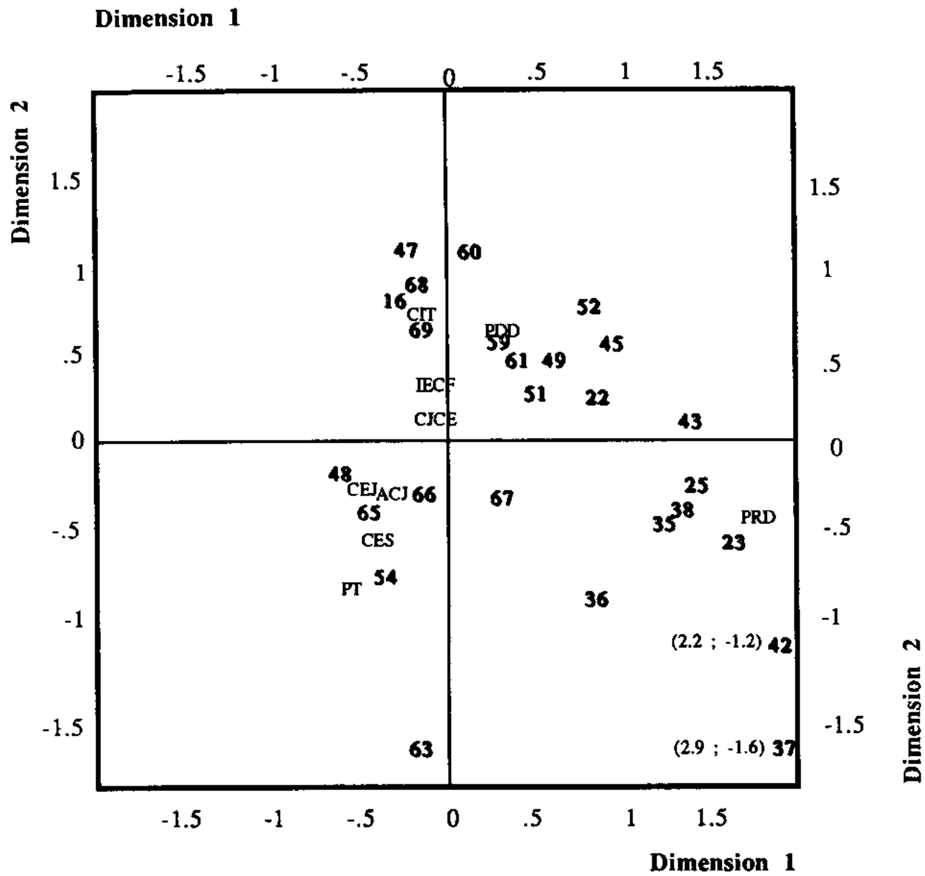

Figure 5: Correspondence analysis of a Journal-to-field matrix for the set of top-journals in the period 1983-1985.

Accounted variance in dimension 1: 49%
2: 16%.

pictured by the first dimension) takes places.
We see that the main structure of the journal network given by the j-f analysis is similar to that resulting from the j-j analyses: a cluster of typical chemical engineering journals with an exceptional position of (Ind Eng Chem) PRD. Note that in the j-j analysis the set of journals was larger (21) than the set of journals in the j-f analysis (9), because in the first analysis we also aimed at picturing the relation of chemical engineering journal with important general chemistry journals. Therefore, this cluster of general chemistry journals (in the j-j analysis) is not present in the j-f analysis. Including such general chemistry journals in a j-f analysis would of course have increased the number of involved (sub)fields enormously, making a j-f display rather chaotic.
Comparison of j-j and j-f results in the second dimension should be made with great care: the accounted variance in the second dimension for the j-j 1980 analysis is only 11%. This is also 'visible' in Figure 1: the chemical engineering journals are completely scattered in this dimension. In the j-j 1985 analysis (Figure 2) the situation has considerably improved, now 23% of the total information is accounted for the second dimension. We see that Chemie-Ingenieur-Technik (CIT) and Ind Eng Chem Proc Des Dev (PDD) take rather extreme positions, with Ind Eng Chem Fund (IECF) and Can J Chem Eng (CJCE) taking intermediate positions between CIT and PDD at the one side and the other chemical engineering journals at the other side. This structure is confirmed by the j-f 1983-1985 analysis (Figure 5).
After this comparison of main features between j-j and j-f analysis, we now focus on specific findings in our j-f displays.
A very important CA-section (subfield) is 48: Unit Operations and Processes, since more than 40% of all top-journal publications throughout the years 1977-1985 are assigned to this subfield. It appears to be the very heart of chemical engineering. Next important subfields are 51 (Fossil Fuels, Derivatives & Related Products) and 67 (Catalysis, Reaction Kinetics & Inorganic Reaction Mechanisms). Subfields of minor importance are 22 (Physical Organic Chemistry), 35 (Chemistry of Synthetic High Polymers), 47 (Apparatus & Plant Equipment), 59 (Air Pollution & Industrial Hygiene), 60 (Waste Treatment & Disposal), 65 (General Physical Chemistry), 66 (Surface Chemistry & Colloids), and 68 (Phase Equilibriums, Chemical Equilibriums & Solutions). Some of the sections are becoming increasingly important, for instance section 45 (Industrial Organic Chemicals, Leather, Fats, and Waxes), 51, and 52 (Electrochemical, Radiational & Termal Energy Technology). Sections of decreasing importance are 22, 23 (Aliphatic Compounds), 47, and 65.
We now know the main subfields (CA-sections) for the top-journals, and this knowledge can be used as benchmarks within the various j-f displays. We see, that the position of the major section 48 - near the center of the picture - remains virtually the same for the subsequent periods. This nicely illustrates that the data-analytical technique indeed positions subfield 48 as a central item.
We can also see that together with 48, the subfields 65 and 66 remain near the center with the journals Chemical Engineering Science (CES), Chemical Engineering Journal (CEJ) and the American Institute of Chemical Engineers Journal AIChE), here abbreviated as ACJ. Remarkably related with the position of Ind Eng Chem Proc Des Dev (PDD) are the subfields 51, 59, and, to a lesser extent, 22 and 67. Notice the movement of PDD (with the related subfields) from the 'PRD-side' (in 1977-1979), via the 'chemical engineering center' (in 1980-1982), toward the more or less fixed position of CIT (1983-1985). This last and most recent step is for more reasons remarkable: not only PDD moves in the direction of CIT, but IECF and CJCE also shift from the central chemical engineering cluster into about the same direction. A first interpretation, checking also the raw data, leads us to the conjecture that for the above journals, and in particular for PDD, the subfields 16 (Fermentation & Bio-Industrial Chemistry), 52 (Electrochemical, Radiational & Thermal Energy Technology), and 60 (Waste Treatment & Disposal) have become increasingly important.
Another nice finding from these displays concerns the journal Powder Technology (PT). In the early periods, this journal was rather an outsider, strongly related

with the subfield 48[1]) 54 (Extractive Metallurgy) and, surprisingly, 63 (Pharmaceuticals). This subfield 63 is (for all three displayed structures) an outer province of the chemical engineering subfields family. But it is evident that powders plays a role in engineering (metallic powders), but also in medical drugs! What however leads PT in the most recent period toward the center of chemical engineering? Inspection of the raw data reveals that in this last period (1983-1985) there was more than a doubling of PT publications in subfield 48 (the core subfield, Unit Operations & Processes) as compared to the periods 1980-1982 and 1977-1979. No other journal shows such a strong increase of publications in this subfield.

3.1.3 Word and Co-word Analysis

We performed for the periods 1978-1980, 1981-1983, and 1984-1986 a two-step frequency analysis of specific words for the publications in the nine top-journals. Words in titles of publications, keywords, and controlled terms given by CA were involved. The numbers of publications in this analysis are 3475, 3578, and 4408 for the successive time periods.
The aim of the first step is to find characteristic words or combinations of words that appear in the most recent period (1984-1986) more frequently than in the foregoing periods. Since these words originate from publications in top-journals, we assume that these (combinations of) words are indicative for, what we would call the 'high quality mainstream' of chemical engineering. We listed these words in part A of Table II.
After identifying these first-step words, we performed a second word frequency analysis for all publications in the nine top journals in the period 1984-1986, but now successively for each first-step word in the title of the publications involved. This second step frequency analysis thus reveals the words that are most close to the main, first step words. A (second) word frequency analysis of the database controlled terms (which are word-combinations rather than single words) of the publications concerned confirms these findings and adds additional information. In this way, important research themes and subjects are identified. They are listed (no order of importance implied) in part B of Table II.
Of course, for an expert in the field the results will not be new. We like to emphasize however that with a relatively simple technique valuable information about recent developments in a particular field can be revealed. Nobody is expert in all fields, or even in all subfields of his own field. Scientists could be in need of information about new developments in neighbouring (sub)fields. Research committees might have similar problems with need for information. Having done a second word frequency analysis, one can construct a matrix with the same (first step) words in the rows and in the columns of the matrix. The rows indicate, for each word, the frequency with which the other words occur in combination with the word to which the rows belongs. We added the following words which appeared to be of general importance: absorp-, alumin-, bed-, extract-, gas-, mix-, model-, particl-, porous-, separat-, slurr-, stirr-. As a suitable multivariate analysis technique to display a complete 50 x 50 'co-word' matrix, the multi-dimensional scaling technique SMACOF (Stoop and De Leeuw, 1982) was choosen. The result of this co-word analysis is presented in Figure 6. From this figure it is clear that with the applied data-analytical technique practically no clusters are formed. However, an important check is the relation of the words 'fluidized' (11) and 'bed' (48), and also 'Fischer Tropsch' (36) and 'synthesis' (21). A minimum outcome of each applied technique must be a very strong coupling of these two above word-pairs. There is no doubt that this is

---

1 To see this one must take into account the projection of PT on the most important (i.e. first) dimension.

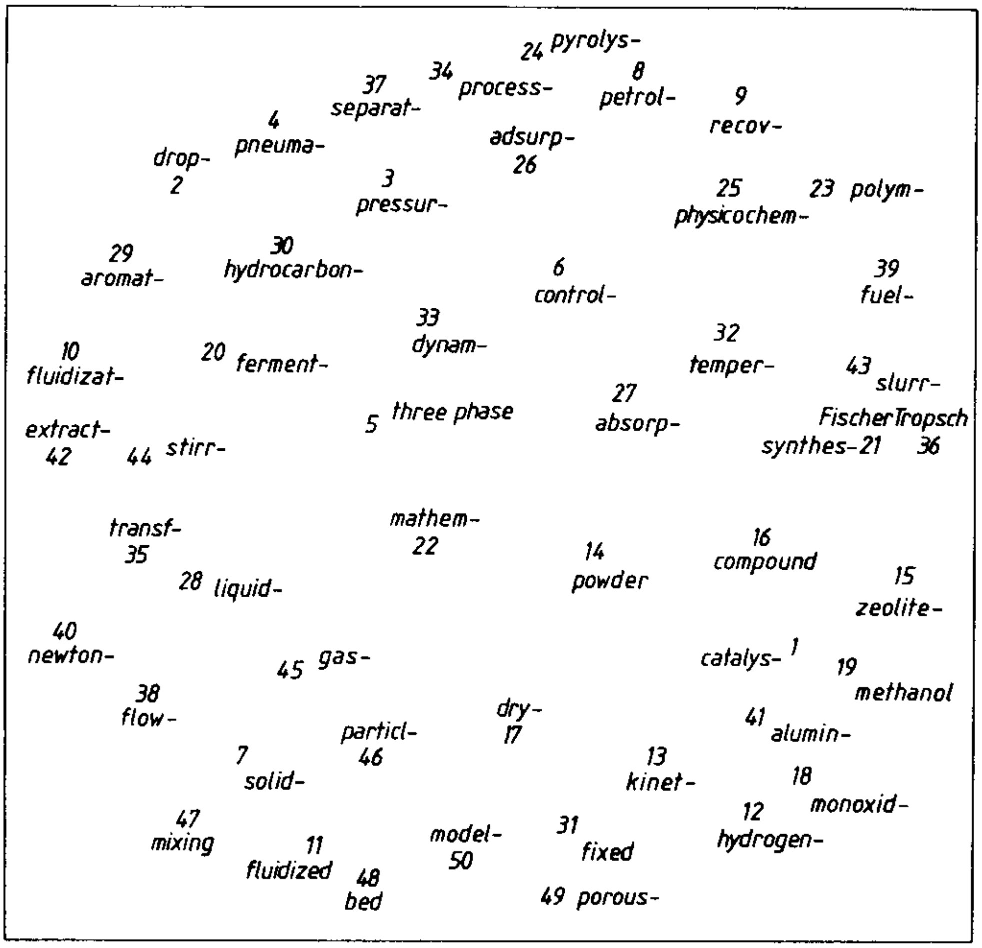

Figure 6 : Multidimensional Scaling display of a co-word matrix for top-journals in the period 1984-1986.

Table II: Word Analysis for Top-Journals 1984-1986 (alphabetical order)

A. First-Step Words

adsorp-	fluidized	polymeriz-
aromat-	fuel-	powder-
catalys-	hydrocarb-	pressure-
compound-	hydrogen-	process-
control-	kinetic-	pyroly-
drop-	liquid-	solid-
dry-	mathemat-	synthesis-
dynamic-	methanol-	temperature
ferment-	monoxide	three phase-
Fischer-Tropsch	newton-	zeolit-
fixed-	petroleum	
flow-	physicochem-	
fluidizat-	pneuma-	

B: Research Subjects Constructed After Second Step

Kinetics of Catalysis (Fischer-Tropsch Synthesis, oxidation, hydrogenation, deactivation by coke deposition, fixed bed reactors, porous and cellular materials, methanol, zeolites, nickel/alumina/iron/platinum/cobalt ...)

Particle Transport (pneumatic, pressure drop, solutions, suspensions, bubbles, diffusion (porous media), viscosity, heat and mass transfer, particle interaction with non-newtonian fluids, dispersion of solids in liquids, size of droplets, ...)

Separation Techniques (distillation, adsorption, absorption, membranes (permeability), sieves, drying, process simulation, vapor pressure prediction, (supercritical) extraction (columns), zeolites, electro- and ionophoresis ...)

Three-Phase Systems (fluidized bed, mass and heat transfer, Patulin production, gas-liquid-solid equilibria, gas slurry contact in bubble columns, mixing, ...)

Process Control & Dynamics (destillation, mathematical/physicochemical models, reactor design, process simulation, fermentation process monitoring, foaming control, hazard management, ...)

Petroleum Refining (aromatic hydrocarbon production, alkanes, thermodynamics/equation of state, petroleum recovery ...)

High Temperature Processes (pyrolysis (flash-), kinetics of thermal decomposition, (wood) gasification, fuel gas manufacturing, oil shale ...)

Powder Technology (fluidization, pharmaceutics, particle size and shape, compressibility, ceramics, particle flow, porosity ...)

Fermentation (bioreactors, ethanol production, penicillin, ...)

indeed the case: words 11 en 48, and words 21 en 36 are positioned very closely. We hoped to find a sort of hierarchical structure in the word-clustering, which would provide us with a 'natural' scheme of interrelated chemical engineering research topics.
In our opinion, it is too early to judge the applicability of co-words analysis on the basis of these results (for a detailed discussion of this technique see Callon et al 1986). First, our 50 x 50 co-word matrix is, because of the applied two-step word-frequency analysis, rather 'empty'. Knowing this, we will repeat this analysis up till low frequencies, thus filling the matrix and distinguishing between small co-occurrences of words and 'real zeroes'. It is important to investigate whether such 'low-threshold' values indeed improve the co-word analysis, or that it only will add noise. Secondly, and probably more important, is a further improvement of our data-analytical techniques, thereby concentrating on the 'not-missing values', i.e., eliminating the influence of the numerous missing values. Thirdly, 50 words is probably too few for structuring a large field like chemical engineering.
Although no clustering of words is achieved yet (couldn't it be so that in chemical engineering 'everything' is connected with 'everything', i.e. is it a field which is very broad and 'multi-topic' in character?), we undoubtedly can see a 'kinship' of subjects (words) in the figure. For example, the lower right quadrant is typically catalysis (1): we see neighbour-relations with 12 (hydrogen-), 13 (kinet-), 15 (zeolite-), 18 (monoxide-), 19 (methanol), 41 (alumin-), and, on a somewhat larger distance (probably because these subjects have interrelations with subjects in other quadrants) we see 36 (Fisher-Tropsch) connected with 21 (synthes-) at the one side, and 49 (porous-), 31 (fixed-(bed, 48)) at the other side. These latter two subjects (porous- and fixed (bed)) form a bridge to the lower left quadrant. In this quadrant we see central activities of chemical engineering: fluidized (11) bed (48) systems, closely interrelated with subjects as 46 (particle-), 47 (mixing-), 50 (model-) and with a sub-group containing 7 (solid-), 45 (gas-), 38 (flow-), 28 (liquid-), and 35 (transfer-). Subject 40 (newton-) is, as to be expected, closely related with 28 (liquid-), 35 (transfer-) and 38 (flow-). An artefact of our word-frequency analysis is that 'newton' shows up, in stead of 'non-newton-'. A slight improvement of the technique will correct for this (not unimportant!) omission, as we already checked. The upper two quadrants are less easy to understand. There is a clear neighbour-relationship (left side) between 3 (pressure-), 2 (drop-), 4 (pneuma-), 37 (separat-), but improvement of the analysis is necessary to understand the positions of, for example, 20 (ferment-), 32 (temper-), 43 (slurr-) and, in the lower quadrants, 14 (powder) and 17 (dry).

3.2 Trends in Chemical Engineering Revealed by the Work of Top-Scientists

3.2.1 General overview of Top-Scientists Publication Activities

The distribution of research output of top-scientists over various types of publications can be characterized as follows. Journal articles form the major source of publications: 78.4% of all publications, with proceedings papers in second place: 16.9%. Only 5% is published in other ways, for example patents. Of these journal papers, 51% is published in our earlier defined set of top journals. Like in the case of top-journals, section 48 is of major importance: 50.1% of all publications from top-scientists is assigned to this section; 11.3% of the publications is assigned to section 51, which is about the same as for the top-journals. Only 4.0% is assigned to section 67, against 9.1% in case of the top-journals. Other important sections for the top-scientists, as compared to the top-journals are 16 (Fermentation & Bioindustrial Chemistry), 22, and 60. All other publications are assigned to sections of minor importance, i.e. 47, 42, 66, and 68. In general we can conclude that the top-scientists are active in almost the same subfields as all scientists publishing in our set of top-journals.

3.2.2 Citation Analysis for Top-Scientists

Publication counts give a measure of scientific 'productivity'. Counting how often these publications are cited, may add operationalization of at least one important aspect of quality, namely 'impact'.
A manifold of methodological and technical problems is involved in citation analysis. In our group extended and very detailed studies has been performed on this subject for several years (The Leiden Indicators Project, see Moed, Burger, Frankfort, and Van Raan, 1983;1985). The analytical method discussed in this section is entirely based on this earlier work.
For a group of twelve top-scientists we collected online (using ISI's online version of the Science Citation Index, available from the host DIMDI at Cologne) data on articles citing the ('target') publications of the top-scientists.
The next step was a very careful, manual matching of the target-publications with data of the citation records. In this way we counted per year the total number of publications citing one specific target-publication.
With the thus collected citation data, we constructed 'impact indicators' (i.e., bibliometric indicators based on publication and citation data). For instance, we counted the number of citations received by all 1977 to 1980 publications of the 12 top-scientists, in the period from their year of publication to six years thereafter. Figures 7 and 8 represent the average number of citations received by different types of publications as a function of their age.
For basic research (Moed et al 1983,1985), it appears that most papers reach their highest citation-rate within three years of publication; then the number of citations received gradually decreases. In our case, however, it is unclear, for any type of publication, when they reach their highest citation rate, except the journal articles not published in the top-journals, which seem to reach a maximum in about the 5th year after publication. Books reach a high citation-level compared to other types, but we also immediately see that books take a long time to reach their citation peak, yet two years after publication books are better cited than any other type of publication.
Are citation-scores a valuable operationalization of research impact? If this is indeed the case, publications from top-scientists should be cited more often than those from other scientists. It also would mean that top-journals should be cited more often than other journals. These two hypotheses were tested in the following way. We restricted ourselves to short-term impact, which refers to the impact of researchers (or journals) at the research front just a few years after the publication of research results. A citation-window of three years after publication was used, which made it possible to assess impact of publications up to 1984. We collected citations of all ISI-journals (i.e., journals covered by ISI) articles published by the 12 top-scientists in the years 1977 to 1984. For example, we counted how many times all articles published in 1977 were cited in the years 1977, 1978, and 1979. Using these three-year citation-windows, we calculated the number of citations, for two different sets of publications, published during the years 1977 to 1984: (a) articles published in the set of top-journals and published by the 12 top-scientists; and (b) articles published by the 12 top-scientists, but in other ISI-journals.
Numbers of citations are not always easy to interpret. What we need is a (field-specific) reference value, to compare with. Following the lines of our earlier work, we calculated for the years 1977-1983 a weighted average Journal Citation Score (JCS) for the group of nine top-journals in which our top-scientists have published. We call this dataset (c). Thus an 'expected' value, i.e. the number of citations received by an average article in a specific (set of) journal(s), is constructed. To calculate these expected values, we used citation data from the ISI Journal Citation Reports (JCR).
Results of the above analysis are presented in Figure 9. Comparison of the curves for datasets (a) and (c) verifies our first hypothesis, namely that publications from top-scientists are cited more frequently than those from other scientists. This also indicates that our method to choose top-scientists converges with the citation analysis method. Comparing curves (a) and (b), we also see that top-journals are cited more often than other journals, which again is an

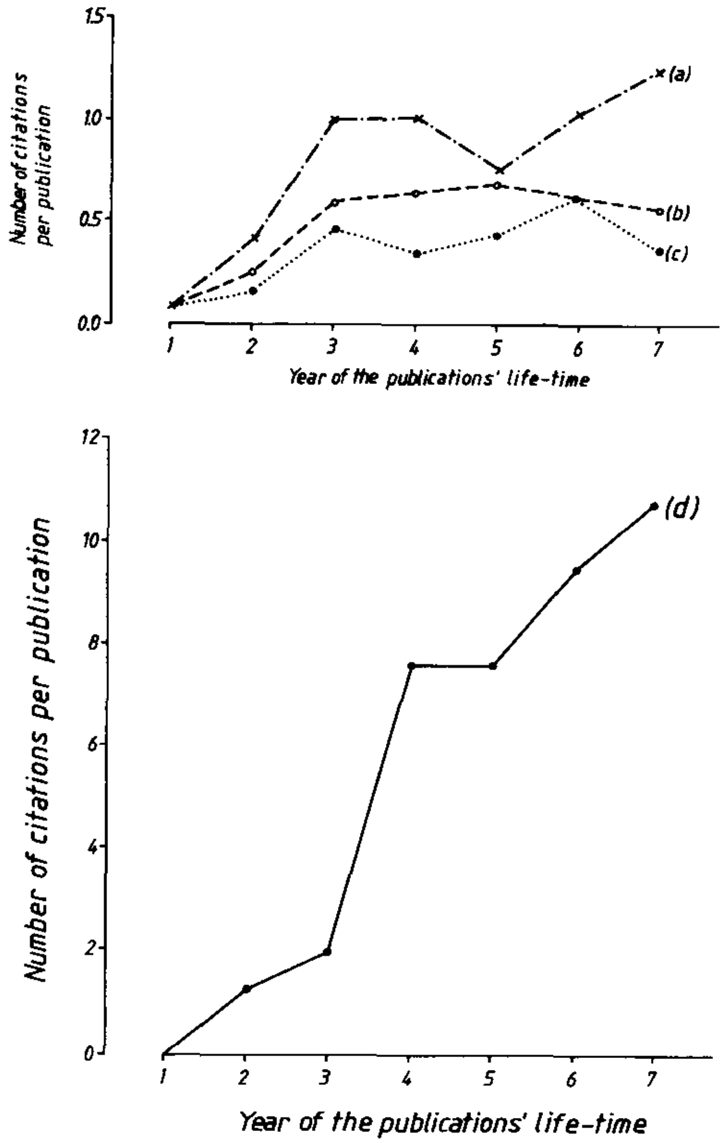

Figure 7,8: Curves representing the average number of citations received by different types of publications as a function of their age. Publication counts for the period 1977-1980 are aggregated. The citations received by these publications starting from the year of publication (1) to the seventh year of their lifetime (7) are counted, excluding in-house citations. Each curve represents a different type of publication:

a. articles published in the set of top-journals, except review-articles (n=126);
b. articles published in other journals, but covered by ISI (n=57), again with the exception of review-articles;
c. conference articles (n=40);
d. books (n=7).

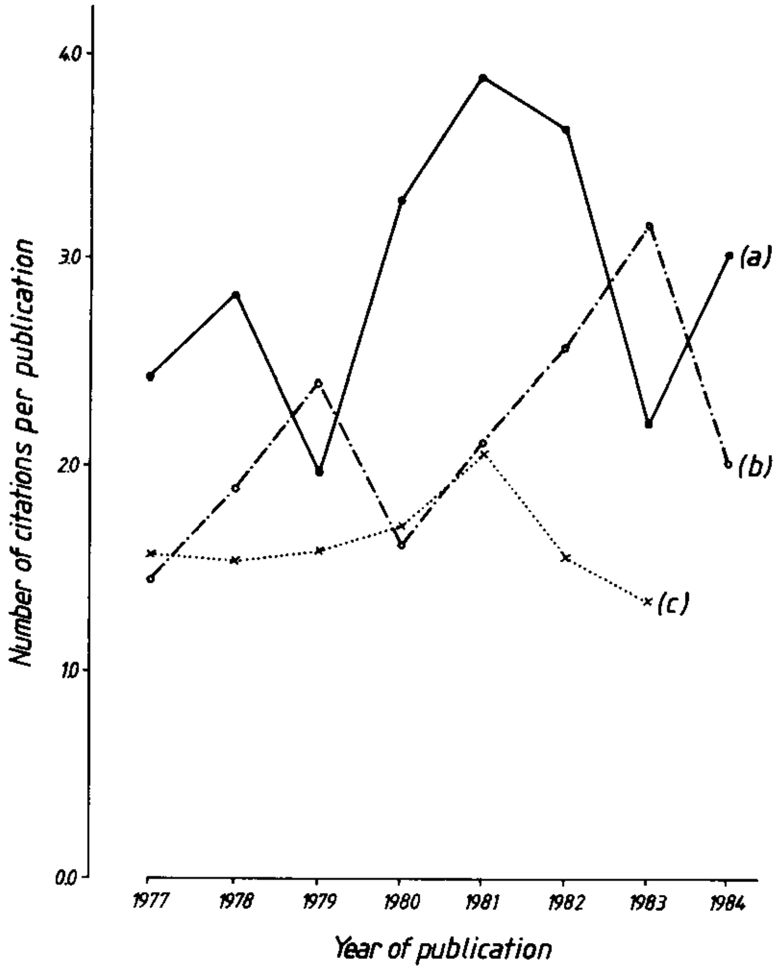

Figure 9: Curves representing the number of citations per publication received by different sets of publication during the first three years of their lifetime, including in-house citations for the years 1977-1984. The three curves represent:

curve a) articles published in the set of top-journals and published by one of the twelve top-scientists (n=271);
curve b) articles published by one of the 12 top-scientists, but in other ISI-journals (n=153);
curve c) the weighted-average Journal Citation Score (JSC) (or the "expected" publication impact) for the packet of top-journals of top-scientists concerned.

interesting convergence of our top-journal choice and citation-based results. Till now, we did not distinguish between citations and self-citations. In fact, in all the calculations so far, self-citations have been included. We here introduce a rather broad concept of self-citation which we call 'in-house citation' (Moed et al. 1983, 1985): all citations by one or more co-authors of one specific top-scientists are considered to be in-house citations. Our semi-automatically performed citation analysis can easily take these considerations into account.
In the Leiden Indicators Project we found that approximately 30% of all citations appears to be in-house citations. Here about 42% of all citations appears to be in-house citations.
A further step in citation analysis is the identification of highly cited work and a synthesis of these results with the findings from other approaches, such as a (co-)word analysis. This work is in progress and will be published elsewhere. Preliminary results on (co-)word analysis are presented in the next section.

3.2.3 (Co-) Word Analysis: Leading Research Fronts

In order to pursue a detailed comparison of identifying recent scientific progress by citation analysis and by (co-)word analysis, we also performed a (co-)word analysis with all publications (1135) in the ten-year oeuvres of our top-scientists. From the title-words, keywords, and controlled terms, the most frequent words were used as 'research front identifier terms' (first-step word analysis). Instead of an online second-step word-analysis (like we did for top-journals) we performed a more extensive manual (second-step) analysis since we have the complete CA publication records for all top-scientists available. For the last three-year period (1984-1986), we compared the first-step words with information (title of publication, uncontrolled terms, controlled terms) contained in the records of the publications in that period, in order to construct meaningful description of the research fronts. This analysis yielded detailed and interesting information about recent developments in chemical engineering. An example of our analysis for a few selected main topics is presented in Table III. We see that these results are more specific (more 'specialty-like') than the top-journal (more 'mainstream') co-word results. A detailed presentation of the results will be published elsewhere.

Table III: Example of some Top-Scientists Research Topics

Catalytic Processes:
- desulfurization, denitrification
- effect of water
- simulation models
- Pt, Al, Co catalysts
- H recovery from gas mixtures by metal hydrides in slurry
- catalyst deactivation by coke deposit
- kinetic (in)stability
- fused-magnitite catalysts in Fischer-Tropsch synthesis
- effects of liquids in catalytic pores on slurry Fischer-Tropsch synthesis

Separation Techniques:
- permeability of membranes
- flow of liquid through screens
- distillation by diffusion
- deep bed filtration
- separation in desorption
- segregation of solids in fluidized beds
- Fischer-Tropsch induced growth of chemical chains
- crystal attrition in stirred vessels
- drying of porous particles with binary mixtures

Fischer-Tropsch Kinetics:
- Fischer-Tropsch synthesis in bubble column slurry reactor on suspended iron catalysts (fused-magnetite catalysts)
- Fischer-Tropsch synthesis in slurry reactor, effects of liquids in pores.

4. CONCLUSIONS

Bibliometric analysis has shown to be a valuable tool for monitoring chemical engineering. In particular, citation analysis is applicable, although the citation-level is below that for the basic science and, even more important, the time scale is longer. This does not mean that citation analysis is useful to the same extent in all subfields of chemical engineering. Further research is needed to investigate more in more detail why scientists do have a high (local?) reputation but are 'bibliometrically invisible'.
Publications from internationally recognized top-scientists are cited more often than those from other scientists: the results of our method of chosing top-scientists converges with the results of our citation analysis method.
Top-journal articles are, on the average, much better cited than articles in other chemical engineering journals. Again, our method of choosing a set of top-journals converges with citation analysis. For the top-scientists we see, in general, a clear emphasis on journal articles, published mainly in top-journals.
Since journals play such an important role, further bibliometric techniques can be used: journal-to-journal citation data to picture the structure of chemical engineering and related fields by means of a network structure of journals; journal-to-field relations to position these journals in a multitude of subfields.
Journal-to-journal citation data revealed three main clusters of journals: the 'core' of chemical engineering, more general chemistry-oriented journals, and catalysis journals; journal-to-field analysis shows that Unit Operations & Processes is the most important chemical engineering subfield for the set of top-journals. The same is true for top-scientists. This subfield is strongly related to journals such as Chem Eng Sci, AIChEJ and Chem Eng J.
Finally, useful combinations of conventional bibliometric indicators and a more content-specific typology can be made. The crucial point here is that such a synthesis must discriminate mainstream research from that which really advances scientific specialties. To our opinion, the technique of a two-step word-frequency analysis shows promising possibilities.

A crucial next step is a linkage between findings from these various data sources. This linkage will guide the construction of a conceptual typology on scientific developments so that qualification like 'high performance' can be operationalized in different senses: thorough reviewing and overviewing of existing knowledge in a field; progress by synthesis of specific fields of research; progress by 'changing' or 'growing' of a field; progress by new applications from basic research, etcetera.
We think that the construction of this linkage is a major task in the near future.

REFERENCES

Callon, M., Law, J. and Rip, A., (eds.), Mapping the Dynamics of Science and Technology (Macmillan, London, 1986).

Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J. On the Measurement of Research Performance: The Use of Bibliometric Indicators (Leiden, Science Studies Unit, 1983).

Moed, H.F., Burger, W.J.M., Frankfort, J.G. and Van Raan, A.F.J., The Use of Bibliometric Data as Tools for University Research Policy, Research Policy 14 (1985) pp. 131-149.

Stoop, I. and De Leeuw, J., How to use SMACOF-IB. (Department of Datatheory, University of Leiden, 1982).

Tijssen, R.J.W., De Leeuw, J. and Van Raan, A.F.J., Quasi-correspondence Analysis on Scientometric Transaction Matrices, Scientometrics 11 (1987) pp. 351-366.

Tijssen, R.J.W., De Leeuw, J., Van Raan, A.F.J., A Method for Mapping Bibliometric Relations Based on Field-Classifications of Articles, these proceedings (1988).

Van Raan, A.F.J., Indicators of National Strength and Weakness by Micro-Scale Monitoring of Research Performance (OECD Report DSTI 2957S, Paris, 1986).