Correspondence Factor Analysis of the Publication Patterns of 48 Countries over the Period 1981–1992

Jean-Christophe Doré*
CNRS URA 401, Muséum National d'Histoire Naturelle, 75231 Paris Cedex 05, France

Tiiu Ojasoo
Université Pierre et Marie Curie, 75252 Paris Cedex 05, France

Yoshiko Okubo, Thomas Durand, Gérard Dudognon, and Jean-François Miquel
Laboratoire Stratégie et Technologie, Ecole Centrale de Paris, 92295 Châtenay-Malabry, France

This study illustrates the application of a descriptive multivariate statistical method, Correspondence Factorial Analysis (CFA), to the analysis of a dataset of over 6 million bibliometric entries (data from ISI). CFA is used to show how the 48 most prolific nations stand in relation to each with regard to their publication interests in 17 specific disciplinary areas and one multidisciplinary field over the period 1981–1992. The output of a CFA is a map displaying proximity among variables (countries and disciplines) and constitutes an impartial working document for experts interested in the evaluation of science. The present study focuses on three aspects of a CFA: (1) The normalized "publication patterns" of countries with a common feature (e.g., that belong to the same geopolitical zone, economic union, etc.) can be pooled in order to highlight the position of the union with respect to individual countries; (2) complex CFA maps can be simplified by selecting reference countries or disciplines and observing how the remaining countries and disciplines relate to these references; (3) data on additional countries (new publication profiles) or on additional variables (e.g., socio-economic data on all the countries under study) can be introduced into the CFA maps used as mathematical models. Our CFA of the ISI dataset reveals the scientific interests of nations in relative terms. The main cleavage (the first factorial axis) is between countries that still concentrate on the disciplines of the industrial revolution such as physics and chemistry (or that have turned toward their offspring, materials sciences) and those that have veered toward more "modern" disciplines such as the life sciences (e.g., clinical medicine), the environment, and computer sciences. The second cleavage, along the second factorial axis, is between countries that focus on the agricultural sciences (the land surface) and those interested in the geosciences (the sea, earth's mantle, and mining). The third and fourth axes discriminate even further between earth, life, and abstract sciences highlighting the ostensible relationship between (organic) chemistry and all life science disciplines and between physics and disciplines related to engineering, materials sciences, etc. The CFA maps disclose the specific behavior of each country with respect to these cleavages.

* To whom correspondence should be addressed at 63 rue de Buffon, 75231 Paris Cedex 05, France.

Received December 13, 1994; revised May 18, 1995; accepted May 18, 1995.

1. Introduction

The Institute for Scientific Information (ISI) (Philadelphia, PA) has recently made available a databank, comprising more than 6 million entries, that gives the numbers of scientific publications (articles, proceedings, papers, notes, and reviews) in 17 specific disciplinary areas and one multidisciplinary field, which have been published by over 200 countries, geographical regions, and institutions in 3,200–3,500 upper-echelon journals over a 12-year period (1981–1992). The aim of the present study is to analyze the inherent structure of this databank with regard to the publication output of the 48 most prolific nations by using a descriptive cartographic factorial method, known as Correspondence Factor Analysis (CFA), that we shall explain in some detail. As elegantly outlined in the introduction of a recent study by Engelsman and van Raan (1994), a cartographic approach to structure analysis is fruitful because maps can aggregate data, are easier to grasp than tables, yield an instantaneous picture, appeal to visual memory, and can translate effects occurring over time into differences in location. High-performance cartographic methods, such as CFA, achieve appropriate data reduction, filter out the noise within the data set, and objectify correlations among variables.

By using CFA, we shall show how individual countries and disciplines are related, how a simple, comprehensive, and objective map of the dataset can be drawn, and how data on further variables such as population numbers (with reference to secondary education and employment in different economic sectors) can be introduced into CFA maps used as mathematical models.

2. Description of the Dataset

Our dataset corresponds to a 3D-system formed of the cube 18 (disciplines) × 48 (countries) × 12 (years). Several 2D-matrices within this cube could be analyzed but we shall focus here on the matrix 18 (disciplines) × 48 (countries) with the intention of deriving the main correlations between these two study fields over the entire 12-year time-span. The 17 specific disciplinary areas are listed in the insert to Figure 1. A multidisciplinary field (MUL), defined by ISI, covers publications in Nature, Science, and in the Proceedings of National Academy of Sciences that could not be assigned to any one of the above 17 areas. MUL, in fact, includes the national proceedings of several countries.

Although we shall restrict our analysis to the 48 countries or regions with the highest publication output (Fig. 1), we nevertheless cover over 97.9% (6,582,457 publications) of total world production. There is an estimated redundancy of about 9% due to the attribution of publications to the country of origin of each co-author and also of about 9% due to the allocation of journals and/or articles to more than one disciplinary area. This redundancy is minimized by the use of a multivariate method which extracts the most pertinent relationships within the dataset and is probably part of the background noise.

Figure 1 is the conventional description of such a databank. It gives the percent output per country in decreasing order (histogram) and the overall distribution among disciplines (insert). The US holds the leading position with a publication volume (2.3 million articles) that corresponds to about a third of the total and that is five to ten times greater than that of its closest peers (Japan, England, ex-USSR, ex-West Germany, France, and Canada). If the values for England, Scotland, and Wales are summed, the UK takes second place. If those of the US, UK, and Canada are summed, this North Atlantic trio is seen to produce more than half of the world output thus revealing either the preeminence of their scientific culture and/or language. Reuniting the two Germanies places them on a par with the ex-USSR. The most prolific country in the South Pacific is Australia (10th position), on the African continent, South Africa (25th) and in South America, Brazil (26th). The 48th country is Thailand which published 5,400 articles over the 12 year span, i.e., an annual average of 450 articles. The major discipline is clinical medicine (1.2 million articles), representing 19% of the total volume, followed by general biology and biochemistry, chemistry, and

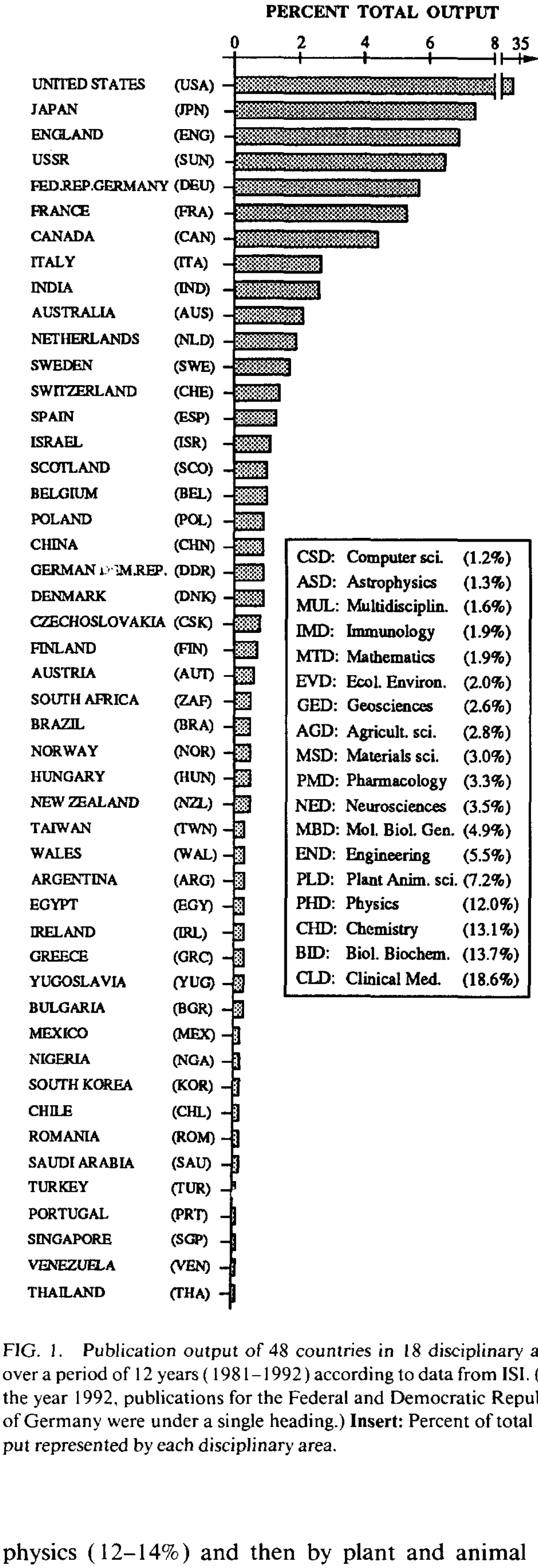


FIG. 1. Publication output of 48 countries in 18 disciplinary areas over a period of 12 years (1981–1992) according to data from ISI. (For the year 1992, publications for the Federal and Democratic Republics of Germany were under a single heading.) **Insert:** Percent of total output represented by each disciplinary area.

physics (12–14%) and then by plant and animal sciences, engineering, molecular biology, and neurosciences (3.5–7%). All other disciplines represent 3% or less

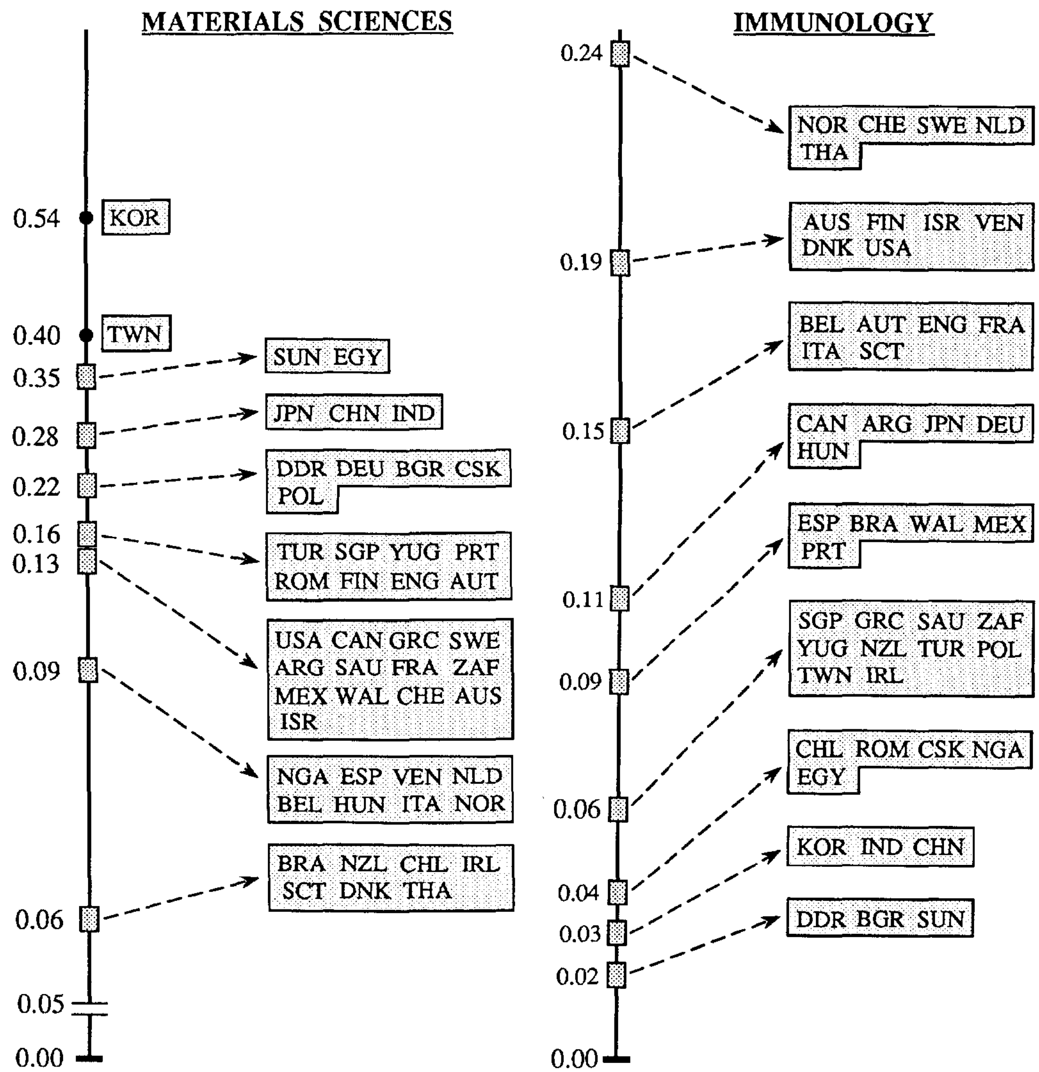


FIG. 2. Associations between the 48 countries and either materials sciences or immunology within the 48 × 18 multidimensional space.

of the total. The lowest outputs are recorded for emerging areas such as ecology/environmental sciences, immunology, and computer sciences (78,500 publications, i.e., 1.2%).

3. Conversion of the Dataset and the Correspondence Factor Analysis (CFA) Method

A conventional analysis of such a dataset is to calculate the relationships between country-discipline pairs (a total of 864 (18 × 48) pairs). The probable relation of each of the 48 countries with either materials sciences or immunology is plotted on the scales shown in Figure 2. The country that was most closely associated with materials sciences between 1981 and 1992 was South Korea. The one least interested in this discipline was Thailand. Immunology was a speciality of Norway but disregarded by the Soviet Union.

Because 18 scales like the ones in Figure 2 are needed to cover the entire information content of the dataset and because, even then, between-country and between-discipline relationships cannot be established in this way, we opted for a different approach and converted the data into a frequency table giving the probability of a statistical link between countries and disciplines using χ^2-metrics (see Appendix). The χ^2-distance is the canonical Euclidian distance between two absolute probability levels divided by a weighting factor in order to take into account the fact that a small distance between two small items can be as important as a large distance between two large items. This procedure obliterates the element of absolute size.

The most common descriptive factorial method for establishing relationships between quantitative variables is Principal Components Analysis (PCA) which uses a covariance matrix (based on Euclidian distances) for data reduction. PCA is not applicable to categorical variables. When dealing with counts (publications) rather than measurements, it is helpful to convert the data into χ^2-frequency distributions, as mentioned above, and then into a matrix of normalized profiles. These patterns of publication by the various countries in different disciplines (normalized profiles) (Miquel, Ojasoo, Okubo, Paul, & Doré, 1995) can then be compared by an abstract form of pattern recognition known as Correspondence Factor Analysis (CFA) (see Appendix for statistics).

CFA has been extensively studied by Benzécri's team (1973, 1982) and reviewed by Greenacre (1983, 1993), Lebart, Morineau, & Fénelon (1979), Lebart, Morineau, & Warwick (1984), Moser (1989), van der Heijden, de Falguerolles, & de Leeuw (1989), Jambu (1991), Pack & Jolliffe (1992) and others. Although initially a specialist area essentially confined to French scientists, it now has a broader following and has proven its worth in several branches of basic and applied science (Doré & Ojasoo, 1994, and references therein) as, for example, in ecology and environment studies (Gauch, 1982; Bobee & Lachance, 1984; Renou, Lalanne-Cassou, Michelot, Gordon, & Doré, 1988; Almendros, 1989; Devillers et al., 1991; Avila & Myers, 1991). It is also applied in social sciences (Lebart et al., 1984; Pirela, Rengifo, Mercado, & Arvanitis, 1993; Cortinovis, Vella, & Ndiku, 1993), marketing research (Hoffman & Franke, 1986), and scientometrics (Miquel & Doré, 1981; Karsky, Doré, & Miquel, 1988; Moed, 1989; Okubo, Miquel, Frigoletto, & Doré, 1992; Miquel & Okubo, 1994; Ojasoo, Doré, & Miquel, 1994). These references are but a small selection among many.

In the present study, we ran an in-house CFA program adapted for BASIC (Microsoft Language) from FORTRAN ANACOR software on a PC-AT compatible microcomputer. The outputs of our CFA program, which is more flexible than the commercial programs listed in the appendix, are (i) the mean marginal weights of the variables (countries and disciplines) and (ii) their distances from the center of gravity of the multidimensional system calculated before diagonalization of the symmetric matrix. After diagonalization, one obtains (iii) the distribution of the variance (τ) over the number of factorial axes required to describe the dataset (these axes are orthogonal and thus independent), (iv) the coordinates of all countries and disciplines to each factorial axis, and (v) their absolute (AC) and relative (RC) contributions to these axes.

Because CFA deals with a contingency or frequency table, its outcome is not a cut-and-dry answer but subject to the laws of probability. However, the method filters out redundancy and noise and thus highlights the most legitimate correlations among the variables. These correlations can be best seen on biplots of the factorial axes that describe ever-decreasing proportions of the total variance (information content) of the system under study. The first factorial map ($\varphi_1\varphi_2$) reveals the strongest correlations among the variables, the $\varphi_3\varphi_4$ map weaker but equally meaningful correlations, and so on. The coordinates (see iv above) are used to plot these maps and the contributions (see v) attribute meaning to these axes. The ACs of the variables describe how well a particular axis represents the variance of the system (ΣACs = 100%) and the RCs how a variable is dispersed across all the axes (ΣRCs = 1). A theoretical index (λ) describes the quality of the factorial representation. A maximum value of 1 for λ means that the two study fields (i.e., countries and disciplines) are totally unrelated. In essence, a CFA provides graphic overviews of correlations for use as working documents by panels of experts.

4. Comparison of Publication Patterns by CFA

By following the procedure broadly outlined above and in more detail in the Appendix, we obtained:

Normalized Publication Profiles

We calculated normalized publication profiles for each country by setting the overall production per country at 100. The discipline patterns for six countries are illustrated in Figure 3.

Marginal Weights

A country's mean marginal weight is its overall publication rate in all disciplinary areas. The marginal weight of a discipline is the overall publication rate of all countries. These weights in fact correspond to the histograms of Figure 1.

Distribution of the Variance over the Factorial Axes

There are 17 (n − 1) factorial axes in this study because there are 18 disciplines. Figure 4 shows the fraction of the total variance (τ) of the multidimensional system, i.e., the information content, that is accounted for by each axis. The first factorial axis embodies 52.8% of the variance, the second 15.1%, and so on. From the 5th axis onwards, the variance of each axis is less than 4%.

The $\varphi_1\varphi_2$ Correspondence Factorial Map

The projections of the disciplines and countries onto the two main factorial axes are given in the superimposable maps of Figures 5A and B. How does one read a factorial map? The degree of proximity between two variables (clustering) indicates the extent to which they are correlated (or alike). Neighboring countries have similar publication profiles; neighboring disciplines are a sign that countries consider these disciplinary areas in the same light. Vicinity between a country and a discipline means that the country publishes with a high degree of selectivity in that area but does not mean that the absolute level of publication is highest in that discipline

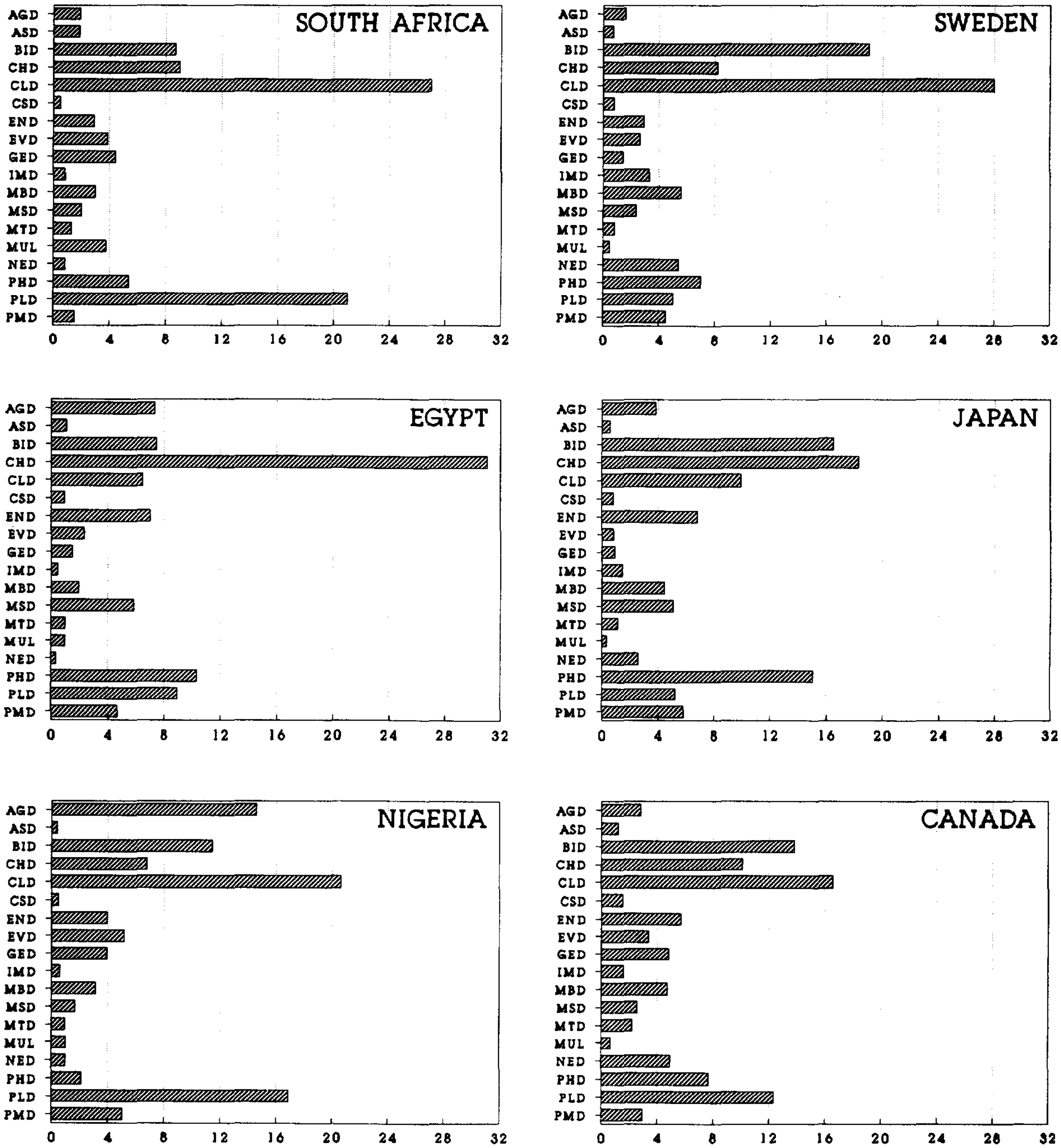


FIG. 3. Normalized profiles of the scientific output of six countries in all 18 disciplinary areas. The sum of each profile totals 100%.

because, as indicated above, CFA compares patterns and not absolute publication activity levels. Countries or disciplines near the origin of the axes are unrelated to the others (e.g., countries with a mean publication profile or countries that are described by lower factorial axes). Variables in diametrically opposed positions with respect to the origin of an axis are anticorrelated. An instance when proximity cannot be interpreted in quite such a clear-cut manner is when the variable is not really in or near the plane of the 2D-plot but well above or below. This can be checked by referring to the absolute (AC) and relative (RC) contributions of the variables to the factorial axes. It is not possible to list all contributions but some are given in brackets in the text.

The first factorial axis (φ_1), which embodies 52.8% of the information within the ISI databank, highlights its dominant feature which is an antithesis between two types of discipline, ancient and modern (left-hand vs. right-hand quadrants of panel 5A). Cluster A on the left groups together three closely related disciplines that deal with the constitution and function of matter. Chemistry (AC = 24%, RC = 0.78) and physics (AC = 16%, RC = 0.81) flourished during industrialization in the late 19th and early 20th century and have either phagocyted their

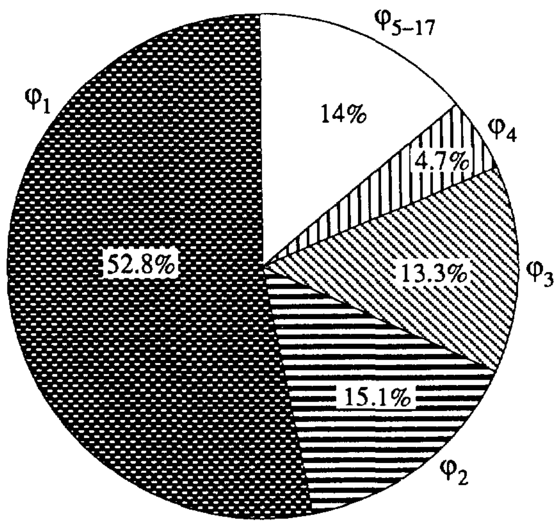


FIG. 4. Distribution of the variance over the factorial axes required to describe the dataset.

offspring, materials sciences (AC = 4.2%, RC = 0.60), or had not yet given it its full credentials in the 1980s. Cluster A lies opposite cluster D which englobes several modern life science disciplines (clinical medicine (AC = 14.3%, RC = 0.63), neurosciences (AC = 4.4%, RC = 0.74), and immunology (AC = 3.0%, RC = 0.72)) together with environment/ecology (AC = 2.4%, RC = 0.49) and computer sciences (AC = 1.0%, RC = 0.39). These are today's cutting edge disciplines and, as shown below, are the prerogative of the more advanced nations. The older biological disciplines (biochemistry/biological sciences (AC = 4.7%, RC = 0.58), plant and animal sciences (AC = 1.9%, RC = 0.14)) are nearer to the origin of the first factorial axis and close to other timeless disciplines of universal interest such as agricultural sciences, pharmacology, engineering, mathematics, astrophysics, and geosciences. These do not take part in the ancient versus modern bipolarity suggested by the first factorial axis. However, rather surprisingly, molecular biology/genetics (AC = 0.49%, RC = 0.29) is also close to the origin.

The second factorial axis (φ_2) describes 15.1% of the information and essentially opposes the agricultural sciences (AC = 12%, RC = 0.33) and chemistry (AC = 14%, RC = 0.14), which have positive coordinates (upper quadrants), with geosciences (AC = 4%, RC = 0.23) and clinical medicine (AC = 16%, RC = 0.20) with negative coordinates (lower quadrants). This implies that the countries that indulge in the former have a correspondingly lower interest or capacity in the latter. In other words, are energy resources (GED) and public health (CLD) the concerns of nations different from those which deal preferentially with the land (AGD) and transformation industries (CHD)?

A variable that partakes substantially in the formation of the φ_1 (AC = 23%, RC = 0.58) and φ_2 (AC = 45%, RC = 0.32) axes is the multidisciplinary field (MUL). MUL is not related to any other discipline but has a great impact on the publication patterns of several countries. The countries with the highest MUL activity (percent of total output) are Bulgaria (16.3%), Peoples' Republic of China (7.6%), ex-USSR (4.2%), Thailand (3.0%), India (2.9%), South Africa (2.3%), France (1.9%), Venezuela (1.8%), New-Zealand (1.5%), Ireland (1.1%), and

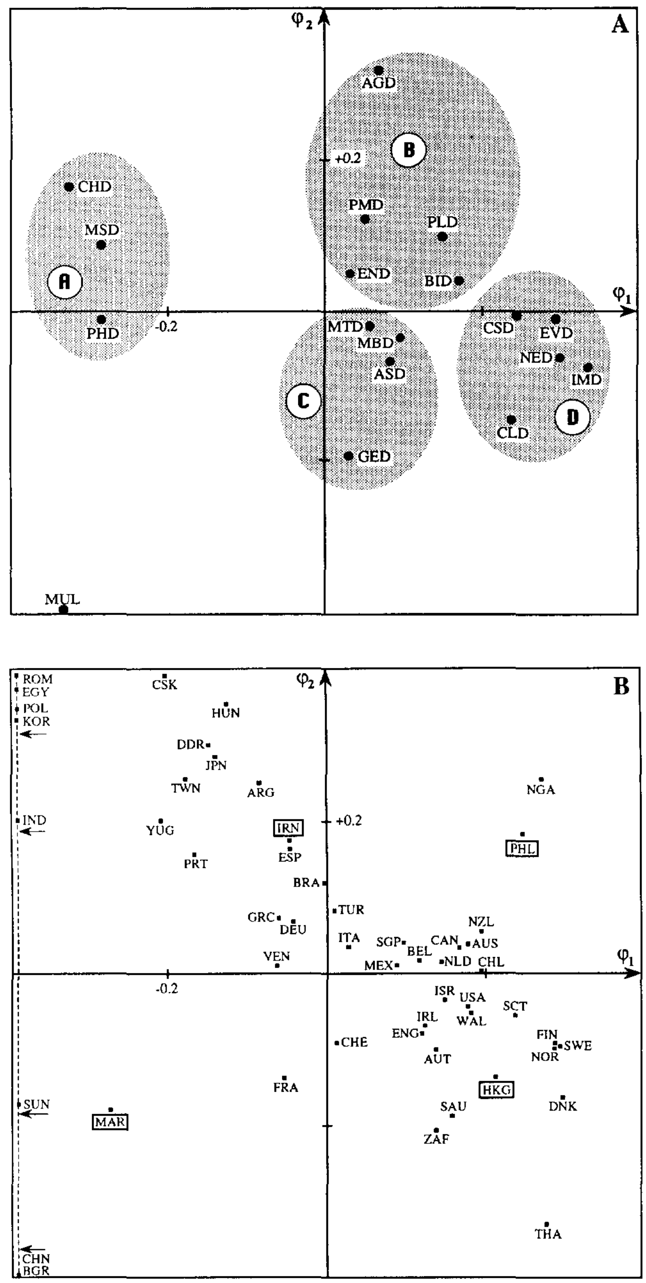


FIG. 5. $\varphi_1\varphi_2$ Correspondence factorial map of the disciplines (**A**) and countries (**B**). Panels A and B are totally superimposable and have only been separated for reasons of legibility. The values of λ and τ are, respectively, 8.5×10^{-2} and 52.8% for φ_1 and 2.4×10^{-2} and 15.1% for φ_2. In Panel A, the disciplines have been encircled into 4 major groups (A to D) on the basis of the results of an ascending hierarchical classification (as explained in the Appendix).

Saudi-Arabia (1.0%). The historical or political influence of the national academies of science of these countries has apparently had important repercussions on their national publication patterns over the period 1981 to 1992. Although the contribution of MUL is waning in most countries (a notable exception is China), we suggest that future CFA studies disregard this activity in order not to relegate important correlations to lower-order axes.

Which countries are responsible for the clustering of disciplines in Figure 5A? The answer lies in panel 5B which is directly superimposable upon panel 5A, thus disclosing the distance, i.e., correlation, between country and discipline. For example, the top left-hand quadrant of Figure 5B likens the publication profiles of Asian countries (South Korea, Taiwan, and Japan) with those of countries of the ex-East European bloc (Romania, Poland, ex-Czechoslovakia, Hungary, ex-East Germany, ex-Yugoslavia). The Asian countries have a high preferential interest in chemistry, physics, and also materials sciences in the case of South Korea; many Eastern European countries are renowned for their chemical expertise. Both Asian and East-European countries are characterized by comparatively little activity in the opposing environmental and life science disciplines. This explains their location in this quadrant.

A brief exposé should further accredit the superposition of the two maps (The Economist, 1991a, 1991b). One in every four candidates in electrical engineering at American Universities is from Taiwan. Korean conglomerates have led successful incursions into steel, semiconductors, petrochemicals, and more recently into the aerospace field. Apparently, neither country has concentrated on computer sciences although a time-analysis comparing 1981 and 1992 publication rates in the field suggests that their participation is growing fast. Computer sciences has remained the domain of the older industrialized nations, in particular the USA (see right-hand quadrant), which has tapped its youth's creativity enabling it to reconcile a new-generation lifestyle and garage ventures with the existing business world. Cleaning up the pollution generated by their technologies is apparently not yet the concern of these growth-addicted societies but that of the long-industrialized nations. Neither is health care yet a part of their consciousness although it will undoubtedly become a demand as the technological boom continues and life expectancy rises. Japan (see normalized profile in Fig. 3) is situated far right of Korea and Taiwan having entered the international scene earlier. Since the 1980s it has sought active collaboration with the Western world in order to become competitive in the life sciences but without ever conceding the longstanding priority it has accorded to engineering and technology. Not unsurprisingly, from among the European countries, it has the greatest similitude with Germany. Much of Asia's industrialization does depend upon the creation of an agricultural surplus created by rich industrious farmers subsequent to serious land reforms (see upward vertical displacement of these countries within the map).

Unlike the developing Asian countries, the East European countries are already highly industrialized, though their polluted landscape does not seem to have inspired activity in ecology/environment. The industries in which they are most competitive are steel, food, and textiles, an observation in complete coherence with their avowed research activities as measured through publication profiles. Poland, the most eccentric East European country together with Romania, has the highest relative involvement in chemistry. Its basic industries are the chemical, cement, and non-ferrous metallurgy industries. Hungary also has an active iron and steel industry but, like ex-East Germany and like countries with an established pharmaceuticals industry, it also publishes in the biochemical and clinical fields. This accounts for the greater proximity of Hungary and DDR to the life sciences compared to Poland. Nevertheless, they are still situated wide off the mark compared to the most advanced nations.

The top right-hand quadrant of the $\varphi_1\varphi_2$ map, which principally reflects publication profiles that combine interest in the agricultural sciences with plant and animal sciences, is dominated by a single country, Nigeria (see profile in Fig. 3). Nigeria's closest neighbors, albeit at a considerable distance, are New Zealand and Brazil. Their positions lower down the φ_2 axis are due to an additional concern for biochemistry and clinical medicine. Brazil also devotes considerable effort to publication in the field of physics as seen by the displacement toward the left (φ_1 axis) compared to New Zealand. The highly eccentric positions of Bulgaria, the People's Republic of China, and of the Soviet Union in the lower left-hand quadrant are in total accord with their very high multidisciplinary activity already mentioned above. The only other country situated in this quadrant is France which has an illustrious academy of sciences which lays great emphasis on national publication. Most countries are situated in the right-hand quadrants in a pool centered around the φ_1 axis and publish in the majority of disciplines in a way which is relatively homogeneous compared to the world average. Certain preferences can, however, be highlighted and to do this, we shall use a technique known as the creation of barycenters.

Creation of Barycenters and Introduction of Data on Additional Countries

The symmetric treatment of matrix rows and columns (distributional equivalence) in a CFA means that countries and disciplines can be represented on the same or on superimposable factorial maps. It also means that selected rows and columns of the χ^2-distance matrix can be combined (e.g., countries belonging to a geographic area or economic union) and the mean normalized pub-

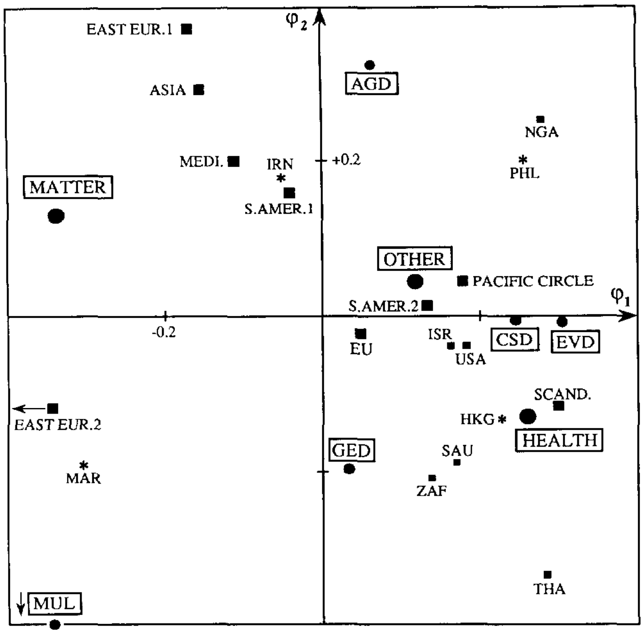


FIG. 6. Selected centers of gravity (barycenters) representing either groups of disciplines or countries that have been calculated by combining the normalized profiles of these disciplines and countries within Figure 5. Introduction of the normalized profiles of four additional countries represented by stars (Iran, Philippines, Morocco, and Hong Kong). *Discipline barycenters:* MATTER = chemistry, materials science, physics; HEALTH = Clinical medicine, immunology, neuroscience; OTHER = disciplines closest to the origin, namely, pharmacology, plant and animal sciences, biology/biochemistry, mathematics, molecular biology and genetics, astrophysics. *Country barycenters:* East Europe 1 = DDR, HUN, CSK; East Europe 2 = BGR, SUN, POL, ROM; Asia = TWN, JPN, KOR; Mediterranean = PRT, ESP, YUG, GRC, TUR, EGY; South America 1 = BRA, ARG, VEN; South America 2 = CHL, MEX; Pacific circle = AUS, NZL, CAN; UK = ENG, SCO, WAL; EU = UK, DEU, FRA, ITA, NLD, ESP, BEL, DNK, GRC, IRL, PRT; Scandinavia = SWE, NOR, FIN, DNK.

lication profile of the combination introduced into the factorial map used as a mathematical model. This creates centers of gravity ("barycenters") within the factorial maps. (The RC of a barycenters equals 1 since the information is preexistent in the map).

Figure 6 is a replica of Figures 5A and B in which disciplines relating to matter, to health, and other disciplines have been grouped into barycenters and in which countries with either common geographic, economic, historical, or political characteristics have also been combined (see legend to figure). The barycenter of countries bordering the Mediterranean (excluding Italy but including Portugal) is situated between the east and the far west (South America), within the zone of influence of the agricultural sciences, and relatively far off from Western Europe. This situation seems to reflect the historical influences of Spanish and Portuguese culture in South America, a similar regard for the land but possibly for different reasons (difficult biotopes of the Mediterranean vs. the vast expanses of South America), and differences in economic development. The European Economic Union (excluding Luxembourg which is not among the 48 nations with the highest publication output) lies virtually at the origin of the axes (lowest variance) probably because it unites a mixed bag of countries with different cultures and priorities. It is some distance away from the trio—US, UK, and Israel—(see also Fig. 5B) that seems to be the epicenter of the advanced nations. Around this epicenter, one notes countries of the Pacific circle (AUS, NZL, CAN) that are offset toward agriculture, two Central and South American countries (Chile and Mexico), and the Scandinavians whose focus is manifestly on health. As might be expected, South Africa and Saudi Arabia, countries rich in mining and oil resources, respectively, lie near the geosciences pole. An odd-man-out is Thailand which, curiously, is closer to Europe, and in particular Scandinavia, than to its Asian neighbors.

The normalized publication profiles of additional countries can be introduced into a CFA map as illustrated in Figure 6 for four countries with a low publication output, namely, Hong Kong (RC = 0.76), Morocco (RC = 0.22), Iran (RC = 0.16), and the Philippines (RC = 0.035). Each falls within a different quadrant of the map. The RCs are low and indicate that the major part of the information on all these countries except Hong Kong, resides within the lower axes. The lower axes explain their eccentricities in behavior that arise principally from absence of activity in several disciplines whereas their position in the $\varphi_1\varphi_2$ map highlights their strong points. Over the 1981 to 1992 period, Iran published in chemistry and physics but neither in agriculture nor in the disciplines of the right-hand quadrants. The Philippines favored agricultural sciences. Hong Kong manifestly shared the same interests as the UK and other highly-advanced nations. Morocco lies not far from France. Its scientists are known to publish extensively in the Proceedings of the French National Academy of Science. Publication in multidisciplinary journals was 7.6% over the study period.

The $\varphi_3\varphi_4$ Correspondence Factorial Map

The lower-order $\varphi_3\varphi_4$ map (Fig. 7) accounts for 18% of the variance (13.3% for φ_3; 4.7% for φ_4). The multidisciplinary field still partakes in the formation of φ_3 (AC = 11%, RC = 0.07) but not of φ_4. This map opposes along the φ_3 axis those countries that publish preferentially in disciplines related to the land (mainly agricultural (AC = 17%) and plant/animal sciences (AC = 39%) but also ecology/environment (AC = 6.5%) and geosciences (AC = 6%)) with those that publish more extensively in physics (AC = 8%) and/or clinical medicine (AC = 5%). Along the φ_4 axis, it opposes a hard core of related "high-tech" disciplines (computer sciences (AC = 11%), engineering (AC = 45%), and materials sciences (AC = 7%)) with chemistry (AC = 11%) and clinical medicine (AC = 11%). These antitheses are not

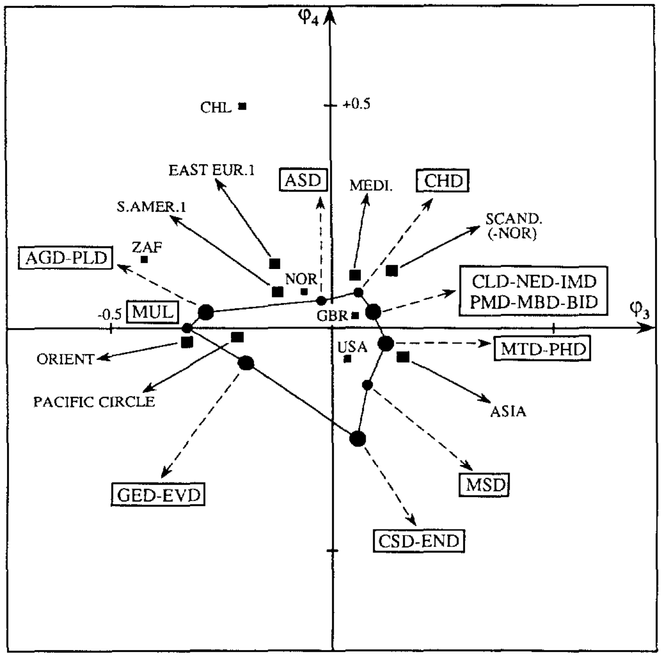


FIG. 7. $\varphi_3\varphi_4$ Correspondence factorial map showing country and discipline barycenters. The disciplines were combined into barycenters on the basis of their location within the $\varphi_3\varphi_4$ map. The country barycenters were the same as those in Figure 6 except that Norway was not included with the other Scandinavian countries because of its atypical behavior in this map. A further barycenter Orient was formed of EGY, SAU, IND.

unexpected. Who in the 1980s would not have associated chemistry (especially organic chemistry) with the life sciences and physics with high-tech? Thus, in rather simplistic terms, whereas the $\varphi_1\varphi_2$ map contrasts the ancient and modern disciplines (φ_1), which is the most important (anti)correlation within the dataset, and agriculture with mining and medicine (φ_2), which is a corollary, the $\varphi_3\varphi_4$ map distinguishes between the earth, life, and abstract sciences.

A specific interest in disciplines to do with the land is a dominant feature of advanced nations with considerable natural resources or specific flora and fauna, such as the countries of the Pacific circle and South Africa (see profiles in Fig. 3) but also of the UK regions (Scotland and Wales) (Ojasoo et al., 1994). They are also a major preoccupation of poor countries with desert lands as revealed by the barycenter of India, Egypt, and Saudi Arabia (denoted Orient) which juxtaposes this pole. During the 1981 to 1992 period, neither Eastern Europe nor South America forfeited the land for advanced technologies. In this lower-order map, the Scandinavian countries reaffirm their interest in clinical medicine but Norway adopts a specific stance, visibly giving priority to earth sciences because of its more extensive and varied natural resources (crude oil, natural gas, fisheries, and paper). Japan and the Asian "dragons" still favor high-tech. The aspect of physics disclosed by this map is no longer that of a traditional discipline apparent in the $\varphi_1\varphi_2$ map but that part of the discipline that is correlated with modern high-tech science and that is opposed to chemistry. The outlying country with respect to the astrophysics pole is Chile. Although at first sight this may appear to be an unusual correlation, one should not forget the presence of the ESO telescope at La Silla and of the US Cerro-Tololo telescope at La Serena which are probably at the origin of many of Chile's publications in collaboration with other countries. This correlation is even clearer on the φ_9 axis which accounts for a third of the variance of Chile and which is explained to 44% by astrophysics.

5. Subpopulation (Triad) Studies

The above CFA maps give an overview of the global structure of the ISI dataset. Often, however, one wishes to focus on a set of closely-related disciplines or countries. Simplified CFA maps representing 100% of the variance can be created by selecting three rows or columns as reference variables. (In a CFA, 100% of the information is contained in ($n - 1$) factorial axes, i.e., for three variables, in a 2D-plot). The panels of Figures 8A to D indicate the positions of the countries with respect to three disciplines (engineering, physics, and materials sciences in panel 8A; pharmacology, immunology, and molecular biology in panel 8B). Panels 8C and 8D illustrate the relative attitudes of three countries in Africa (Egypt, South Africa, and Nigeria) and of three highly prolific nations (Japan, Canada, and Sweden (see Fig. 1)) toward publication in all disciplines. The publication patterns of the six countries are given in Figure 3. As a visual aid, we have linked the countries or disciplines into a shortest-distance network which was calculated by applying an algorithm for a minimum spanning tree (Prim, 1957) to the χ^2-distance semi-square matrix.

Figure 8A indicates that the activity of most countries in either physics or engineering over the 12-year span largely predominated over their effort in materials sciences. The relative emphasis on materials sciences was greatest in the case of Egypt, substantial in the case of South Korea, ex-East Germany, and Czechoslovakia, and non-negligible in the case of Finland, Sweden, South Africa, India, and Japan. The central position of West Germany shows that it has published equally in all three disciplines. The most traditional biological discipline, pharmacology (Fig. 8B), is correlated with the newly-industrialized nations whereas the US and several countries of the EU show no preference. Immunology is a specific attraction for the Norwegians, the Australians, the Swiss, and also the Venezuelians and the Thai. However, the interest of the Thai seems to stem from an approach closer to pharmacology, whereas the Venezuelians and the Israeli have taken an apparently more molecular biology/genetics stance. The outlying positions of Czechoslovakia, Portugal, and Brazil reflect a general

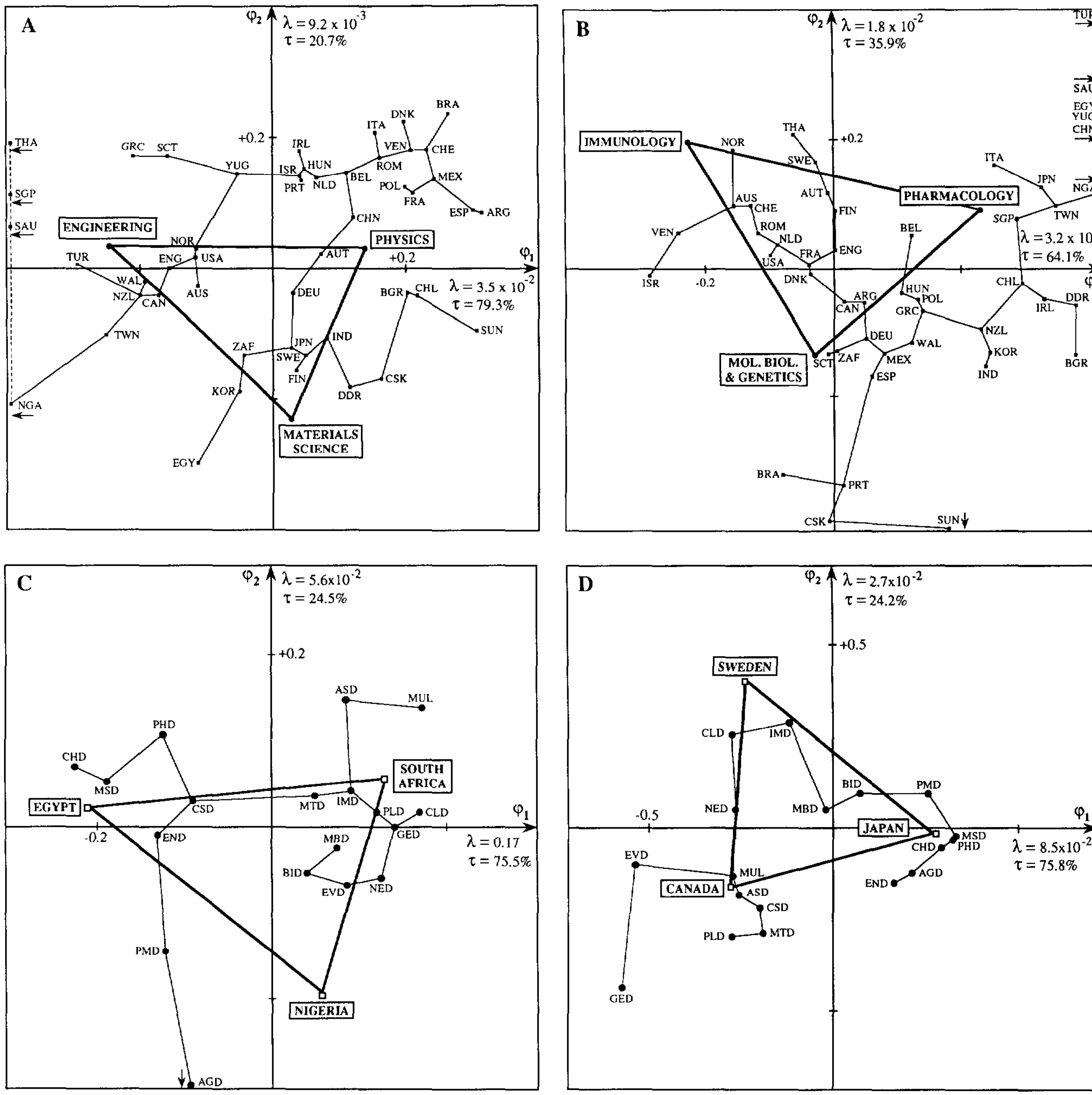


FIG. 8. **A, B:** Triads of disciplines (engineering, physics, materials sciences; immunology, pharmacology, molecular biology-genetics). **C,D:** Triads of countries (Egypt, South Africa, Nigeria and Canada, Sweden, Japan). Minimum spanning trees illustrate the shortest-route network among countries and disciplines (as explained in the Appendix).

lack of publications in immunology and pharmacology rather than a highly-developed interest in molecular biology.

The plots comparing the publication patterns of the countries are self-explanatory. On the African continent (Fig. 8C), South Africa publishes in the largest variety of disciplines although with a decided emphasis on the life sciences compared to Egypt and Nigeria. Nigeria favors the agricultural sciences. Plant and animal sciences is a discipline more specifically associated with Egypt and Nigeria than with South Africa although, as noted above, this discipline separates these three countries from much of the rest of the world. The chemical sciences are the prime domain of Egypt. Thus, to caricature, one could say that South Africa represents the life sciences, Egypt the chemical sciences, and Nigeria the land. It would be tempting to relate this different focus to the degree of industrial development of these nations.

Each of the three industrialized nations, Sweden, Canada, and Japan (Fig. 8D), which are among the 12 most prolific nations of the world as regards publication in all 18 disciplines (see Fig. 1), has its own specialities.

Canada focuses on the geosciences, environment, plant and animal sciences, as expected of a country with such a large surface area and considerable resources, but also on computer sciences possibly as a result of a mathematical bent. Sweden definitely favors the life sciences, namely, clinical medicine, immunology, but also neurosciences, a discipline that also attracts Canada. These disciplines are not Japan's choice although molecular biology and biochemistry are interests it shares with both Sweden and Canada (at the center of the plot). Compared to the two other countries, Japan publishes preferentially in the physical and agricultural sciences. In this instance, it is tempting to relate the different focus of these nations to a caricature of their culture. The pragmatic "peasant" outlook of Canada, the social "humanitarian" outlook of Sweden, and the "high-tech" stance of Japan.

6. Correlations over Time

The variations in the relationships among countries and disciplines year-by-year during the 12-year time span merit an in-depth analysis. This is, however, beyond the scope of the present study which aims to give an overview of the application of CFA to bibliometric studies. Of the three variables—country, discipline, and time—time is the least influential. The structure of the dataset is relatively stable over time especially with regard to the hard core of highly-developed countries (not shown). The greatest variations are undergone by the outlying countries within the $\varphi_1\varphi_2$ map and can be either quite erratic or a gradual move toward the more advanced nations as observed for Hungary and Nigeria.

7. Correlations with Socio-Economic Data

Using CFA to structure a dataset has a singular advantage. CFA maps can be used as mathematical models in order to discover how other variables, not part of the initial analysis, might be related to the dataset. Since a link between publication profiles and socio-economic data is plausible, we introduced into our CFA information on the percent population with secondary education within a country and on the percent workforce within each economic sector (agriculture, mining, industry, and services) of a country (data from Atlaseco, 1987). The countries were grouped into very broad categories on these bases. Figure 9 shows that the φ_1 axis that was created by publication profiles distinguishes countries with high employment in the services (>60%) (a feature of highly-industrialized nations) from those with a large workforce employed in other sectors. The φ_2 axis distinguishes countries with above 30% of their workforce in agriculture from countries with above 6% in mining. As the university-educated population increases, there is a move away from agriculture toward mining/industry and then toward the services. These crude socio-economic data are thus in agreement with the publication data.

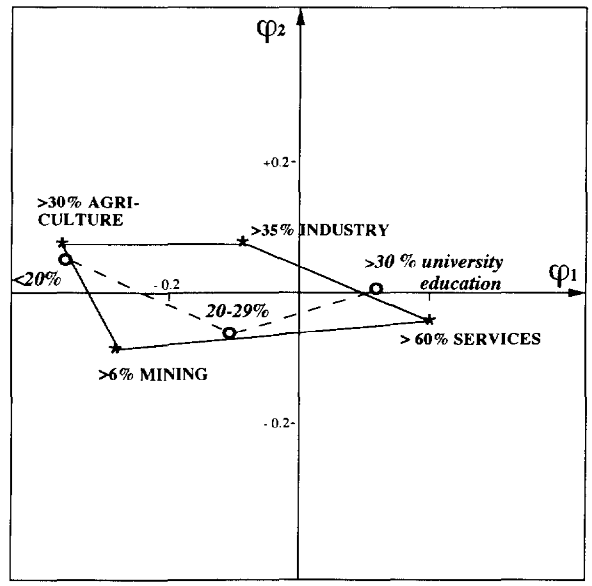


FIG. 9. Introduction of socio-economic data into the CFA map of Figure 5 used as a mathematical model. Locations of all countries (O) with less than 20%, from 20–29%, and above 30% of the population with university education. Locations of countries (*) with above 30% of the workforce in agriculture, above 6% in mining, above 35% in industry, and above 60% in service industries. (Data for 1986 from Atlaseco.)

8. Conclusions

There have been many approaches to bibliometric analyses (Garfield, 1972; Small, 1973; Carpenter & Narin, 1981; Moed, Burger, Frankfort, & van Raan, 1983; Chubin, 1986; Callon, Law, & Rip, 1986; Schubert, Glänzel, & Braun, 1989, and others). The above study has shown how a powerful multivariate method, CFA, can be used to translate a large dataset into a series of increasingly simple maps. CFA deals with the sheer volume of data by reducing the influence of redundant information, artefacts, and background noise and incurs no loss in information because it does not rely on averaging. The maps obtained highlight the correlations among the variables and the governing trends within the dataset. CFA can be combined with other multivariate methods such as dendrograms and hierarchical classifications calculated from the χ^2-distance semi-matrix used as a basis for the CFA (see Appendix). A multiplicity of statistical methods always adds strength to the conclusions that are drawn. Furthermore, because new variables can be introduced into a CFA used as a mathematical model, CFA is an ideal interface between fields of knowledge that sometimes have difficulty in finding a common ground. For instance, we have illustrated how basic economic indicators can be introduced quite licitly into a bibliometric analysis. Future studies will refine the time-dependence of the model factorial maps and also consider the introduction of more sophisticated economic indicators with the aim of ultimately developing an effective strategic tool for decision-making on the basis of bibliometric analyses.

Appendix on Statistical Methods

Tables of the distribution of a population of items (e.g., publications) between two sets of discontinuous variates (countries, disciplines) are termed either contingency matrices (no overlap among classes), frequency matrices (some overlap), or binary tables (0/1 data). When the data are exhaustive and dependent, the table constitutes a SYSTEM of correlations between two features and embodies the STRUCTURE or ORGANIZATION of the system. We shall use the accompanying example (Fig. 10) to show different ways of analyzing such a table to derive correlations and reveal the structure. The example deals with 1,565 publications distributed among five countries (1 to 5) and three disciplines (A to C). Each cell of the matrix Kij gives the number of papers published by country (i) in discipline (j) (e.g., $K_{1A} = 500$).

a) The Chi²-Test

The usual way of analyzing tables of counts is to determine the probability of a global association between the rows and columns. The reality of the association is tested by using the χ^2-test which essentially finds out whether the observed frequencies in a distribution differ significantly from the frequencies which might be expected according to some assumed hypothesis. This is illustrated in the left-hand column of the example. According to the χ^2 formula, the matrix cell K_{1A} has a partial χ^2-contribution of 40.32 corresponding to the discrepancy between the expected value if row and column were independent (376.74) and the observed value (500). The partial contributions of the cells yield, as illustrated, a χ^2-distribution which enables the establishment of the percent deviant items (1%, 5% etc.) and of the association α between rows and columns for a cut-off level x. The sum of the partial χ^2-contributions gives a χ^2-value of 420.87 for our system with 8 degrees of freedom (df). This is considerably greater than the expected values of 15.5 and 20.1 for 5 and 1% probability levels, respectively, and indicates a strong association between countries and disciplines.

This global analysis does not reveal which are the significant individual associations (i.e., $i \times j$ pairs) and can be pursued by transforming the data table, as shown, into normalized frequency profiles for each discipline (on the left) and each country (on the right). (The sum total of each distribution equals 100%.) This enables the analysis of $i + j$ histograms. However, this approach has several disadvantages: (i) The sheer number of histograms for very large data tables; (ii) a prejudicial bias when the sums of the rows and columns are very different. For instance, the distribution profile of country 1 attributes greater weight to discipline A than disciplines B and C because the sum total of column A (880) far exceeds that of columns B (310) or C (375); (iii) it provides a series of bivariate correlations but no comprehensive nor stratified understanding of the structure of the system described by the data set.

b) Correspondence Factor Analysis (CFA)

A more appropriate approach is to perform a multivariate analysis and use a method for the reduction of dimensionality in order to highlight the main structure of the system and individual correlations. The factorial method in common use is Principal Components Analysis (PCA). PCA uses covariance (Euclidian metrics) for data reduction and is therefore applicable to the analysis of measurements but not of discontinuous variates (e.g., numbers of publications). Correspondence factor analysis (CFA), on the other hand, uses χ^2-metrics and can analyze frequencies (see center column of example). Like PCA, CFA establishes the optimal projections of the multidimensional system onto a series of ranked orthogonal factorial axes describing ever-decreasing proportions of the variance (information content) of the system. Moreover, it enables the simultaneous projection of both variables (countries and disciplines) on the same axes or on common factorial plots.

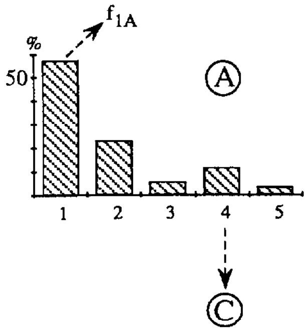
In practice, in a CFA, the (j) disciplines are projected into the multi-dimensional space (R^i) formed by the (i) countries and vice-versa the (i) countries are projected into the space formed by the (j) disciplines (R^j). The position of a point within the R^i space is given by the probability that, for country (i), discipline (j) has a publication volume of f_{ij}. This probability is defined by the ratio $f_{ij}/f_{.j}$ where $f_{ij} = k_{ij}/\sum_{ij}k_{ij}$ and where $f_{.j} = \sum_j f_{ij}$. A symmetrical calculation gives the position of points ($f_{ij}/f_{i.}$) within the R^j space. The χ^2-distances between pairs of disciplines within the cloud of points are set out as a symmetrical semi-square probability matrix $\{S_{jj'}\}$.

$$R = \{S_{jj'}\} = \left[\sum_{i=1}^{n} \frac{1}{f_{i.}} \frac{f_{ij}f_{ij'}}{\sqrt{f_{.j}f_{.j'}}}\right]$$

Diagonalization of this matrix to obtain eigenvalues (λ) and eigenvectors (V_x) by solving equations of the type $[R] - \lambda[x] = 0$ and $[R][V_x] = \lambda_x[V_x]$ provides orthogonal principal projection axes, i.e., factorial axes (φ_α). The coordinates of the disciplines for these axes are calculated by the formula $\varphi_{\alpha j} = \lambda_\alpha^{1/2}V_{\alpha j}/f_j^{1/2}$ where $\lambda_\alpha^{1/2}$ is the square root of the non-trivial eigenvalue λ_α, $V_{\alpha j}$ the corresponding eigenvector, and $f_j^{1/2}$ the square root of the marginal relative frequency between discipline (j) and

MULTIVARIATE ANALYSES OF THE RELATIONSHIPS BETWEEN COUNTRIES AND DISCIPLINES

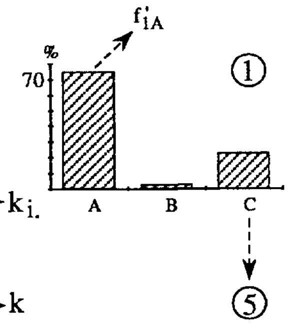
A theoretical example of 5 countries (1-5) and 3 disciplines (A-C)



DATA TABLE

$k_{ij} = k_{1A}$

Country (i)	Discipline (j) A	B	C	$k_{i.}$
1	500	20	150	670
2	200	100	60	360
3	50	150	80	280
4	100	30	80	210
5	30	10	5	45
$k_{.j}$	880	310	375	1565 → k



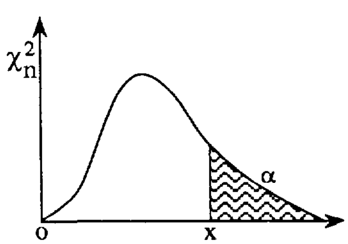
χ^2-DISTRIBUTION

Partial χ^2-contribution

40.32	95.70	0.69
0.03	11.50	7.99
73.30	161.13	2.48
2.76	3.23	17.50
0.87	0.13	3.10

$$\chi^2_{ij} = \frac{(k_{ij} - k'_{ij})^2}{k'_{ij}}$$

where k_{ij} observed value and k'_{ij} expected value



For 8 df, where df = (i-1)(j-1):

$\alpha = 0.05 \rightarrow \chi^2 = 15.5$

$\alpha = 0.01 \rightarrow \chi^2 = 20.1$

$\sum_{i,j} \chi^2 = 420.87$

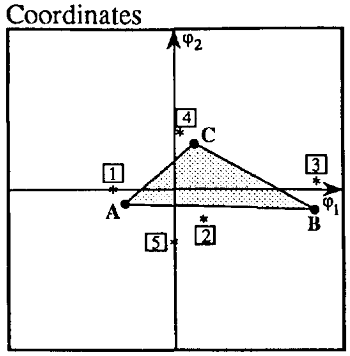
CORRESPONDENCE FACTOR ANALYSIS

$\{S_{jj'}\}$ Probability matrix

0.637	0.226	0.349
0.226	0.372	0.224
0.349	0.224	0.259

Diagonalization

Coordinates



$\tau_1 = 90.5\%, \ \lambda_1 = 0.243$

$\tau_2 = 9.5\%, \ \lambda_2 = 0.025$

Absolute contributions (AC)

	φ_1	φ_2
A	29.27	14.49
B	70.28	9.90
C	0.44	75.59
Σ	100%	100%

Relative contributions (RC) ($\cos^2\phi$)

	φ_1	φ_2	Σ
A	0.95	0.05	1
B	0.98	0.02	1
C	0.05	0.95	1

CLUSTERING METHODS

$d_{jj'}$

	A	B	C
A	0		
B	1.285	0	
C	0.558	0.949	0

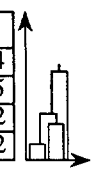
$d_{ii'}$

	1	2	3	4	5
1	0				
2	0.623	0			
3	1.371	0.804	0		
4	0.545	0.542	0.987	0	
5	0.501	0.224	1.023	0.632	0

ASCENDING HIERARCHICAL CLASSIFICATION

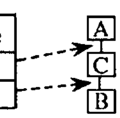
Node	Disciplines	υ
1	C+A	0.558
2	(C+A)+B	1.257

Node	Countries	υ
1	5+2	0.224
2	4+1	0.545
3	(5+2)+(4+1)	0.692
4	(5+2+4+1)+3	1.342



MINIMUM SPANNING TREE

	Link	Distance
1	A-C	0.558
2	C-B	0.948



1	1-5	0.501
2	5-2	0.224
3	2-4	0.542
4	2-3	0.804

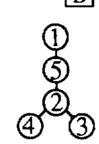


FIG. 10. Example illustrating the difference between the multivariate statistical methods used in this publication and a traditional χ^2-test.

the (i) countries. The correspondence between the countries and disciplines is given by the transition formulae:

$$\varphi_{\alpha i} = (1/\lambda_\alpha^{1/2}) \sum_{j=1}^{P} (f_{ij}/f_{i.})\varphi_{\alpha j} \quad \text{for the countries}$$

$$\varphi_{\alpha j} = (1/\lambda_\alpha^{1/2}) \sum_{i=1}^{n} (f_{ij}/f_{.j})\varphi_{\alpha i} \quad \text{for the disciplines.}$$

The factorial axes are ranked by their order of importance in accounting for the total variance (τ) of the system ($\varphi_1, \varphi_2, \varphi_3 \cdots \varphi_{n-1}$) (n − 1 because the first latent root is not significant) and factorial plots displaying the projections of the points are drawn highlighting the organization of the system.

In our example, the first factorial axis φ_1 embodies 90.5% of the variance of the table and the second factorial axis φ_2 the remaining 9.5%. There are only two axes because the table has just 3 dimensions (n = 3). The values of λ (0.243 and 0.025) are sufficiently different from 1 to indicate that the two sets of variables are not totally distinct and sufficiently different from 0 to indicate that the association is not random. The CFA plot reveals the relationships among the disciplines, the countries, and between countries and disciplines. In particular, it highlights the close association between country 1 and discipline A, country 4 and discipline C, and country 3 and discipline B. It also reveals that countries 2 and 5 favor discipline A over B at the expense of C.

A CFA also calculates the absolute (AC) and relative (RC) contributions of each variable to all factorial axes in order to assess how well a particular axis represents the variance of the system (ACs) and how a variable is dispersed across all the axes (RCs). For discipline j, $AC\alpha(j) = f.\varphi_{\alpha i}^2/\lambda_\alpha 100$ (ΣACs = 100% for any axis α) and $RC\alpha(i) = \varphi_{\alpha i}^2/d_p^2(i, G)$ (ΣRCs of each variable to all axes = 1) where G is the distance from the center of gravity of the points. RC is in fact the square of the cosine of the projection of discipline j onto axis α. In our example, the ACs for the first factorial axis show that discipline C (AC = 0.44) contributes very little to this axis which reflects to 99.5% an anticorrelation between disciplines A (AC = 29.27) and B (AC = 70.28). The relationship is an anticorrelation because the coordinates of the projections of these two disciplines onto this axis are of opposite sign. On the other hand, the second factorial axis reveals an anticorrelation between C and the two other disciplines. The RCs indicate that for information on disciplines A and B, one should refer to the first factorial axis whereas for information on discipline C, one should refer to the second. ACs and RCs for the countries are not shown.

Several textbooks explain how to write CFA programs (Lebart et al., 1979; Foucard, 1982; Greenacre, 1993). Standard CFA programs are available in France (from ITCF (Institut Technique des Céreales et des Fourrages, 8 ave du Président Wilson, 75116 Paris)), slp Statistiques ((Program STATlab) 51-59 rue Ledru Rollin, 94853 Ivry-s-Seine), DP Tool Club (Program ADSO (Dubus, 1992), BP 745 59657 Villeneuve d'Ascq), but also in other countries (Kovach Computing Service, Wales, UK; Professor M. J. Greenacre (SimCa version 2), P.O. Box 567, Irene, 1675 South Africa; BMDP Statistical Software Inc (PC-90 User's Guide 1990), Los Angeles, CA; SPSS Inc (Categories Reference Guide 1990), Chicago; SAS Institute Inc (SAS/STAT User's Guide, Vol 1: ANOVA-FREQ, Version 6, 1990, Cary, NC). The programs can be run on mainframe and personal computers as well as on UNIX workstations.

c) Automatic Clustering Methods

As shown in the right-hand column of our example, χ^2-metrics can also be used to obtain, by a simple iterative calculation starting from the initial k_{ij} frequency table, matrices that give the distances between pairs of disciplines ($d_{jj'}$) or pairs of countries ($d_{ii'}$) in the multidimensional space. For instance, discipline A is nearer to discipline C (0.558) than to discipline B (1.285). The diagonal is 0 since it corresponds to the distance of an item from itself.

$$\delta^2(i, i') = \sum_{j=1}^{3} \left[\frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}}\right)\right]^2 \quad \text{for the countries}$$

$$\delta^2(j, j') = \sum_{i=1}^{5} \left[\frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}} - \frac{f_{i'j}}{f_{.j'}}\right)\right]^2 \quad \text{for the disciplines.}$$

Two different types of clustering algorithm have been applied to these χ^2-distance matrices. The algorithm for an ascending hierarchical classification uses the aggregation criterion of Lance and Williams (1966) with standard coefficients of $\alpha = 0.625$ and $\beta = -0.25$ and yields hierarchical trees in which correlated variables are clustered beneath interconnected nodes of different heights (ν). In brief, the two closest countries (or disciplines) are united into a single group and the dissimilarity of this newly-formed group with each of the other countries is calculated. The two closest countries or groups of countries are again united and the process is iterated a total of i − 1 times. The resultant hierarchical tree is a sequence of partitions (see example). The algorithm for a minimum spanning tree, initially devised by Prim (1957) at the Bell telephone company for laying telephone cables, provides the shortest distance network, with no loops nor backtracking, that links the variables (disciplines or countries) within the system (see example). Neither of these clustering methods is detailed in the present article but the results of the cluster analyses have been used to encircle or link together variables on the CFA plots.

Thus, in conclusion, multivariate methods provide descriptive and objective representations of large volumes of data that can be used as the basis of discussions among experts.

References

Almendros, G. (1989). An analysis of some wheat straw humidification factors and their bearing on the response to compost soil and plant. *Science of the Total Environment, 81–82,* 569–578.

Atlaseco de poche (1987). Les Editions SGB (diffusion Robert Laffont). (Data from the World Bank and the OECD)

Avila, F., & Myers, D. E. (1991). Correspondence analysis applied to environmental data sets: a study of Chautauqua Lake sediments. *Chemometrics and Intelligent Laboratory Systems, 11,* 229–249.

Benzécri, J. P. (1982). *Histoire et préhistoire de l'analyse des données.* Paris: Bordas/Dunod.

Benzécri, J. P., et al. (1973). *L'analyse des données. Tome I: La taxinomie. Tome II: L'analyse des correspondances.* (1st ed.) Paris: Bordas/Dunod.

Bobee, B., & Lachance, M. (1984). Multivariate analysis of parameters related to lake acidification in Quebec. *Water Resources Bulletin, 20/4,* 545–556.

Callon, M., Law, J., & Rip, A. (Eds.). (1986). *Mapping the dynamics of science and technology.* London: Macmillan Press.

Carpenter, M., & Narin, F. (1981). The adequacy of the Science Citation Index (SCI) as an indicator of international scientific activity. *Journal of the American Society for Information Science, 32,* 430–439.

Chubin, D. (1986). Noneconomic quantitative measures: The "output" of science. In *Research funding as an investment: Can we measure the returns?* (pp. 29–44). Washington DC: Office of Technology Assessment.

Cortinovis, I., Vella, V., & Ndiku, J. (1993). Construction of a socio-economic index to facilitate analysis of health data in developing countries. *Social Science and Medicine, 36/8,* 1087–1097.

Devillers, J., Steiman, R., Seigle-Murandi, F., Sage, L., Benoit-Guyot, J. L., & Doré, J. C. (1991). Biotic and abiotic factors governing the antifungal activity of microfungi. A multivariate approach. *Systematic and Applied Microbiology, 14,* 196–204.

Doré, J. C., & Ojasoo, T. (1994). Molecular taxonomy by correspondence factorial analysis (CFA). In H. van de Waterbeemd (Ed.), *Advanced computer-assisted techniques in drug discovery,* Vol. 3 ("Methods and principles in medicinal chemistry") (pp. 190–227). Weinheim: VCH.

Dubus, A. (1992). *Méthodes et pratique du traitement statistique en sciences humaines. Logiciel ADSO.* Lille: Atelier des Trois Monts.

The Economist (1991a, September 21). *A survey of business in Eastern Europe.*

The Economist (1991b, November 16). *A survey of Asia's emerging economies.*

Engelsman, E. C., & van Raan, A. F. J. (1994). A patent-based cartography of technology. *Research Policy, 23,* 1–26.

Foucard, T. (1982). *Analyse factorielle. Programmation sur micro-ordinateurs.* Paris: Masson.

Garfield, E. (1972). *Citation indexing—its theory and application in science, technology and humanities.* New York: John Wiley & Sons.

Gauch, H. G. (1982). *Multivariate analysis in community ecology.* Cambridge, Cambridge University Press.

Greenacre, M. J. (1983). *Theory and application of correspondence analysis.* New York: Academic Press.

Greenacre, M. J. (1993). *Correspondence analysis in practice.* New York: Academic Press.

Hoffman, D. L., & Franke, G. R. (1986). Correspondence analysis: Graphical representation of categorical data in marketing research. *Journal of Marketing Research, 23,* 213–227.

Jambu, M. (1991). *Exploratory and multivariate data analysis.* New York: Academic Press.

Karsky, M., Doré, J. C., & Miquel, J. F. (1988). Understanding chemical reactivity tendencies of research. In *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics, Vol. 1.* (pp. 712–721). Beijing: International Academic Publishers.

Lance, G. N., & Williams, W. T. (1966). Computer programs for hierarchical polythetic classification ("similarity analyses"). *Computer Journal, 9,* 60–64.

Lebart, L., Morineau, A., & Fénelon, J. P. (1979). *Traitement des données statistiques. Méthodes et programmes.* Paris: Dunod.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis.* Chichester, UK: John Wiley & Sons.

Miquel, J. F., & Doré, J. C. (1981). Research on new drugs. A system analysis approach. *Drugs of Today, 17,* 487–501.

Miquel, J. F., Ojasoo, T., Okubo, Y., Paul, A., and Doré, J. C. (1995). World science in 18 disciplinary areas: Comparative evaluation of the publication patterns of 48 countries over the period 1981–1992. *Scientometrics, 33,* 149–167.

Miquel, J. F., & Okubo, Y. (1994). Structure of international collaboration in science: Part II: Comparisons of profiles in countries using a LINK indicator. *Scientometrics, 29,* 271–297.

Moed, H. F. (1989). New developments in online bibliometric analysis. In *Science and technology indicators* (pp. 129–146). Leiden: DSWO Press.

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. (1983). On the measurement of research performance: The use of bibliometric indicators. Leiden: Center for Science and Technology Studies, University of Leiden.

Moser, E. B. (1989). Exploring contingency tables with correspondence analysis. *Computer Applications in the Biosciences, 5,* 183–189.

Ojasoo, T., Doré, J. C., & Miquel, J. F. (1994). Regional Cinderellas [Letter]. *Nature, 70,* 172.

Okubo, Y., Miquel, J. F., Frigoletto, L., & Doré, J. C. (1992). Structure of international collaboration in science: Typology of countries through multivariate techniques using a LINK indicator. *Scientometrics, 25,* 321–351.

Pack, P., & Jolliffe, I. T. (1992). Influence in correspondence analysis. *Applied Statistics, 41,* 365–380.

Pirela, A., Rengifo, R., Mercado, A., & Arvanitis, R. (1993). Technological learning and entrepreneurial behaviour: A taxonomy of the chemical industry in Venezuela. *Research Policy, 22,* 431–453.

Prim, R. C. (1957). Shortest connection networks and some generalisations. *Bell System Technical Journal, 36,* 1389–1401.

Renou, M., Lalanne-Cassou, B., Michelot, D., Gordon, G., & Doré, J. C. (1988). Multivariate analysis of the correlation between *Noctuidae* subfamilies and the chemical structure of their sex pheromones or male attractants. *J Chem Ecol, 14,* 1187–1215.

Schubert, A., Glänzel, W., & Braun, T. (1989). Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields, 1981–1985. *Scientometrics, 16,* 3–478.

Small, H. (1973). Co-citation in the scientific literature. A new measure of the relationship between publications. *Journal of the American Society for Information Science, 24,* 265–269.

van der Heijden, P. G. M., de Falguerolles, A., & de Leeuw, J. (1989). A combined approach to contingency table analysis using correspondence analysis and log-linear analysis (with discussion). *Applied Statistics, 38,* 249–292.