Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 73, No. 3 (2007) 265–279
DOI: 10.1007/s11192-007-1798-5

Scientometric analysis of geostatistics using multivariate methods

FENG ZHOU,[a] HUAI-CHENG GUO,[a] YUH-SHAN HO,[a] CHAO-ZHONG WU[b]

[a] *College of Environmental Sciences, Peking University, Beijing (P. R. China)*
[b] *Intelligent Transport System Research Center, Wuhan University of Technology, Wuhan, Hubei (P. R. China)*

Multivariate methods were successfully employed in a comprehensive scientometric analysis of geostatistics research, and the publications data for this research came from the Science Citation Index and spanned the period from 1967 to 2005. Hierarchical cluster analysis (CA) was used in publication patterns based on different types of variables. A backward discriminant analysis (DA) with appropriate statistical tests was then conducted to confirm CA results and evaluate the variations of various patterns. For authorship pattern, the 50 most productive authors were classified by CA into 4 groups representing different levels, and DA produced 92.0% correct assignment with high reliability. The discriminant parameters were mean impact factor (MIF), annual citations per publication (ACPP), and the number of publications by the first author; for country/region pattern, CA divided the top 50 most productive countries/regions into 4 groups with 95.9% correct assignments, and the discriminant parameters were MIF, ACCP, and independent publication (IP); for institute pattern, 3 groups were identified from the top 50 most productive institutes with nearly 88.0% correct assignment, and the discriminant parameters were MIF, ACCP, IP, and international collaborative publication; last, for journal pattern, the top 50 most productive journals were classified into 3 groups with nearly 98.0% correct assignment, and its discriminant parameters were total citations, impact factor and ACCP. Moreover, we also analyzed general patterns for publication document type, language, subject category, and publication growth.

Received January 15, 2007

Address for correspondence:
HUAI-CHENG GUO
Room 208, Old Geosciences Building, College of Environmental Sciences
Peking University, Beijing 100871, China
E-mail: hcguo@pku.edu.cn

0138–9130/US $ 20.00

Introduction

Geostatistics, as a branch of applied mathematics and statistics, first appeared during the mineral exploration of the early 1950s as a means of improving ore reserve estimation (WACKERNAGEL, 2003). The original work in this area was conducted in South Africa by D. G. Krige and H. S. Sichel (KRIGE, 1952; SICHEL, 1952), but the field later came to be associated with the work of Georges Matheron, the founder of the discipline of geostatistics. Matheron set Krige's innovative concepts into a single, powerful theoretical framework of regionalized variables that could be used to describe spatial autocorrelation (MATHERON, 1993; WACKERNAGEL, 2003). For the past years, geostatistics has been deeply developed and widely applied as a advanced statistical methodology in petroleum engineering, environmental science, agriculture, and epidemiology, as well as in ecological studies (DELHOMME, 1979; ROSSI et al., 1992; PETIT et al., 1994; HERZFELD & HIGGINSON, 1996; BARABAS et al., 2001; KRIEGER et al., 2002). Therefore, it is necessary to investigate and identify the quantitative characteristics of its methodologies development and applications.

In recent years, scientometrics has been used for quantitative analysis and for obtaining statistics that measure the contribution of scientific publications to the advancement of knowledge within a given topic or country (CHIU & HO, 2005; NWAGWU, 2006). Indeed, scientific publications represent current research trends and can be used to identify the focus of present, past, or future research (GARFIELD, 1970). Moreover, the Social Science Citation Index (SSCI), the Science Citation Index (SCI), and the Arts and Humanities Citation Index (A&HCI) databases provide a ready data source for scientometric analysis. To a certain extent, previous scientometric analyses have nearly been limited to considering a single variable, such as document type, subject category, or publication pattern, and few studies have considered using various multivariate statistical models to identify the patterns of authors, countries/regions, institutes, or journals (FERNANDEZ-CANO & BUENO, 2002; ENACHESCU & POSTELNICU, 2003; FABA-PEREZ et al., 2003). Moreover, appropriate statistical tests have always been ignored in multivariate analysis (ZHOU et al., 2006).

In this study, we sought primarily to apply multivariate methods, such as cluster analysis (CA) and discriminant analysis (DA), in publication pattern recognitions of authors, countries/regions, institutes, and journals that were present in geostatistical methodologies development and its applications during the period from 1967 to 2005. In addition, we also attempted to analyze general publication patterns, including document type, language, subject category, and publication growth trends.

Materials and methods

Data collection

The documents considered in this study come from the Science Citation Index (SCI), which is maintained by the Institute for Scientific Information (ISI) Web of Science in Philadelphia. We used the keyword "geostatistics," "geostatistic," and "geostatistical" to search titles, abstracts, and keywords. We determined the impact factor (IF) of a given journal by using the *Journal Citation Reports* (JCR) published in 2005. To assess exactly the impact of an article, author, country/region, or institute in the following analysis, we created and used a new variable, annual citations (AC), or annual citations per publication (ACPP), instead of total citations (TC), because TC tends to be highly correlated with the length of time since publication.

Cluster analysis

CA is an unsupervised pattern recognition method that divides a large group of cases into smaller groups, or clusters, of relatively similar cases that are dissimilar to other groups. Hierarchical CA, the most common approach, starts with each case in a separate cluster and joins the clusters together, step-by-step, until only one cluster remains (ZHOU et al., 2006). In particular, we used Ward's method with squared Euclidean distances, which can be expressed as follows (LATTIN et al., 2003):

$$D_{ij}^2 = \sum_{k=1}^{n} (z_{ik} - z_{jk})^2 , \qquad (1)$$

$$z_{ik} = \frac{x_{ik} - \bar{u}_k}{\sigma_k} , \qquad (2)$$

where D_{ij}^2 is the squared Euclidean distance between object i and object j, and z_{ik} and z_{jk} are the normalized values for variable k (for $k = 1, \ldots, n$), which minimizes the effects of differences in measurement units and variances and renders the data dimensionless. Moreover, x_{ik}, $\bar{u}_k$, and σ_k are the measured value, average value, and standard deviation of the variable k, respectively. We performed hierarchical CA to identify the similarities among the different authors, countries/regions, and institutes, as well as journals. The corresponding parameters used include the impact factor (IF), the mean impact factor (MIF), the ACPP, the number of publications by the first author (FP), the number of publications by the corresponding author (RP), independent publication (IP), and international collaborative publication (CP).

Discriminant analysis

DA is a method of analyzing dependence that is a special case of canonical correlation used to discriminate between two or more groups on the basis of discrimination parameters (JOHNSON et al., 2002). DA operates on original data, and the method constructs a discriminant function for each group, as follows:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_j \cdot p_{ij} \ , \tag{3}$$

where i is the number of groups determined by CA, k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, and w_j is the weight coefficient assigned by DA to a given parameter (p_{ij}). We used a backward stepwise approach to construct the best discriminant functions (DFs). We then confirmed the groups determined by means of CA and evaluated the variations in the patterns for authors, countries/regions, institutes, and journals. We constructed the best discriminant functions by considering the quality of the classification matrices and the number of parameters by means of appropriate statistical tests, Wilks' λ, the F-test, and the p-level.

Results and discussion

General patterns

The 2,866 publications identified by the ISI between 1967 and 2005 were of 12 document types, though most were articles. More specifically, 2,691 (93.9%) were original articles. The remaining publications were reviews (46; 1.61%), meeting abstracts (41; 1.43%), editorial materials (27; 0.94%), letters (23; 0.80%), notes (11; 0.38%), book reviews (9; 0.31%), discussions (9; 0.31%), corrections (6; 0.21%), corrections and additions (1; 0.03%), news items (1; 0.03%), and software reviews (1; 0.03%).

Our language analysis of 2,691 articles indicated that 2,625 articles (97.5%) were published in English. The remainder were published in French (20; 0.74%), Portuguese (20; 0.74%), German (19; 0.71%), Spanish (3; 0.11%), Arabic (1; 0.03%), Chinese (1; 0.03%), Italian (1; 0.03%), or Japanese (1; 0.03%).

In addition, these 2,691 articles spanned 107 subject categories. The top 10 categories with the largest number of articles were Geosciences, Multidisciplinary (698; 25.9%), Water Resources (496; 18.4%), Environmental Sciences (493; 18.3%), Agriculture or Soil Science (333; 12.4%), Mathematics or Interdisciplinary Applications (242; 8.99%), Statistics and Probability (183; 6.80%), Ecology (176;

6.54%), Civil Engineering (150; 5.57%), Petroleum Engineering (150; 5.57%), and Limnology (141; 5.24%). Hence, geostatistics has been successfully developed and applied in diverse fields,

Figure 1 presents publication output growth, TC, MIF, and ACPP for these articles for the period from 1972 to 2005 (no article was published from 1967 to 1971). The production trend rose slowly during the first 18 years but steadily and rapidly increased after 1991, and more than 81% of the articles were published during the period from 1995 to 2005. The annual total citations of articles (TC) peaked in 1992, 1994, and 1996–1998, which was the period of greatest popularity for these publications (Figure 1). The trends for ACPP and MIF indicated that the ACPP (>1.0) was high primarily in 1974, 1983, 1991, 1992, 1994–1998, 2000, and 2002, but that the MIF (>1.0) was high only between 1989 and 2005; the latter increased continuously after 1988. Meanwhile, the trend in ACPP reflected not only the trend in TC but also other information that gave rise to higher points on the graph; this discrepancy illustrates the advantage of using ACPP.

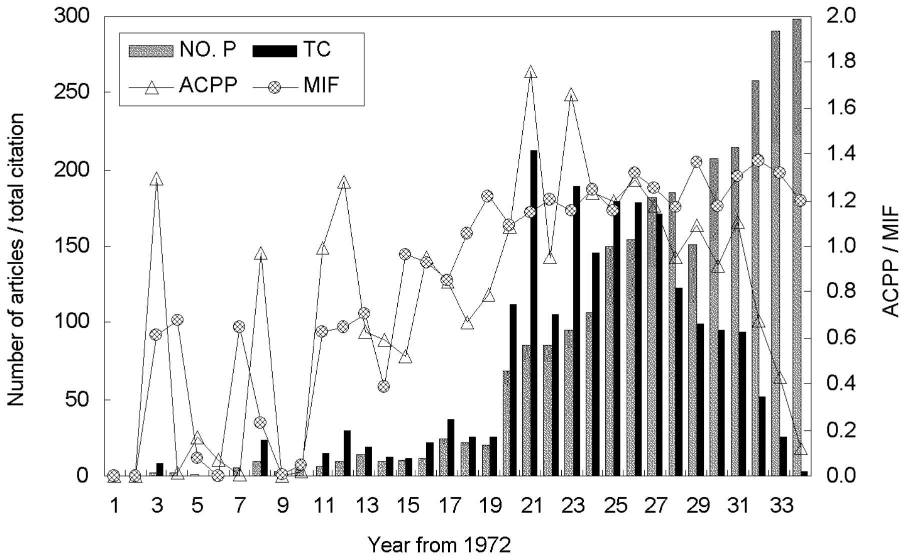

Figure 1. Annual articles (NO. P), total citation (TC), mean impact factors (MIF), and annually citation per publication (ACPP) from 1972 to 2005
Note: "1" and "34" represent 1972 and 2005, respectively

Authorship pattern

Eight documents in the SCI database had no corresponding author information. Based on Table 1, the most prolific first author, however, had a relatively low ACPP score. The top 10 most prolific first authors come from the USA, UK or Canada. Moreover, the maximum MIF and ACPP scores for their publications were 1.89 and 2.99, and the authors of the papers with these scores were R. M. Lark and P. Goovaerts, respectively. The most productive corresponding author and byliner was P. A. Dowd (22 papers) and A. Stein (34 papers). Moreover, the MIF of P. Goovaerts' papers was the highest among those of the top 10 corresponding authors and byliners, and the highest ACPPs for their publications were 2.92 and 2.18.

Table 1. Most productive authors (first author, corresponding author and byliner) in geostatistics research from 1967–2005

Author	Country	TP	MIF	ACPP
First author (a)				
Herzfeld, UC	USA	22	1.01	0.61
Pardo-iguzquiza, E	UK	20	0.92	0.51
Goovaerts, P	USA	14	1.58	2.99
Lark, RM	UK	13	1.89	0.98
Dowd, PA	UK	10	0.33	0.18
Journel, AG	USA	10	0.61	0.93
Kitanidis, PK	USA	10	1.28	1.94
Atkinson, PM	UK	9	1.21	1.12
Carroll, SS	USA	9	1.85	1.38
Desbarats, AJ	Canada	9	1.42	1.76
Corresponding author (b)				
Dowd, PA	UK	22	0.62	0.23
Herzfeld, UC	USA	20	1.07	0.62
Goovaerts, P	USA	15	2.04	2.92
Lark, RM	UK	13	1.89	0.98
Deutsch, CV	Canada	12	0.92	0.82
Stein, A	Netherlands	11	1.17	0.68
Kitanidis, PK	USA	10	1.28	1.94
Atkinson, PM	UK	9	1.21	1.12
Carroll, SS	USA	9	1.85	1.38
Oliver, MA	UK	9	0.93	1.15
Byliner author (c)				
Stein, A	Netherlands	34	1.13	0.71
Deutsch, CV	Canada	31	0.85	0.62
Goovaerts, P	USA	29	1.94	2.11
Dowd, PA	UK	24	0.64	0.27
Herzfeld, UC	USA	24	1.03	0.58
Kitanidis, PK	USA	24	1.40	1.88
Pardo-Iguzquiza, E	UK	22	0.92	0.47
Journel, AG	USA	20	0.97	2.18
Webster, R	UK	19	1.43	1.48
Yeh, TCJ	USA	19	1.57	1.70

We further used CA to identify the authorship pattern of the top 50 most productive authors based on the four variables MIF, ACPP, FP, and RP. This analysis produced a dendrogram with four groups at $(D_{link}/D_{max}) \times 100 < 75$, and the difference between the clusters was significant (Figure 2). To validate the results of the CA, we conducted a backward DA with statistical tests (Table 2a). The value of Wilks' λ, the *F*-value, and the *p*-level for this analysis were 0.09097, 19.99, and <0.0000, respectively. These results indicated that the DA methodology had great discriminatory ability. In particular, the DA produced three DFs that approached 92.0% correct assignment using three discriminant parameters (MIF, ACCP, and FP), which corroborated the results derived by CA. The MIF, ACPP, FP, and RP scores for group 1 were the lowest, as these authors were the relatively least productive; these authors in group 2 had high MIF and ACPP scores but low FP and RP scores, since their work had the relatively highest IFs and were most frequently cited; and group 4 had high FP and RP scores but low MIF and ACPP values, since they were the most productive authors; Group 3, who had high MIF, ACPP, FP, and RP scores, were the leading authors during the period.

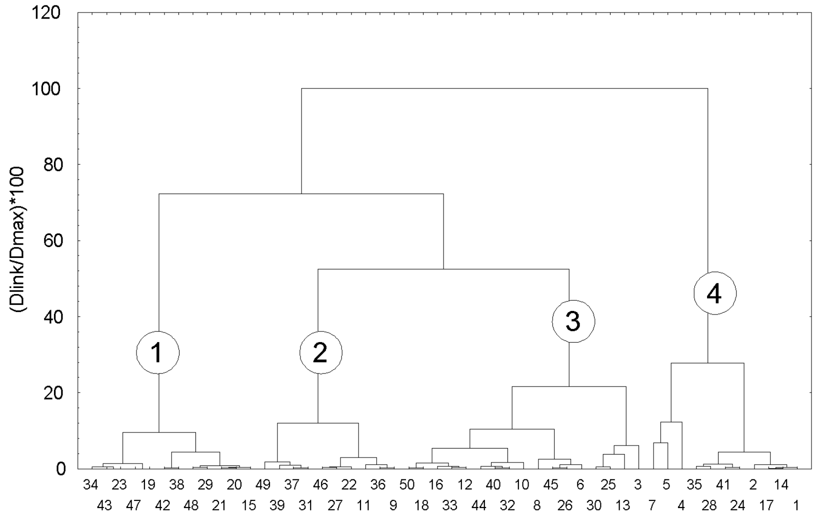

Figure 2. Cluster analysis of the top 50 productive authors based on total publications
Note: Ward's methods with square Euclidean distances
1-Stein, A; 2-Deutsch, CV; 3-Goovaerts, P; 4-Dowd, PA; 5-Herzfeld, UC; 6-Kitanidis, PK; 7-Pardo-Iguzquiza, E; 8-Journel, AG; 9-Webster, R; 10-Yeh, TCJ; 11-McBratney, AB; 12-Christakos, G; 13-Lark, RM; 14-Oliver, MA; 15-Cressie, N; 16-Wen, XH; 17-Atkinson, PM; 18-Bardossy, A; 19-Van Meirvenne, M; 20-Delay, F; 21-Marcotte, D; 22-Myers, DE; 23-Bogardi, I; 24-Caers, J; 25-Carroll, SS; 26-Desbarats, AJ; 27-Fogg, GE; 28-Gneiting, T; 29-Gomez-Hernandez, JJ; 30-Kyriakidis, PC; 31-Neuman, SP; 32-SCHOTZKO, DJ; 33-Fuentes, M; 34-Gertner, G; 35-Hu, LY; 36-Istok, JD; 37-Liebhold, AM; 38-Noetinger, B; 39-Odeh, IOA; 40-Rossi, JP; 41-Vargas-Guzman, JA; 42-Vieira, SR; 43-Anderson, AB; 44-Castrignano, A; 45-Chappell, A; 46-de Marsily, G; 47-McKenna, SA; 48-POSA, D; 49-Rubin, Y; 50-Wang, GX

Table 2. Classification matrices of authorship, country/region, institute and journal using backward DA

Objective	% Correct	A	B	C	D	Statistical tests summary	
(a) Author variation							
A	100	12	0	0	0	Wilks' λ	0.09097
B	90.0	0	9	1	0	F	F (9, 107) = 19.99
C	88.2	1	1	15	0	p	< 0.0000
D	90.9	0	0	1	10	DV [a]	MIF, ACCP, and FP
Total	92.0	13	10	17	10		
(b) Country/region variation[b]							
A	95.2	20	0	1	--	Wilks' λ	0. 1113
B	100	0	8	0	--	F	F (6, 88) = 29.31
C	95.0	0	1	19	--	p	< 0.0000
Total	95.9	20	9	20	--	DV	MIF, ACCP, and IP
(c) Institute variation							
A	89.5	17	2	0	--	Wilks' λ	0.1121
B	85.7	4	24	0	--	F	F (8, 88) = 21.86
C	100	0	0	3	--	p	< 0.0000
Total	88.0	21	26	3	--	DV	MIF, ACCP, SP, and CP
(d) Journal variation							
A	100	2	0	0	0	Wilks' λ	0.04122
B	100	0	17	0	0	F	F (9, 107) = 32.26
C	95.8	0	1	23	0	p	< 0.0000
D	100	0	0	0	7	DV	TC, IF, and ACCP
Total	98.0	2	18	23	7		

[a] DV is short for discriminant variables identified by backward DA.
[b] not including USA when performing DA.

Country/region pattern

Author address information was not available for 31 (1.2%) articles in the ISI. Thus, only 2,660 articles were included in this part of the analysis. Ninety-three countries or regions took part in geostatistics research; 2,162 (81.3%) papers were produced independently by 63 countries/regions, while 498 (18.7%) papers were the product of international collaborations involving 86 countries/regions. The most productive country in the sample was the USA (Table 3), followed distantly by France, UK, Canada, Germany, Spain, Netherlands, Italy, Australia, Brazil, Belgium, China, Sweden, Switzerland, and Japan. Moreover, the countries/regions with the highest MIF and ACPP scores for independent and collaborative papers were Switzerland (1.59) and the USA (1.24), and Belgium (1.70) and Sweden (1.45), respectively.

Like the authorship analysis, the top 50 most productive countries/regions was classified based on the four variables MIF, ACPP, IP, and CP. Hierarchical CA generated a dendrogram with four groups at $(D_{link}/D_{max})\times 100 < 60$ (Figure 3). The data set used for the DA did not include Group 4 because this group comprised a single object, USA.

Table 3. Most productive countries or regions (total, single and co-operation) in geostatistics research from 1967–2005

Country/region	TP	Independent			International collaborative		
		IP	MIF	ACPP	CP	MIF	ACPP
USA	1105	862	1.39	1.24	243	1.39	1.08
France	239	144	1.21	0.84	95	1.38	1.05
UK	228	148	1.18	0.86	80	1.37	1.33
Canada	212	139	0.98	0.79	73	1.25	0.97
Germany	139	82	0.95	0.63	57	1.10	0.73
Spain	128	92	1.35	0.80	36	1.15	0.96
Netherlands	117	75	1.24	0.88	42	1.57	1.24
Italy	112	78	1.23	0.52	34	1.46	0.94
Australia	95	57	1.09	1.10	38	1.08	1.18
Brazil	59	38	0.53	0.22	21	1.17	0.79
Belgium	54	29	1.36	1.01	25	1.70	1.06
China	48	27	1.06	0.39	21	1.51	0.57
Sweden	46	27	1.41	0.48	19	1.54	1.45
Switzerland	46	21	1.59	0.70	25	1.86	1.00
Japan	39	25	1.10	0.35	14	1.19	0.90

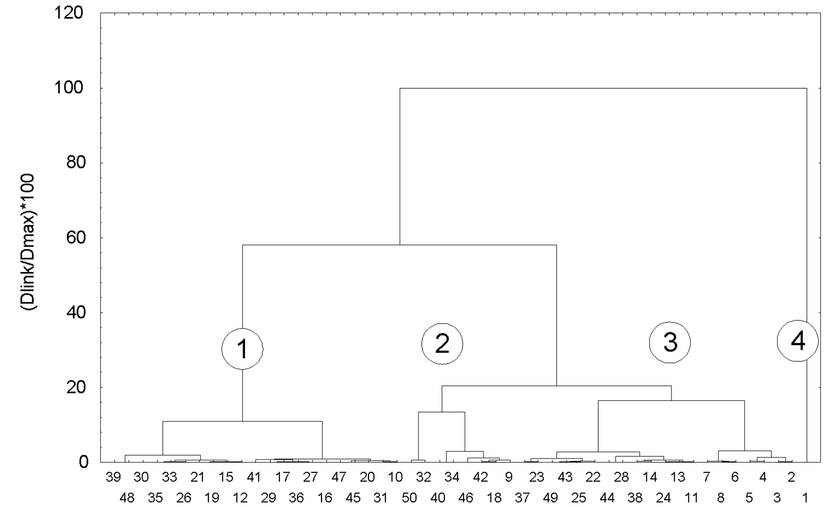

Figure 3. Cluster analysis of the top 50 productive countries/regions based on total publications.
Note: Ward's methods with square Euclidean distances
1-USA; 2-France; 3-UK; 4-Canada; 5-Germany; 6-Spain; 7-Netherlands; 8-Italy; 9-Australia; 10-Brazil; 11-Belgium; 12-China; 13-Sweden; 14-Switzerland; 15-Japan; 16-India; 17-Turkey; 18-Austria; 19-Norway; 20-South Africa; 21-Taiwan; 22-Mexico; 23-Portugal; 24-Denmark; 25-New Zealand; 26-Greece; 27-Saudi Arabia; 28-Israel; 29-Russia; 30-South Korea; 31-Chile; 32-Poland; 33-Hungary; 34-Czech Republic; 35-Finland; 36-Iran; 37-Ireland; 38-Colombia; 39-Morocco; 40-Argentina; 41-Kuwait; 42-Pakistan; 43-Philippines; 44-Venezuela; 45-Slovakia; 46-Kenya; 47-Nigeria; 48-Cameroon; 49-Cuba; 50-Martinique

The value of Wilks' λ, the F-value, and the p-level were 0.1113, 29.31, and <0.0000, respectively. Moreover, this DA produced two DFs that made correct assignments 95.9% of the time using three discriminant parameters (MIF, ACCP, and IP). The countries/regions in group 1 had the lowest MIF and ACPP values but high IP and CP values, since the publications from these countries had the lowest IFs and were cited the least; group 2 had the highest ACPP values, since their publications were the most frequently cited, while Group 3 had high MIF, FP, and RP scores but low ACPP values, since their publications had the greatest IFs. Undoubtedly, the USA was the leading country in geostatistics research.

Institute pattern

Table 4 provides a list of the top 10 most productive Institutes for the period between 1971 and 2005. Nine of these Institutes were located in USA while another one was in Canada. A total of 1,789 Institutes participated. Moreover, 1,368 (51.4%) papers were independently produced by 610 Institutes while 1,292 (48.6%) papers were cooperatively produced by 1,507 Institutes. Stanford University had the highest publication frequency for the period. However, the highest MIF for independent and collaborative papers was achieved by the University of Michigan (1.64 and 2.05), while the highest ACPPs for independent and collaborative publications were achieved by the University of North Carolina (2.01) and the University of Arizona (1.77), respectively.

Table 4. Most productive institutes (total, single and co-operation) in geostatistics research from 1967–2005

Institute	Country/ region	TP	Independent			International collaborative		
			IP	MIF	ACPP	CP	MIF	ACPP
Stanford University	USA	91	52	0.92	0.95	39	1.47	1.767
University of Arizona	USA	60	20	1.57	1.84	40	1.47	1.771
USDA Agricultural Research Center	USA	51	13	1.35	1.22	38	1.53	1.60
University of Michigan	USA	36	16	1.64	1.08	20	2.05	0.99
University of California, Davis	USA	35	12	1.23	0.64	23	1.44	1.17
University of Alberta	Canada	30	11	0.72	0.28	19	1.08	0.75
University of Leeds	UK	30	19	0.50	0.48	11	0.93	0.83
University of Nebraska	USA	29	13	1.12	1.39	16	1.54	0.86
University of Illinois	USA	28	5	1.47	1.34	23	1.46	0.83
Texas A&M University	USA	27	6	1.09	1.96	21	1.19	0.91

We used a hierarchical CA to classify the top 50 most productive institutes based on the MIF, ACPP, IP, and CP. This analysis produced a dendrogram with three groups at $(D_{link}/D_{max}) \times 100 < 75$ (Figure 4). According to a backward DA (Table 2c), the value of Wilks' λ, the F-value, and the p-level were 0.1121, 21.86, and <0.0000, respectively. These results confirmed the validity of the DA. Moreover, this DA produced two DFs with nearly 88.0% correct assignment based on four discriminant parameters (MIF, ACCP, IP, and CP). These Institutes in group 1 had high CP values but low MIF and ACPP values, since the publications from these Institutes had the lowest IFs and were cited the least; while group 3 had the highest MIF, ACPP, IP, and CP scores, since these were the most significant leading institutes; compared to those of group 1, group 2 had higher MIF and ACPP scores, as the publications from these organizations had the highest IFs and were most frequently cited.

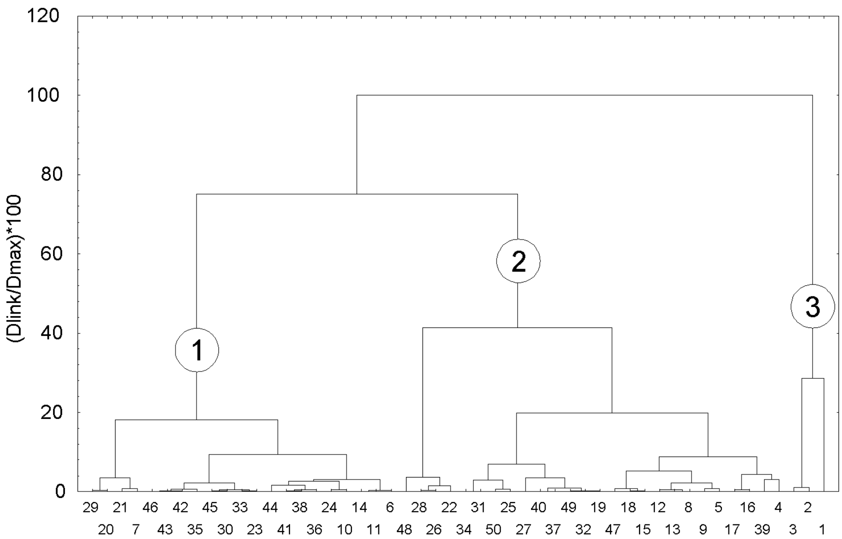

Figure 4. Cluster analysis of the top 50 productive institutes based on total publications
Note: Ward's methods with square Euclidean distances
1-Stanford University; 2-University of Arizona; 3-USDA Agricultural Research Center; 4-University of Michigan; 5-University of California, Davis; 6-University of Alberta; 7-University of Leeds; 8-University of Nebraska; 9-University of Illinois; 10-Texas A&M University; 11-Chinese Academy of Sciences; 12-Consejo Superior De Investigaciones Científicas (CSIC); 13-I'institut National De La Recherche Agronomique (INRA); 14-University of Texas; 15-Universities of North Carolina; 16-US Geology Survey; 17-Oregon State University; 18-University of Paris 06; 19-Wageningen University; 20-Ecole Polytechnique; 21-Institut Francais du Petrole; 22-Michigan State University; 23-Ohio State University; 24-University of Colorado; 25-University of Sydney; 26-University of California, Berkeley; 27-University of Idaho; 28-Washington State University; 29-Chevron Petroleum Technology Corporation; 30-National Taiwan University; 31-University of Washington; 32-University Waterloo; 33-US Environmental Protection Agency; 34-Ecole des mines de Paris; 35-Pennsylvania State University; 36-University of British Columbia; 37-University of Minnesota; 38-University of Nevada; 39-University of Wisconsin; 40-US Forest Service; 41-Centre National De La Recherche Scientifique (CNRS); 42-University of Reading; 43-University of Southampton; 44-University of Trier; 45-University of Wageningen & Research Center; 46-Bureau de Recherches Géologiques et Minières; 47-Sandia National Laboratories; 48-University of California, Santa Barbara; 49-University of Utrecht; 50-University of Iowa

Journal pattern

In total, 2,691 articles were published in 593 journals during the period from 1972 to 2005. Table 5 lists the top 10 journals for each of four criteria: TP, TC, IF, and ACPP. *Mathematical Geology*, which published 172 articles, was the most popular journal in the sample, followed by *Water Resources Research* (133) and *Geoderma* (68).

Table 5. The journals of the top 10 number, citation, impact factor of publication

Journal	TP	TC	IF*	ACPP
Top 10 most productive journals (a)				
Mathematical Geology	172	1293	0.75	0.71
Water Resources Research	133	3017	1.94	2.06
Geoderma	68	659	1.77	1.39
Journal of Hydrology	68	692	1.75	1.31
Computers & Geosciences	66	262	0.78	0.62
Soil Science Society of America Journal	50	824	1.83	1.53
Ground Water	45	346	1.42	0.89
International Journal of Remote Sensing	36	236	0.93	0.88
Environmetrics	34	125	0.77	0.49
Stochastic Environmental Research and Risk Assessment	29	70	0.70	0.52
Top 10 most cited journals (b)				
Water Resources Research	133	3017	1.94	2.06
Mathematical Geology	172	1293	0.75	0.71
Soil Science Society of America Journal	50	824	1.83	1.53
Journal of Hydrology	68	692	1.75	1.31
Geoderma	68	659	1.77	1.39
Journal of Applied Meteorology	10	557	1.70	4.81
Ecology	15	436	4.51	3.51
Ecological Monographs	1	409	4.86	29.21
Journal of Ecology	5	380	4.28	6.12
Ground Water	45	346	1.42	0.89
Top 10 highest IF journals (c)				
Trends in Ecology & Evolution	1	57	14.86	7.13
Proceedings of the National Academy of Sciences of the USA	1	126	10.23	14.0
FEMS Microbiology Reviews	1	21	10.00	2.33
Environmental Health Perspectives	4	5	5.34	1.13
Ecology Letters	1	19	5.15	4.75
American Journal of Epidemiology	3	62	5.07	5.42
BMC Bioinformatics	1	4	4.96	2.00
Ecological Monographs	1	409	4.86	29.21
Journal of Applied Ecology	2	29	4.59	2.07
Biostatistics	1	9	4.53	3.00

* According to *Journal Citation Report* (2005)

The most frequently cited journal was *Water Resources Research*, which had 3,017 citations during the period; however, *Ecological Monographs* had the highest ACPP score (29.21). In the area of geostatistics research, the journal with the highest IF was

Trends in Ecology & Evolution (14.86), followed by the *Proceedings of the National Academy of Sciences of the United States of America* (10.23) and *FEMS Microbiology Reviews* (10.00).

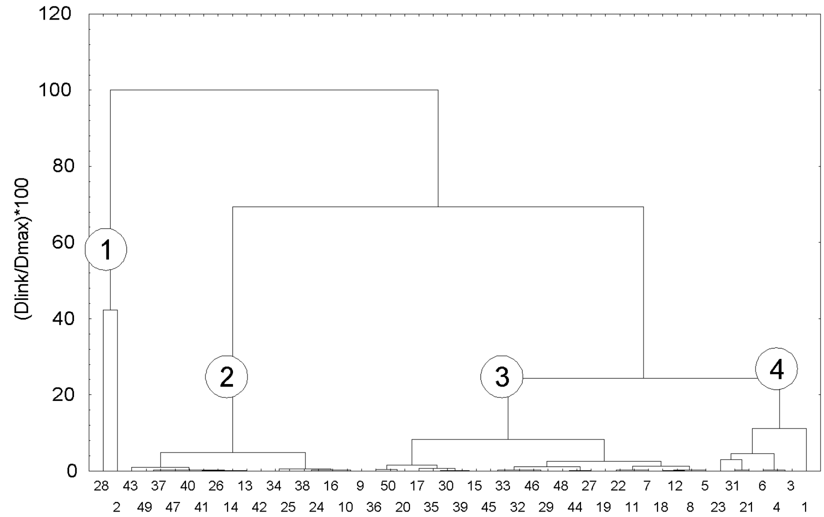

Figure 5. Cluster analysis of the top 50 productive journals based on total publications
Note: Ward's methods with square Euclidean distances
1-*Mathematical Geology;* 2-*Water Resources Research;* 3-*Geoderma;* 4-*Journal of Hydrology;* 5-*Computers & Geosciences;* 6-*Soil Science Society of America Journal;* 7-*Ground Water;* 8-*International Journal of Remote Sensing;* 9-*Environmetrics;* 10-*Stochastic Environmental Research and Risk Assessment;* 11-*Environmental Entomology;* 12-*Soil Science;* 13-*Spe Reservoir Evaluation & Engineering;* 14-*CIM Bulletin;* 15-*Advances in Water Resources;* 16-*Environmental Geology;* 17-*Phytopathology;* 18-*Geophysics;* 19-*Hydrological Processes;* 20-*Journal of Environmental Quality;* 21-*Plant and Soil;* 22-*Ices Journal of Marine Science;* 23-*Remote Sensing of Environment;* 24-*Journal of Petroleum Science and Engineering;* 25-*Petroleum Geoscience;* 26-*Transactions of the Institute of Mining and Metallurgy Section A-Mining Industry;* 27-*Soil & Tillage Research;* 28-*Ecology;* 29-*Forest Ecology and Management;* 30-*Journal of Contaminant Hydrology;* 31-*Weed Science;* 32-*Ecological Modelling;* 33-*Environmental and Ecological Statistics;* 34-*Environmental Monitoring and Assessment;* 35-*Geophysical Journal International;* 36-*Geophysical Research Letters;* 37-*Revista Brasileira De Ciencia Do Solo;* 38-*Communications In Soil Science and Plant Analysis;* 39-*International Journal of Climatology;* 40-*Journal of Petroleum Technology;* 41-*Journal of the International Association for Mathematical Geology;* 42-*Spe Journal;* 43-*Stochastic Hydrology and Hydraulics;* 44-*Water Air and Soil Pollution;* 45-*AAPG Bulletin-American Association of Petroleum Geologists;* 46-*Agronomy Journal;* 47-*E&MJ-Engineering and Mining Journal;* 48-*Fisheries Research;* 49-*Journal of Canadian Petroleum Technology;* 50-*Journal of Geophysical Research-Atmospheres*

We also used a CA to group the top 50 most productive journals based on the three variables TC, IF, and ACCP. This analysis generated a dendrogram with three groups at $(D_{link}/D_{max}) \times 100 < 70$ (Figure 5). The value of Wilks' λ, the F-value, and the p-level determined by the DA were 0.04122, 32.26, and <0.0000, respectively (Table 2d).

Moreover, this DA produced three DFs with nearly 98.0% correct assignment based on three discriminant parameters (TC, IF, and ACCP). The journals in group 1 had the highest TC, IF, and ACPP scores, which were the top journals during the period; compared to group 1, group 2 had lower TC, IF, and ACPP scores, represented relatively ordinary journals; group 4 had high TC and ACPP scores, as their articles were the most frequently cited; and group 3 contained the remaining 24 journals, which had high IFs but low ACPP scores. The articles in these journals had the greatest IFs.

Conclusions

In this study, we successfully applied multivariate methods to the scientometric analysis of worldwide geostatistics research published during the period from 1967 to 2005. Our analysis indicated that CA and DA were effective approaches for identifying patterns. Moreover, we applied a hierarchical CA to classify the authors, countries/regions, Institutes, and journals of the publications considered into four groups, which may have facilitated a comprehensive evaluation. Based on these results, we applied a backward DA with appropriate statistical tests that demonstrated good discriminatory ability; we used its high level of correct assignment to test the CA results and identify significant parameters for the identification of each pattern. With regard to general patterns, we found that the document type "article," the language English, and the field of Geosciences were predominant components. In addition, more than 81% of the articles were published during the period from 1995 to 2005. Finally, the ACPP was more useful than the TC for uncovering peak citation trends.

*

This paper was funded by the China Scholarship Programs (2006100766) for F. Zhou and the National Basic Research Program (No.2005CB724205). The authors would like to thanks to anonymous reviewers for their valuable comments.

References

BARABAS, N., GOOVAERTS, P., ADRIAENS, P. (2001), Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. *Environmental Science & Technology*, 35 : 3294–3301.

CHIU, W. T., HO, Y. S. (2005), Bibliometric analysis of homeopathy research during the period of 1991 to 2003. *Scientometrics*, 63 : 3–23.

DELHOMME, J. P. (1979), Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach. *Water Resources Research*, 15 : 269–280.

ENACHESCU, C., POSTELNICU, T. (2003), Patterns in journal citation data revealed by exploratory multivariate analysis. *Scientometrics*, 56 : 43–59.

FABA-PEREZ, C., GUERRERO-BOTE, V. P., DE MOYA-ANEGON, F. (2003), Data mining in a closed Web environment. *Scientometrics*, 58 : 623–640.

FERNANDEZ-CANO, A., BUENO, A. (2002), Multivariate evaluation of Spanish educational research journals. *Scientometrics*, 55 : 87–102.

GARFIELD, E. (1970), Citation indexing for studying science. *Essays of an Information Scientist*, 1 : 133–138.

HERZFELD, U. C., HIGGINSON, C. A. (1996), Automated geostatistical seafloor classification – Principles, parameters, feature vectors, and discrimination criteria. *Computers & Geosciences*, 22 : 35–41.

JOHNSON, R. A., WICHERN, D. W. (2002), *Applied Multivariate Statistical Analysis*. 5th edition. Prentice-Hall, New Jersey.

KRIEGER, N., CHEN, J. T., WATERMAN, P. D., SOOBADER, M. J., SUBRAMANIAN, S. V., CARSON, R. (2002), Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? The public health disparities geocoding project. *American Journal of Epidemiology*, 156 : 471–482.

KRIGE, D. G. (1952), A statistical analysis of some of the borehole values in the orange free state gold field. *Journal of the Chemical and Metallurgical Society of South Africa*, 53 : 47–64.

MATHERON, G. (1963), Principles of geostatistics. *Economic Geology*, 58 : 1250–1266.

NWAGWU, W. (2006), A bibliometric analysis of productivity patterns of biomedical authors of Nigeria during 1967–2002. *Scientometrics*, 69 : 259–269.

PETIT, R. J., PINEAU, E., DEMESURE, B., BACILIERI, R., DUCOUSSO, A., KREMER, A. (1994), Field-scale variability of soil Properties in Central Iowa soils. *Soil Science Society of America Journal*, 58 : 1501–1511.

ROSSI, R. E., MULLA, D. J., JOURNEL, A. G., FRANZ, E. H. (1992), Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 62 : 277–314.

SICHEL, H. S. (1952), New methods in the statistical evaluation of mine sampling data. *Transactions of the Institution of Mining and Metallurgy*.

WACKERNAGEL, H. (2003), *Multivariate Geostatistics: An Introduction with Applications*. Springer, Berlin.

ZHOU, F., LIU, Y., GUO, H. C. (2006), Application of multivariate statistical methods to the water quality assessment of the watercourses in the Northwestern New Territories, Hong Kong. *Environmental Monitoring and Assessment*, DOI 10.1007/s10661-006-9497-x.