Jointly published by Akadémiai Kiadó, Budapest
and Springer, Dordrecht

Scientometrics,
Vol. 63, No. 2 (2005) 259–276

Analysis of the field of physical chemistry of surfactants with the Unified Scientometric Model. Fit of relational and activity indicators

R. BAILÓN-MORENO,[a] E. JURADO-ALAMEDA,[a] R. RUIZ-BAÑOS,[b]
J. P. COURTIAL[c]

[a] *Departamento de Ingeniería Química. Facultad de Ciencias, Universidad de Granada, Granada (Spain)*
[b] *Departamento de Biblioteconomía y Documentación, Facultad de Biblioteconomía y Documentación, Universidad de Granada, Granada (Spain)*
[c] *Laboratoire de Psychologie – Education – Cognition Développement (LabECD), Université de Nantes, Nantes (France)*

By the information system of CoPalRed© and with the treatment of 63,543 bibliographical references of scientific articles, the field of surfactants has been analysed in the light of the Unified Scientometric Model. It was found that the distributions of actors (countries, centres, and research laboratories, journals, researchers, key words of documents) fit Zif's Unified Law better than the Zipf-Mandelbrot Law. The model showed an especially good fit for relational indicators such as density and centrality. Using the Unified Bradford Law, the three zones fit were: core, straight fraction, and Groos droop. The fractality index was used to verify that Science can present fractal as well as transfractal structures. In conclusion, the Unified Scientometric Model is, for its flexibility and its integrating capacity, an appropriate model for representing Science, joining non-relational with relational Scientometrics under the same paradigm.

Introduction

The Unified Scientometric Model is based on 7 principles from fully accepted theories and models in Scientometrics but combined in a new and ingenious way.[1] The principles are:

1. *Actor-network principle:* Science and Technology (Technoscience) is comprised of networks of actors, as established in the Actor-Network Theory of Callon and Courtial.[2-6]
2. *Translation principle*: The dynamic of technoscientific networks is governed by the Translation Theory of Latour.[7,8].

Received June 28, 2004

Address for correspondence:
RAFAEL BAILÓN-MORENO
Departamento de Ingeniería Química. Facultad de Ciencias, Campus de Fuentenueva
Universidad de Granada, 18701-Granada, Spain
E-mail: bailonm@ugr.es

0138–9130/US $ 20.00

3. *Spatial principle*: Translation implies the existence of a space with temporal and geometric components, of the Hausdorff-Besicovitch type, for which the spatial dimensions are fractionary.[9]
4. *Principle of Translation Quantitativity*: The translation, T, is equal to the variation of the qualities or attributes of the actors, $Q(x)$, as they move in the translation space, x.[10,11] The translation is therefore the derivate or gradient of the function quality with respect to the coordinates of the translation space.

$$T(x) = \frac{dQ(x)}{dx} \tag{1}$$

5. *Principle of Composition of the Translations:* Any translation, regardless of how complex, can be considered to be composed of elemental translations, in a series, in parallel or in combinations of both.[10-12]
6. *Principle of the Centre–Periphery or Principle of Nucleation:* The translation space is the field that generates a point, which can be called centre or nucleus, which all the actors try to approach in order to improve their strategic advantage.[13]
7. *Unified Principle of Accumulated Advantages:* The translation, T, is directly proportional to the strategic advantage, s (function of spatial, temporal, or geometric coordinates), and the intrinsic advantage, q (function of the qualities or attributes of the actor or actors). Mathematically, it is expressed by the so-called Fundamental Equation of the Unified Scientometric Model:[1]

$$T = ksq \tag{2}$$

If Eq. 1 of the 4th principle is taken into account and it is considered that the strategic advantage consists of the following function of the range, r (r is a case of space coordinate x):

$$s = -\frac{1}{(r+m)^{\varphi}} \tag{3}$$

and q is identified with $F(r)$, then the fundamental equation becomes:

$$\frac{dF(r)}{dr} = -k\frac{F(r)}{(r+m)^{\varphi}} \tag{4}$$

This expression is a differential equation of separable variables, which depends on the parameters k, m, and φ. According to the model proposed, k is the inverse of the fractal dimension, D. Thus:

$$D = \frac{1}{k} \tag{5}$$

The parameter m, called the Mandelbrot distance (in honour of this researcher), represents the distance from a range-one actor to the centre of the translation space. In many cases, this distance is nil, but in others it is positive.

Finally, in the description of Eq. 4 the parameter which can be considered of great relevance is φ, or the fractality index. (or co-fractality index, β, equal to $1 - \varphi$). When $\varphi = 1$, the system is purely fractal in character and, on resolving the differential equation, it is possible to deduce all the bibliometric distributions known in all their mathematical expressions: Condon–Zipf Law, Zipf–Brookes Law, Booth–Federowicz Law, Zipf–Mandelbrot Law, and their equivalents of the Lotka–Pareto type, (with an exponent equal to 2 or other than 2), and the Bradford type (Brookes–Ferreiro, Leimkuhler and generalized Leimkuhler). The common characteristic of the Zipf laws for $\varphi = 1$ is that they are of the inverse-power type.

When $\varphi = 0$, the proof of Eq. 4 leads to exponential functions, whereas when $0 < \varphi < 1$ or $\varphi < 0$, the Zipf Law generated is a hybrid between the exponential function and the potential function. Therefore, in these cases in which φ is other than 1, we use the Zipf's Unified Law:

$$F(x) = F(0)e^{-b(x+m)^{\beta}} \tag{6}$$

and its equivalents, Bradford's Unified Law (result of the integration of Zipf's Law):

$$R(r) = \int_{x}^{x} F(0)e^{-b(x+m)^{\beta}}\,dx \tag{7}$$

and Lotka's Unified Law (result of the derivation of the inverse function of Zipf's Law):

$$A(T) = \frac{B}{T(\ln T - \lambda)^{\mu}} \tag{8}$$

It should be emphasized that $\varphi = 0$ on the borders between fractal structures (φ positive) and structures that in the Unified Scientometric Model are termed *transfractal* (beyond fractality) when φ is negative. As will be seen in future works, this differentiation between fractal and transfractal geometry is indispensable to explain the so-called "Matthew Effect" and the "Ortega Hypothesis", as well as to make predictions.

For the case that the spatial coordinate is time, and for $\varphi = 0$, the Law of Exponential Growth of Price's Science and from the Ageing Law of Brookes can be deduced.

Similarly, making use of the Principle of Translation Composition, it has been possible to construct the model of Ageing-Validity of Science which, as particular cases, are the Avramescu equation and again Brookes' Ageing Law.

Objectives

An undeniable success of the Unified Scientific Model is to be able to deduce, for diverse values of the parameters *k*, *m*, and φ, all the bibliometric laws known without any exception. However, it remains to be confirmed whether Zipf's Unified Law and its equivalents of Bradford and Lotka, which are proposed for the first time, make a good fit to the empirical values. Previously, it has been shown that the bibliometric distributions often have problems of fit, and therefore it is useful to test whether the new "unified" versions are more appropriate than the previously known ones.[1]

In addition, there is an unanswered question that we believe should be resolved by the proposed model. Scientometrics is usually divided into two broad fields, one which uses activity indicators and another which uses relational indicators. The mathematical treatments in the two cases appear to be divorced. The model proposed must in some way wed the two and demonstrate that the indicators being handled in both cases have shared behaviour.

Therefore, the two prime objectives of this work are:

a) To determine whether the Unified Scientometric Model is capable not only of generating all the bibliometric laws known but also of correctly fitting the empirical values in those cases in which the laws used to date are not capable of doing so.

b) To unify, under a common behaviour, the activity indicators (production, frequency of appearance, etc.) with relational ones (centrality, density of cowords analysis).

Materials and methods

The scientific field to be studied concerns surfactants and related products. The primary data consist of a set of 63,543 bibliographic references of scientific articles for the period 1993-2002. These references were downloaded through Internet from the Science Citation Index (SCI), Expanded version, of the Institute for Scientific Information (ISI). The search was:

SURFACTANT* OR DETERGENT* OR TENSIDE* OR CLEANER* OR LAUNDRY* OR FRAGRANCE* OR PERFUME* OR FLAVOR* OR ODOR* OR (ESSENTIAL SAME OIL*) OR COSMETIC* OR TOILETR* OR SOAP*

For the data analysis, the well-known CoPalRed© was used, a system enabling, among other things, the treatment of activity combined with a relational analysis. As a preliminary step, a general thematic analysis was performed to determine the broad research areas. It was found that, for an occurrence and co-occurrence of the descriptors equal to or greater than 50, and a cluster size of between 6 and 20, the scientific field of surfactants was divided into the large areas specified in Figure 1. From among these areas, the physical chemistry of surfactants (Topics 2, 4, 5, 8 and 10) was chosen, with a total of 8,678 articles, to perform a more detailed topical study divided into 5 two-year time periods.

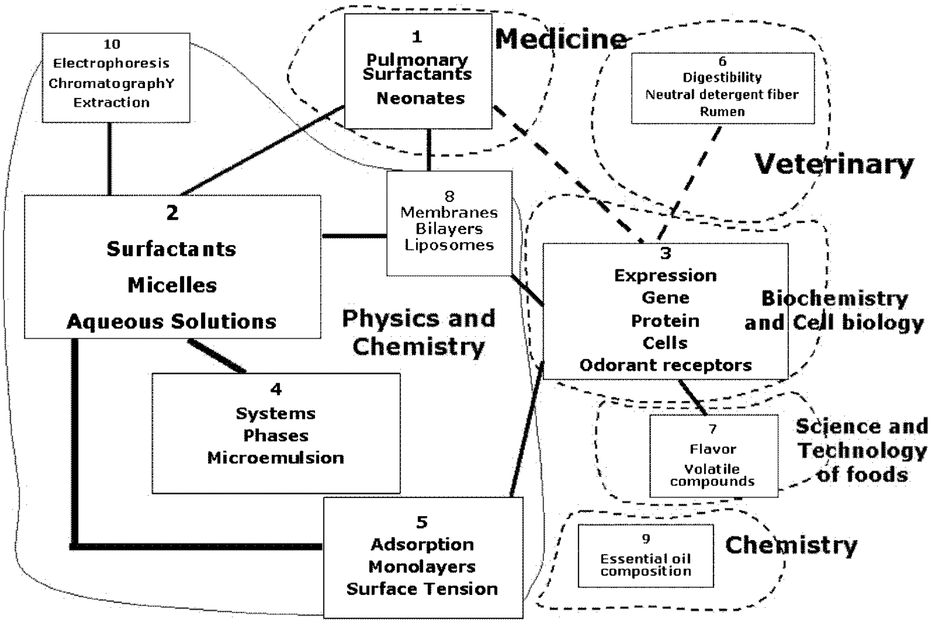

Figure 1. Topical map of the broad areas of research in the field of surfactants and similar products

Overall, of the 63,543 articles from the scientific field of surfactants, the production distributions were determined per countries, per centres and laboratories, per journals, and per researchers, as was the distribution of the frequency of the descriptors. With respect to relational indicators, and from the subjects 2, 4, 5, 8 and 10 (8,678 articles), for each two-year period, a topical map was drawn in order to detail the physical chemistry of surfactants, calculating the centrality and the density of the resulting research topics.

Production per countries, centres, journals and researchers. Frequency of the descriptors

The distributions that to be fit have as a spatial coordinate the range r. If we consider that the fractality index, φ, is equal to unity, the most general equation of Zipf's Law is given by Mandelbrot:

$$F(r) = \frac{k_m}{(r+m)^{\alpha}} \tag{9}$$

However, as we are not going to ensure that the fractality index, φ, is exactly unity, we rule out, in principle, that it is null, and the most adequate equation of the fit will be the proposed Unified Law of Zipf (Eq. 6):

$$F(x) = F(0)e^{-b(x+m)^{\beta}} \tag{6}$$

An attempt was made to fit the data by non-linear regression using the MathCad program, professional version 7.0. It was found that the result is highly sensitive to the initial values proposed. Therefore, given the uncertainty that the result is correct, a specific adjustment module was designed and programmed in CoPalRed, which does guarantee a completely correct and solid result. The program indicates the optimal values of m and β (bearing in mind that $\beta = 1 - \varphi$), which make R^2 maximum. For m and β the program offers the value of k, $F(0)$, and b, resultants of a linear regression.

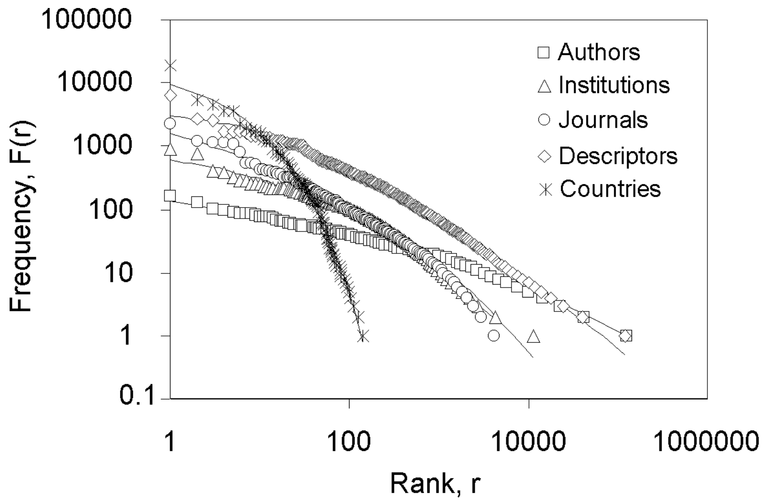

Figure 2. Production of articles by author, centre, journal, and country. Distribution of the frequency of descriptors. Fit by Zipf's Unified Law

Figure 2 shows the empirical values and adjustment curves. Table 1 lists the values of the parameters φ, k, and m for each distribution. In addition, the coefficient of determination is indicated, this being greater than 0.990 in all cases except for that of the centres, which is slightly less (0.986). Nevertheless, the fit is very good and clearly superior in all cases to that using the Zipf–Mandelbrot Law.

Table 1. Parameters of distributions fit to Zipf's Unified Law. Comparison with the Zipf–Mandelbrot Law

	Authors	Centres	Journals	Descriptors	Countries
	Zipf's Unified Law (Eq. 6)				
Zipf K Constant	0.1971	0.2386	0.4264	0.5256	0.3906
Mandelbrot Distance, m	0	0	0	3.82	0
Fractality Index, φ	0.885	0.782	0.857	0.929	0.474
Co-Fractality Index, β	0.115	0.218	0.143	0.071	0.526
$F(0)$	739.5	1848	32286	11.95×10^6	19853
b	1.714	1.094	2.982	7.403	0.7426
R^2	0.997	0.986	0.995	0.996	0.993
Dimension, D	5.07	4.19	2.35	1.90	2.56
% Fractality	88.5	78.2	85.7	92.9	47.4
	Zipf–Mandelbrot Law (Eq. 9)				
R^2	0.891	0.825	0.944	0.883	0.840

The translation space of the descriptors is almost a plane (D = 1.90) while that of the journals and countries can be considered intermediate between a plane and a sphere (D = 2.35 and D = 2.56, respectively). Centres and authors form hyperspheres.

No case was 100% fractal (this being why none perfectly fit the Zipf–Mandelbrot Law), although the closest one was the distribution of the descriptors. On the other hand, the case of production of the countries hardly reached half fractality (φ = 0.474)

The Mandelbrot distance, m, had a value greater than zero only for the descriptors. That is, in the rest of the cases the centre of the translation space was not far from the actor or set of actors of higher frequency.

Figure 3 shows a graphic representation of the calculated frequencies against observed ones, confirming that all the points strongly approach a diagonal, further evidence of the goodness of the Unified Scientometric Model. Meanwhile, Figure 4 reflects the improvement achieved with respect to Zipf–Mandelbrot.

The distribution of residuals (difference between calculated and observed values, expressed in percentage) are distributed uniformly throughout the ranges, indicating that the model is correct and there is no appreciable bias (Figure 5). In addition, the value of these residuals is notably lower than in the case of the fit with Zipf–Mandelbrot (Figure 6).

With the values calculated, it is possible to provide representations of Figure 2 not only of the Unified Zipf type, but also representations of the Unified Bradford as well as Unified Lotka types. The results are analogous and for the sake of brevity are not shown.

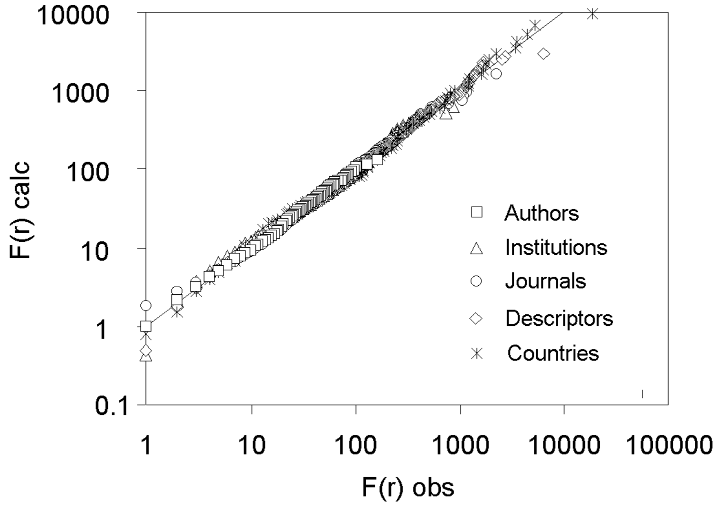

Figure 3. Calculated frequencies against observed ones fit by the proposed Unified Law of Zipf

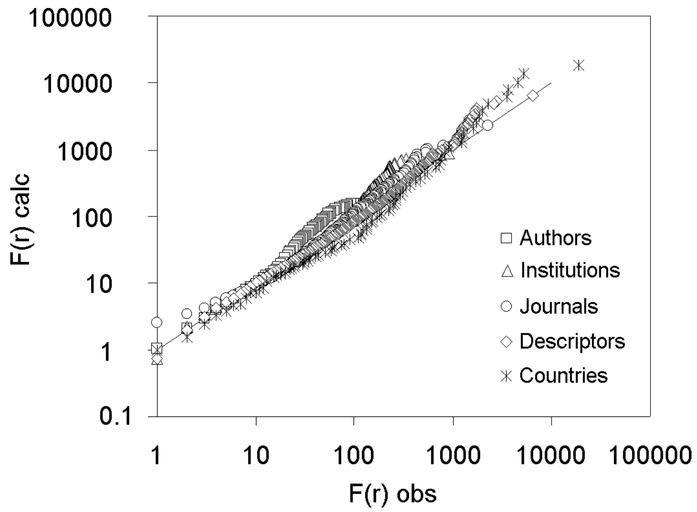

Figure 4. Observed values against calculated ones in the overall distributions of authors, centres, journals, and descriptors for a Zipf–Mandelbrot fit

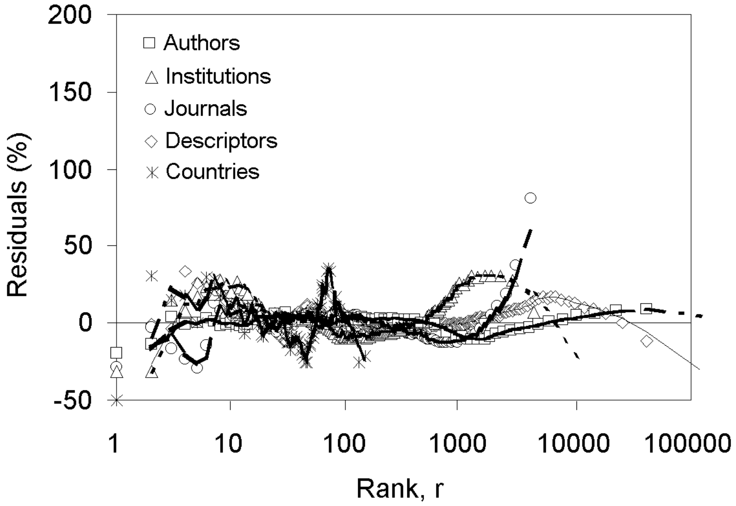

Figure 5. Distribution of residuals for the fit of the proposed Unified Law of Zipf

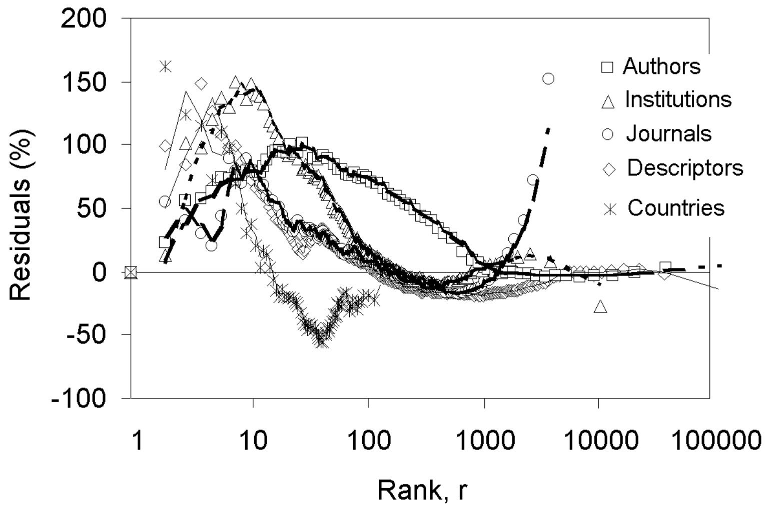

Figure 6. Residuals for Zipf–Mandelbrot fits for authors, centres, journals, and descriptors

The Unified Scientometric Model and coword analysis

The bibliometric distributions of production of authors, journals, etc., which classically generate the different forms of Zipf's Law, Lotka's Law, and Bradford's Law, are based on first-generation indicators that are not relational (activity indicators). The Unified Scientometric Model can be considered truly unified if in addition to being able to represent correctly the non-relational classical distributions, it can also fit the distributions generated from a relational analysis.

In the Coword Analysis, the two most important indices are centrality and density. Clearly, our target should be to determine how the topics are distributed according to these indices. The analysis is directed, finally, towards a behaviour of the axes of the strategic diagram.

Zipf's Unified Law for centrality

If we consider the centrality values of each of the topics generated in the five periods into which the analysis of the physical chemistry of surfactants was divided and if we represent the ordinates against range, we get the curves of Figure 7. The parameters of fit are specified in Table 2, where e proves excellent even in the period 1995–1996, particularly taking into account that these distributions have very few values to fit in comparison with the descriptors authors or journals, and any small divergence of a point abruptly diminishes the coefficient of determination R^2. In any case, the use of the Zipf–Mandelbrot equation is very unsatisfactory in comparison with the Unified Zipf equation proposed.

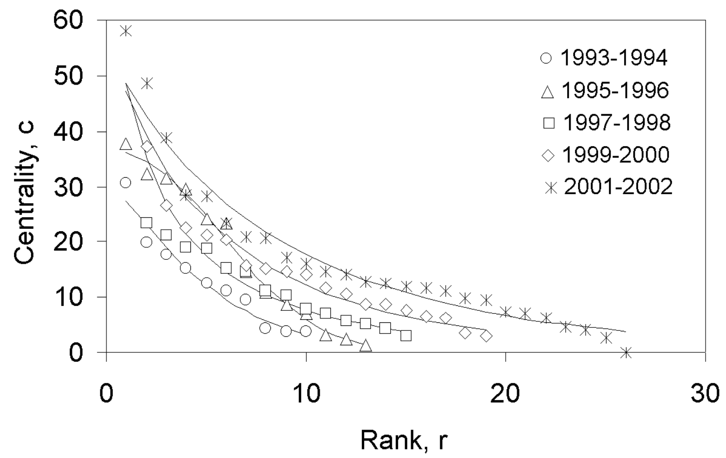

Figure 7. Fit of centrality to Zipf's Unified Law

The dimensions presented by the network with respect to centrality are noteworthy, going from 2.6, intermediate between a circle and a sphere (roughly analogous to a kind of bowl), to a hypersphere of more than 41 dimensions. This variability is surprising, combined also with the values of φ, which in this case even become negative; there is an alternation between fractal structure and transfractal structures.

Table 2. Parameters of fit to Zipf's Unified Law

Unified Zipf

Period	1993–1994	1995–1996	1997–1998	1999–2000	2001–2002
k	0.157	2.43E-02	0.3865	0.2051	0.13577
m	0	0	0	0	0
β	-0.243	-1.221	0.425	0.202	0.124
$F(0)$	31.092	36.382	95.383	61.073	56.676
b	0.1266	0.0109	0.672	0.257	0.155
φ	1.243	2.221	0.575	0.798	0.876
D	6.4	41.2	2.6	4.9	7.4
R^2	0.949	0.992	0.943	0.945	0.954

Figure 8 presents the calculated values against observed ones, showing that all share a good alignment. Additional confirmation of the goodness of the model for the case of centrality comes from Figure 9 with the residuals, which are distributed at random on both sides of the central line.

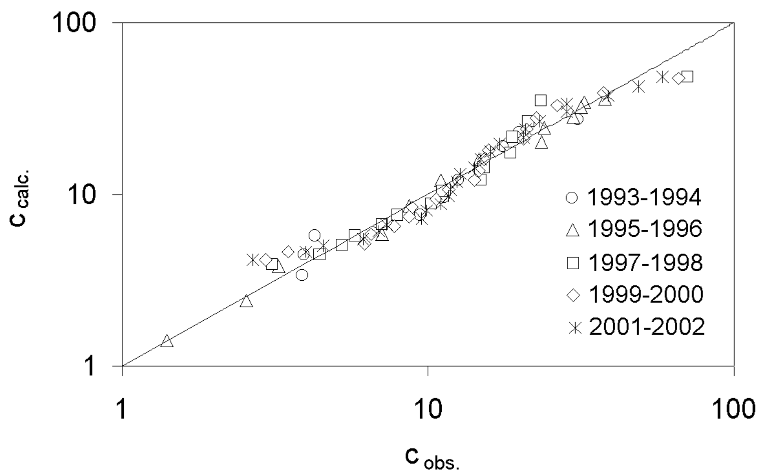

Figure 8. Calculated values against observed ones in the centrality fit to the Zipf's Unified Law

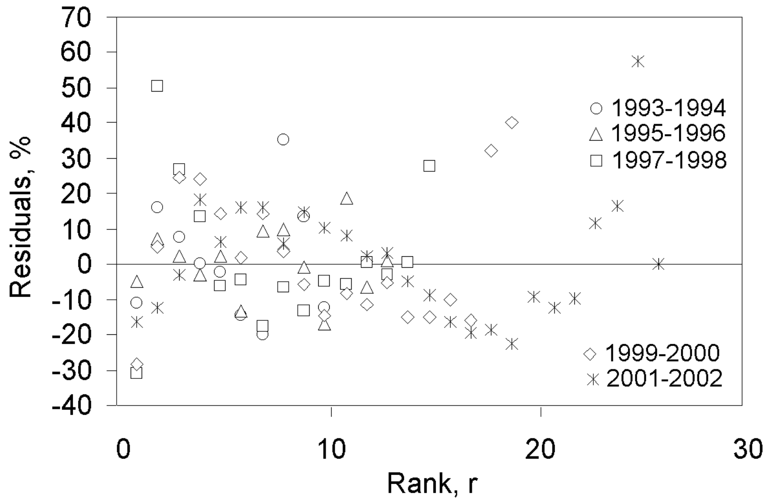

Figure 9. Residuals for the centrality fit to Zipf's Unified Law

Bradford's Unified Law for centrality

If both observed and calculated centrality values accumulate and are represented against range, we get observed and calculated Bradford distributions, respectively, as shown in Figure 10, where the continuous lines represent accumulated centrality values calculated. In all cases, we find the typical "S" shape of the Bradford distribution, which includes the nucleus, straight fraction, and Groos droop. Most striking is the very good correspondence between the observed and calculated values. The coefficients of determination are those of Table 3, the coefficient from 1995–1996 reaching a value of 1,000, when rounded off by a thousandth, or 0.9998 when rounded off by a ten-thousandth.

Figure 11 presents the correlation of the accumulated-centrality values calculated with the Unified Scientometric Model and the observed values. The correlation is impeccable.

Zipf's Unified Law for density

In a way analogous to the fit of centrality, the densities of the topics have been fit with the proposed Unified Law of Zipf in Figure 12, and the empirical parameters of the fit appear in Table 4.

Table 3. Coefficient of determination for fitting the centralities of Bradford's Unified Law

Period	1993–1994	1995–1996	1997–1998	1999–2000	2001–2002
R^2	0.998	1.000	0.990	0.994	0.997

Table 4. Parameters for fitting the densities of Zipf's Unified Law

Zipf's Unified Law					
Period	1993–1994	1995–1996	1997–1998	1999–2000	2001–2002
k	0.698	1.18E-01	4.889E-02	0.459	0.09969
m	0	0	0	0	0
β	0.595	-0.362	-0.743	0.445	-0.117
$F(0)$	138.7	32.293	22.245	78.265	38.325
b	1.724	8.69E-02	2.805E-02	0.827	0.08925
φ	0.405	1.362	1.743	0.555	1.117
D	1.4	8.5	20.5	2.2	10.0
R^2	0.936	0.956	0.991	0.985	0.975

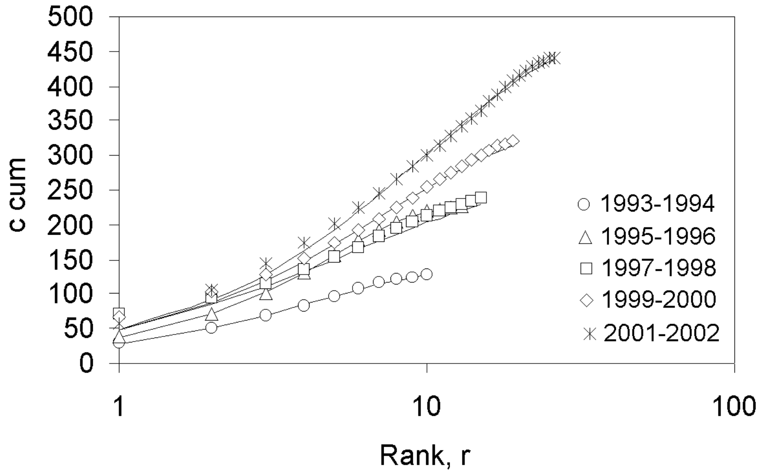

Figure 10. Fit of centralities to Bradford's Unified Law

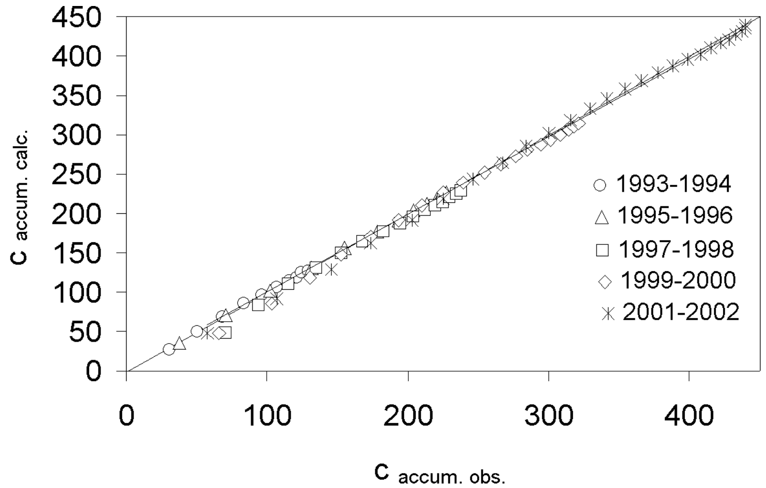

Figure 11. Calculated values of accumulated centralities against observed values in the fit to Bradford's Unified Law

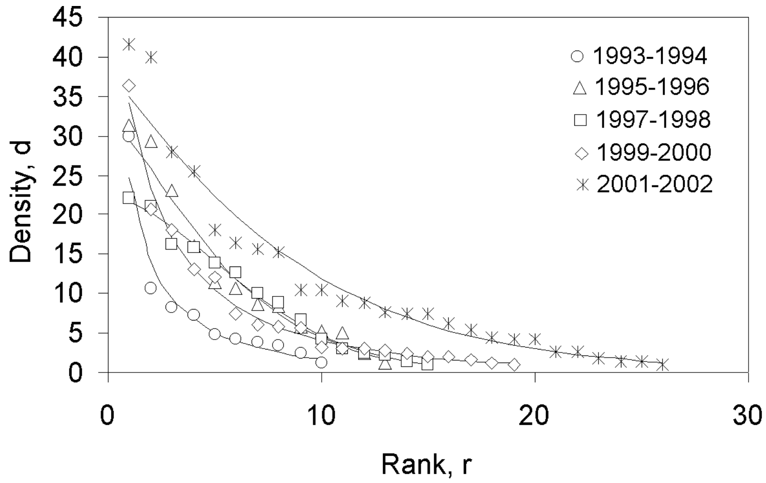

Figure 12. Fit of densities to Zipf's Unified Law

As in the case of the centralities, the dimension of the translation space can present great differences, from 1.4 (broken line) to a hypersphere of dimension 20.5. Similarly, there are not only cases of more or less fractal structure (positive φ), but also cases of more or less transfractal structure (negative φ).

In Figures 13 and 14, the calculated and observed values are well aligned and the residuals distributed completely at random. The Unified Law of Zipf proposed is adequate to represent the density distribution of the topics taken from the Coword Analysis. In other words, the Unified Scientometric Model correctly fits the ordinates of the strategic diagram.

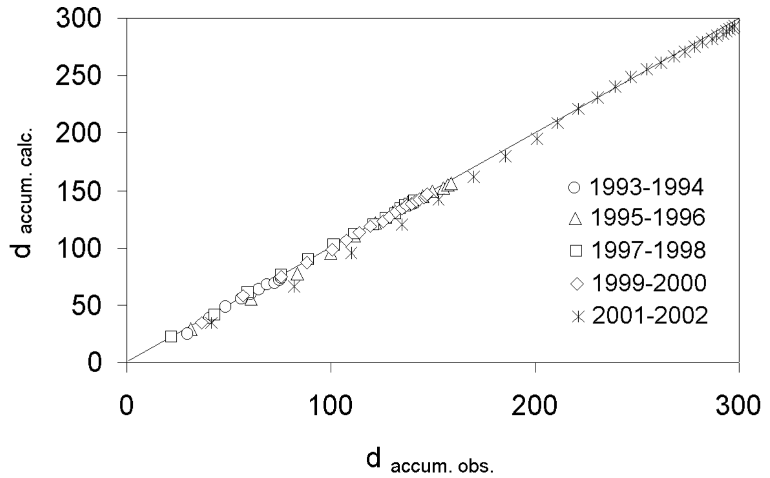

Figure 13. Calculated density values against observed ones in the fit to Zipf's Unified Law

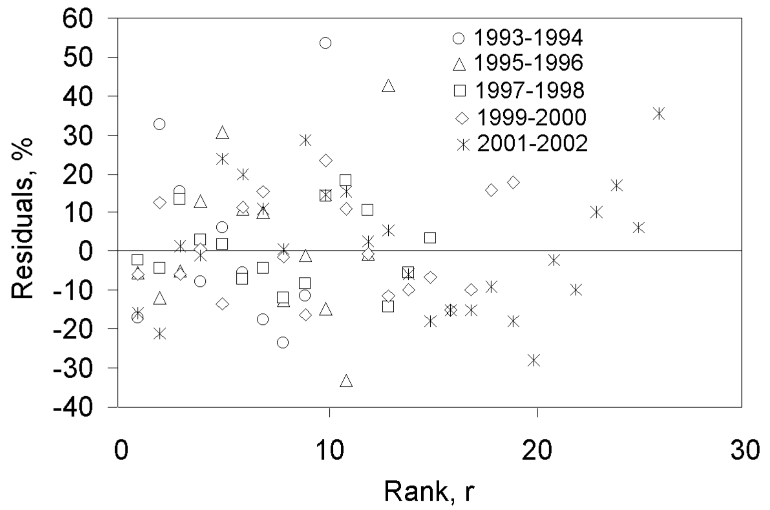

Figure 14. Density residuals in the fit to Zipf's Unified Law

Bradford's Unified Law for densities

As with centrality, observed and calculated values were accumulated to prepare Bradford-type distributions. The result is shown in Figure 15, where the continuous lines represent calculated values and the symbols observed values. Again, the Bradford "S" curve appears with its nucleus, straight fraction, and Groos droop. Also, the fit is successful over the entire range with the correlation coefficients specified in Table 5.

Figure 16 presents the correlation of the accumulated-density values calculated with the Unified Scientometric Model and the observed values. Except for the period 1993–1994 with an R^2 of "only" 0.990, the other periods show a correlation that can be qualified as excellent (R^2=1,000 for 1997–1998 and 1999–2000)

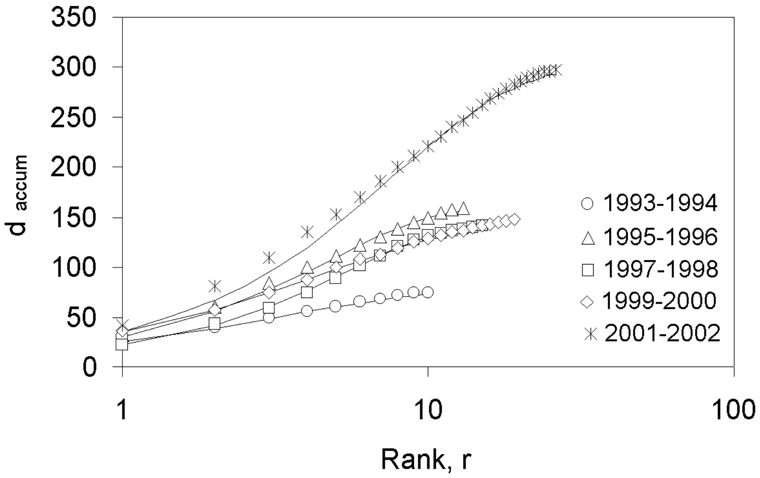

Figure 15. Fit of the densities to Bradford's Unified Law

Table 5. Parameters of fitting the densities to the Unified Bradford Law

Period	1993–1994	1995–1996	1997–1998	1999–2000	2001–2002
R^2	0.990	0.996	1.000	1.000	0.995

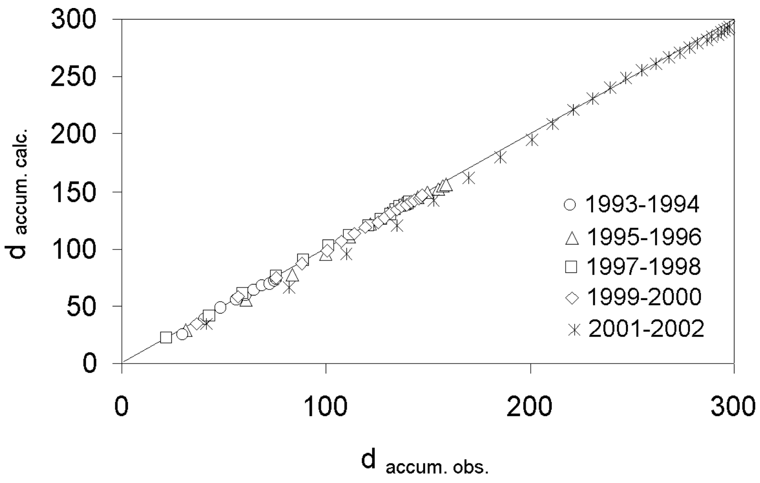

Figure 16. Calculated values of accumulated densities against observed values in the fit to Bradford's Unified Law

Conclusions

As clearly demonstrated throughout this paper, the Unified Scientometric Model can correctly fit the production of articles per country, research centre, journal, and researcher. Also, it provides very good fits for the frequency of the appearance of descriptors.

In addition, it has been confirmed that it is not only capable of fitting activity indicators but it is especially suitable for the distribution of research topics according to their density and centrality. Bradford's Unified Law provided an exceptional fit in all three zones: nucleus, straight fraction, and Groos droop. The fact that the indicators offered by Coword Analysis are distributed according to a common patter similar to that of the indicators of classical activity reaffirms even further the relevance, solidity, and reliability of the Coword Analysis as a means to study techno-scientific networks.

Furthermore, by the fractality index, φ, it has been demonstrated that the field of surfactants (and very probably by extension any field in Science) presents both fractal and transfractal structures.

In short, the Unified Scientometric Model is, for its flexibility and its integrating capacity, a suitable model for the representation of Technical Science, joining under the same paradigm non relational and relational Scientometrics.

*

This study was supported by the Spanish Ministry of Science and Technology through projects no. 1FD97-0931 and PB1998-1293.

References

1. BAILÓN-MORENO, R., *Ingeniería del conocimiento y vigilancia tecnológica aplicada a la investigación en el campo de los tensioactivos. Desarrollo de un modelo ciencimétrico unificado.* Ph. D. Thesis. Universidad de Granada, Granada, 2003.
2. CALLON, M., COURTIAL, J. P., LAVILLE, F., Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. *Scientometrics*, 22 (1991) 155–205.
3. CALLON, M., LAW, J., RIP, A., *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London, The McMillan Press LTD, 1986.
4. CALLON, M., COURTIAL, J. P., PENAN, H., BAUIN, S., From translation to problematic networks: an introduction to coword analysis. *Social Science Information*, 22 (1983) 191–235.
5. COURTIAL, J. P., MICHELET, B., A mathematical model of development in a research field. *Scientometrics*, 19 (1990) 127–141.
6. COURTIAL, J. P., *Introduction à la scientomètrie: de la bibliométrie à la veille techonologique*, Anthropos, Paris , 1990.
7. LATOUR, B., *Ciencia en acción: Cómo seguir a los científicos e ingenieros a través de la sociedad*, Labor, Barcelona, 1992.
8. LATOUR, B., WOOLGAR, S., *La vida en el laboratorio. La construcción de los hechos científicos*, Alianza, Madrid, 1995.
9. MANDELBROT, B. B., *The Fractal Geometry of Nature*, Freeman, New York, 1977.
10. RUIZ-BAÑOS, R., BAILÓN-MORENO, R., JIMÉNEZ-CONTRERAS, E., COURTIAL, J. P., Structure and dynamics of scientific networks. Part 1: Fundamentals of the quantitative model of translation. *Scientometrics*, 44 (1999) 217–234.
11. RUIZ-BAÑOS, R., BAILÓN-MORENO, R., JIMÉNEZ-CONTRERAS, E., COURTIAL, J. P., Structure and dynamics of scientific networks. Part 2: The new Zipf's Law, the cocitations's clusters and the model of the presence of key-words. *Scientometrics*, 44 (1999) 235–265.
12. RUIZ-BAÑOS, R. Las traducciones dinámicas de las series temáticas. Propuesta de una nueva clasificación. *Actas del IV Congreso ISKO España EOCONSID'99. La Representación y Organización del Conocimiento en sus distintas perspectivas: su influencia en la Recuperación de la Información*, pp. 193–198. Capítulo Español de la Sociedad Internacional para la Organización del Conocimiento, 1999.
13. JIMÉNEZ-CONTRERAS, E., Las revistas científicas: el centro y la periferia. *Revista Española de Documentación Científica*, 15 (1992) 174–182.