INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

REPRESENTING CANADIAN GEOPHYSICS : A BIBLIOMETRIC APPROACH

R. TODOROV

Centre for Scientific Information,
Bulgarian Academy of Sciences
BG-1040 Sofia, 7 Noemvri St. No 1

Abstract

Various bibliometric methods are used to obtain an alternative (quantitative) representation of Canadian publication activity in geophysics. First, an online technique for data gathering is applied to create appropriate sets of document records from the 1988 INSPEC file and to count selected bibliographic characteristics by using directly the software implemented in the host system. Then, the appearance and co-appearance of classification subdivisions (headings) in the chosen records are utilized to describe the subject structure of Canadian geophysics in terms of headings and their links. Multivariate technique is applied to display the results in the form of two-dimensional inclusion maps.

INTRODUCTION

Online bibliographic services are predominantly used for retrieval of documents which are relevant to a problem or subject area. In bibliometrics they provide possibilities to avoid manual gathering of data, and more specifically to group a great number of records according to selected address or subject characteristics [1,2]. Additionally, advanced software in many host systems (STN, DIMDI, etc.) allows to count elements in different fields and to list them in descending order of appearance [3]. For further analysis, chosen record fields could be downloaded and processed outside of the host system. Thus, in scientometrics, common appearance of selected field elements is determined to study, for example, the subject structure of a given specialty in terms of frequently appearing topics (keywords, citations, etc.) and their links [4,5].

DATA

Bibliographic databases in science and technology store a great number of document records relevant to a given subject area. The INSPEC (Information Services for the Physics and Engineering Communities) file contains records for the worldwide literature on physics (as well as on electricity and electronics, and computer science). Bibliometric data are extracted mainly from fields including identification bibliographic elements (such as Title, Author, Source, Corporate Source, etc.) or subject characteristics (Abstract Text, Classification Codes, Treatment Codes, Controlled and Supplementary Terms). For the purpose of this study the following fields have been selected for online analysis : Corporate Source (CS), Source (SO), Classification (CC), and Treatment Codes (TC).

The field "Corporate Source" (CS) is first used to select all the records of documents that are authored by Canadian geophysicists and then, to determine highly productive institutes. Counting the different titles in the field "Source" (SO) has permitted to rank geophysics journals according to the number of Canadian articles published in them.

The CC in the INSPEC file are used traditionally to retrieve all documents that are assigned to a given classification subdivision. CC stand for subject headings or topics which fall into the following categories : substances or named objects, and their properties; phenomena, processes, effects; theories, methods, and models; instruments and techniques. Here, the CC for (or headings of) geophysics which appear in Canadian records have been downloaded and then arranged according to their frequency of occurrence. (Separately, the common appearance of pairs of headings has been determined by using an appropriate "external" software).

Treatment Codes (TC) indicate particular research characteristics of the subject treated in a given document (theoretical, applied, practical, etc.). The different TC in 1988 Canadian geophysics records have been counted and used to partition Canadian papers into two main subsets : experimental and theoretical ones.

METHOD

For the purpose of this study all 1988 INSPEC records containing a truncated code from geophysics have been retrieved. 504 of them contained a Canadian institute in the CS-field. All the elements in the fields CC, TC, and SO of these 504 records have been separately listed in descending order of occurrence by using the software implemented in the host system.

As in co-word analysis the common occurrence or co-occurrence of CC (in the set of 504 selected document records) is taken to express the subject link between two headings. More precisely, two classification codes C_i and C_j are said to co-occur if there is at least one record containing both C_i and C_j. An appropriate software has been developed to count the co-appearances of the CC and to place them as off-diagonal values c_{ij} in a triangular matrix $C = \{c_{ij}\}$. (A diagonal element c_{ii} of C is the absolute frequency of appearance of C_i). Then, an appropriate measure of similarity between the CC has been selected and calculated using the values from the matrix C. Here, preference is given to the Inclusion index (I) from co-word analysis [6] with $I = c_{ij}/\min(c_{ii}, c_{jj})$ as the C_i are elements of a (hierarchical) classification scheme [7]. Once this measure has been determined the final step consists in selecting a technique to reveal the links between the C_i. Multidimensional scaling (MDS) is applied to display closeness of headings through their spatial configuration within a two dimensional representation [8]. Since a representation of two dimensions is usually not enough to accommodate all interrelationships, an additional information from hierarchical cluster analysis (HCA) has been introduced on the map by drawing curves around the headings that belong to the same cluster [9]. Additionally, relation strength (based upon the inclusion index values) has been shown by connecting highly co-appearing headings (from the lower to the higher frequency of appearance) with corresponding line type.

The described method has been applied on :
- the set of Canadian records in geophysics as a whole,
- two subsets containing experimental and theoretical papers, respectively, and
- three subsets covering the main geophysics subdivisions : solid Earth geo-

physics (SEG), hydrospheric and atmospheric geophysics (HAG), and geophysical observations and techniques (GOT).

RESULTS

The results of using online bibliometrics, i.e. of counting and ranking (with an implemented software in the host system) are shown in Tables 1-4. A list of geophysics headings or topics which occur frequently in Canadian document records is given in Table 1. In Table 2 Canadian geophysics papers are divided in several groups according to the assigned TC. In Table 3 are enumerated some productive Canadian institutes in the field of geophysics in descending order of published papers. The distribution of Canadian articles over journal titles is shown in Table 4.

As mentioned above, six different sets of Canadian records have been formed and analyzed. Figures 1-6 represent inclusion maps which are the result of application of MDS and HCA on I-matrices. On Figure 1 is shown the structure of Canadian geophysics as a whole : frequently appearing headings are grouped in clusters and interrelated according to the value of the inclusion index. Experimental and theoretical links between topics are revealed by Figures 2 and 3, respectively. Figures 4-6 describe with more details the subject structures of the three main geophysics subfields. All plot symbols are explained in Table 1.

DISCUSSION

Host systems in science usually offer various possibilities for searching in different databases. In bibliometrics a great number of records could be selected and sorted according to their address and subject characteristics with the software implemented in the system. Here, an online method is used to select all Canadian records in geophysics from the 1988 INSPEC file. Then, a sorting function has been applied on this set and several lists of frequently occurring characteristics (headings, author affiliations, source journals, etc.) obtained. Thus, different aspects of Canadian publication activity in geophysics have been detected and displayed. For example, 32 per cent of all Canadian papers in geophysics deal explicitly with North America (see Table 1). It appears that Canadian publications cover all subfields of geophysics with large use of instrumentation and computer technique for measurements as well as for data acquisition and processing. Canadian papers in geophysics are mainly experimental (50 %), theoretical (37 %), and to lesser extent practical (22 %). The sum does not add to 100 % as many papers are assigned to more than one TC (see Table 2). Canadian papers are published mainly by institutes in Ontario (43 %), Nova Scotia (16 %), British Columbia (13 %), and Quebec (11 %). More than 40 % of the papers originate from the 16 institutes listed (in descending order) in Table 3. About 80 % of the papers are published in journals. 238 journal articles (57 %) are distributed over 23 (17 %) titles (see Table 4). The online lists of corporate sources and journals have been additionally edited as they contain abbreviations, spelling variants, and errors, i.e. Tables 3 and 4 are not directly obtained. Most of the articles are in 'visible' and high impact journals published in Canada, USA, UK, and the Netherlands.

Document records in the INSPEC file are assigned to different subdivisions (or headings) according to the International Classification for Physics (ICP) [10]. Each record is allocated on an average to three headings. The co-appearance of such headings in a set of preselected document records is used to describe closeness of topics in Canadian geophysics publications. Thus, a network is created, i.e. an alternative topology to the classification cross-references (attached by experts) could be introduced. In practice, an appropriate software is used to construct a matrix of co-appearance. On the diagonal are placed the frequencies of appearance of the CC in a descending order. The size of this matrix is equal to the number of different codes occurring in the selected records. With a view to a more legible presentation, the initial size is reduced by introducing a threshold c_t ($c_{ii} > c_t$). The practical criteria for determining c_t are the amount of information lost and the number of headings (and their links) to be represented. In Table 5 are given several values for c_t and the corresponding percentage of information lost in case all Canadian publications in geophysics are considered.

The selection of a similarity measure is a compromise between adequate representation of central (high appearance) and peripheral topics. More indices could be taken under consideration : Jaccard index, Strength index, Inclusion index, Proximity index, etc. Here, the Inclusion index has been selected as it preserves high values even when c_{jj} is low and c_{ii} is high.

Once the similarity measure is determined one should select a technique of structure displaying. The choice depends exclusively on the purpose of the study and the final application of results. MDS (a two-dimensional solution from ALSCAL) [11] has been used to visualize the structure of Canadian geophysics. Two-dimensional solutions appear to be simple and easily perceived. However, two dimensions are, in general, not enough to reflect adequately all links. In such cases, it is useful to include the result of HCA as a complementary information about closeness of topics. Thus, related topics that appear far apart on the map could be encircled, i.e. brought together in one cluster. Here the criterion of average linkage between groups is applied in the clustering process. As the choice of a clustering method implicitly imposes a structure on the set of headings it is desirable to "validate" the results.

Additionally, relation strength between headings (based upon the inclusion index values) is shown by connecting highly co-appearing headings (from the lower to the higher frequency of appearance) with a corresponding line type : two threshold values for the inclusion index (0.8 and 0.6) are introduced and indicated respectively by solid and dashed lines. A good representation is achieved if links are concentrated mainly between topics within the same cluster.

Further improvement of the map could be attained by disclosing the size of the topics according to their absolute frequency of occurrence. For reasons of map legibility this information is not shown on the maps.

INTERPRETATION OF THE INCLUSION MAPS

Preliminary interpretation (without expert participation) of the maps involves label selection for the clusters, as well as a description of some (common) features related to the formal structure of headings. Thus, the map of Canadian geophysics (Figure 1) includes

- two SEG clusters : "Structure studies of crust in North America and Geochronology" (1,3,7,8,14,17,24,25,26), and "Hydrology and other SEG topics" (20,29,30);
- two clusters covering physics of the ocean, namely "Dynamics of the upper ocean" (11,12,15,21,23) and "Coastal processes" (5,31,36);
- two GOT clusters : "Computer technique and Oceanographic measurement" (4,6,34) as well as "Instrumentation and Geometrical observation" (2,13,19);
- three meteorology clusters : "Winds and their effects" (9,35), "Weather analysis" (10,16,28), and "Atmospheric studies" (22,27).

By selecting records with the same treatment code a network with a given research characteristic has been created. Thus, experimental and theoretical networks (see Figure 2 and 3) reveal slightly different structures. The experimental map is characterized by a splitting of the main SEG cluster and a combination of the two GOT clusters (see Figure 2). Clusters like "Weather analysis" and "Atmospheric studies" as well as "Seismology" (and "Explosion seismology" in particular) appear only on the theoretical map (Figure 3). Thus, some of the links in such clusters could be labeled as purely experimental (3-24, 22-27, etc.) or as theoretical ones (11-12, 9-35, etc.).

A detailed map of geophysics subfield could be created by selecting only records with a truncated code from a given subdivision. Thus, three particular maps of the main geophysics subfields (SEG, HAG, GOT) have been obtained (see Figures 4-6). The map on Figure 4 shows the splitting of the main SEG cluster from Figure 1 and the emergence of a smaller cluster "Explosion seismology". This process could be discerned on the maps of HAG and GOT as well. Thus, on Figure 5 a new cluster "Water resources and limnology" has appeared and a greater one ("Weather analysis") disintegrated. The GOT-map includes, for example, two new clusters related to measurements of "Water in the atmosphere" and "Paleomagnetism", and one which could be labeled as "Studies of thermohaline structure" (see Figure 6).

CONCLUSION

This study is based on online bibliometrics and a more sophisticated co-occurrence method. It has been designed to show some capacities of such combined approach, namely to describe "quickly" publication activity of Canadian geophysicists and mainly the subject structure of the papers in terms of headings or classification subdivisions and their links. These results could help information scientists (or people with no or little substantive knowledge) to select appropriate titles for journal collections as well as to improve classification schemes by introducing 'objective' cross-references. Lists of ranked headings, institutes or journals could be used for creating indices to meet the needs of an alternative view in such areas as science policy and evaluation.

The structure representation is based on a co-occurrence model which is introduced here outside of a conceptual framework [12], and without a special analysis of the "indexer effect" and some of the technical limitations [13]. In this sense, the obtained results should be considered critically and taken as preliminary. Furthermore, they should be used in conjunction with expert judgements and other co-occurrence methods (co-citation and co-word analyses) [14].

REFERENCES

[1] O. Persson, Online Bibliometrics. A Research Tool for Every Man. Scientometrics, 10(1986) 1-2, 69-75.

[2] R. Todorov, Representing a Scientific Field : A Bibliometric Approach. Scientometrics, 15(1989) No 5-6, 589-601.

[3] O. Persson, Measuring Scientific Output by Online Techniques, In : A.F.J. Van Raan (Ed.), Handbook of Quantitative Studies of Science and Technology, North-Holland, Amsterdam (1988), 229-252.

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

[5] J.J. Franklin and R. Johnston, Co-Citation Bibliometric Modeling as a Tool for S & T Policy and R & D Management : Issues, Applications, and Developments, In : A.F.J. Van Raan (Ed.), Handbook of Quantitative Studies of Science and Technology, North-Holland, Amsterdam (1988), 325-389.

[6] M. Callon, J. Law and A. Rip, Qualitative Scientometrics, In : M. Callon, J. Law, A. Rip (Eds.), Mapping the Dynamics of Science and Technology, Mcmillan, London (1986), p.112.

[7] J.-P. Courtial, Technical issues and developments in methodology, In : M. Callon, J. Law, A. Rip (Eds.), Mapping the Dynamics of Science and Technology, Macmillan, London (1986), 189-193.

[8] R.N. Shepard, Introduction, In : R.N. Shepard, A.K. Romney, S.B. Nerlove (Eds.), Multidimensional scaling : Theory and applications in the behavioral sciences, Vol.1, Theory, New York : Seminar Pr., 1973, p.1-20.

[9] R.N. Shepard, A taxonomy of some principal types of data and of multidimensional methods for their analysis, In : R.N. Shepard, A.K. Romney, S.B. Nerlove (Eds.), Multidimensional scaling : Theory and applications in the behavioral sciences. Vol.I, Theory, New York : Seminar Pr., 1973, pp.21-47.

[10] A. Berthelot, P. Clague, S. Schiminovich and W. Zwirner, The ICSU AB International Classification System for Physics : Its History and Future. J. Am. Soc. Inf. Sci. 1979, 30, 343-352.

[11] Y. Tacane, F.W. Young, J. Deleeuw, Nonmetric individual differences multidimensional scaling : An alternating least squares method with optimal scaling features. Psychometrika 42(1977), 7-67.

[12] A. Rip, Mapping of Science : Possibilities and Limitations, In : A.F.J. Van Raan (Ed.), Handbook of Quantitative Studies of Science and Technology, North-Holland, Amsterdam (1988), 253-273.

[13] J.E.J. Oberski, Some statistical aspects to co-citation cluster analysis and a judgement by physicists, *In : A.F.J. Van Raan (Ed.), Handbook of* Quantitative Studies of Science and Technology, North-Holland, Amsterdam (1988), 431-462.

[14] R.R. Braam, H.F. Moed and A.F.J. Van Raan, Comparison and Combination of Co-Citation and Co-Word Clustering, In : A.F.J. Van Raan, A.J. Nederhof and H.F. Moed (Eds.), Science and Technology Indicators, DSWO Press, University of Leiden (1989), 307-337.

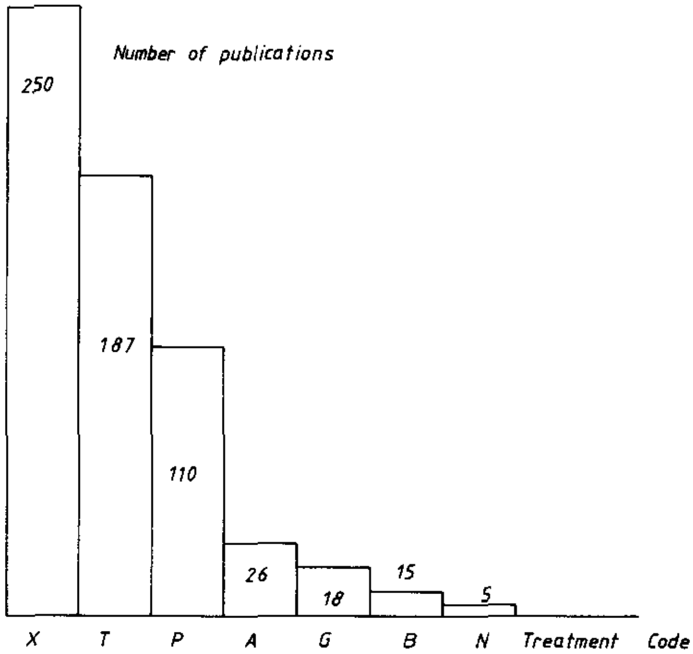
Table 1 : Topics (headings) of 1988 Canadian Geophysics arranged according to their occurrence

Plot symbol	Heading	Frequency of occurrence
1	North America	162
2	Instrumentation and techniques	152
3	Structure of crust and upper mantle	60
4	Computer techniques	47
5	Atlantic Ocean	42
6	Oceanographic and hydrographic measurement	42
7	Geochronology	40
8	Geophysical aspects of geology	40
9	Winds and their effects	33
10	Water in the atmosphere	32
11	Dynamics of the upper ocean	32
12	Coastal and estuarine oceanography	32
13	Geophysics instruments and equipment	30
14	Large islands	30
15	Surface waves, tides, and sea level	25
16	Weather analysis and prediction	24
17	Physics of plate tectonics	23
18	Explosion seismology	23
19	Topography : geometrical observations	23
20	Other topics in solid Earth geophysics	23
21	Thermohaline structure and circulation	22
22	Atmosphere	21
23	Regional seas	21
24	Composition of Earth's interior	21
25	Paleomagnetism	20
26	Plate Tectonics	19
27	Air quality and air pollution	19
28	Climatology	18
29	Erosion and sedimentation	17
30	Rivers, runoff, and streamflows	16
31	Beach, coastal and shelf processes	16
32	Sea ice	16
33	Underwater sound	16
34	Data acquisition and storage	16
35	Boundary layer structure and processes	15
36	Turbidity currents : sedimentation	15
37	Earth's electricity	14
38	Water quality and water resources	14
39	Water	13
40	General circulation	13
41	Glaciers	13
42	Heat flow; geothermy	13
43	Arctic Ocean	13
44	Limnology	12
45	Physical properties of seawater	12
46	Radar systems and equipment	12
47	Interactions between exterior sources and interior properties	12
48	Europe	11
49	Volcanology	11
50	Storms	11
51	Cloud physics	11
52	Chemical composition and interactions	11
53	Surface and body waves	11

Table 1 (Cont.) :

Plot Symbol	Heading	Frequence of occurrence
54	Seismology	11
55	Ocean bottom processes	10
56	Groundwater	10
57	Meteorology	9
58	Photogrammetry	9
59	Harmonics of gravity potential field	9
60	Physics of the oceans	9

Table 2 : Classification of Canadian geophysics papers according to their Treatment Code (TC)



X EXPERIMENTAL	the document describes an experimental method, observation or result
T THEORETICAL	the treatment of the subject is of a theoretical nature
P PRACTICAL	the document is intended to be of direct practical use and is likely to be of interest particularly to engineering and design staff
A APPLICATION	the document describes the actual or potential use of an instrument, device, method or technique, computer program, or a physical effect where some specific application is described or envisaged
G GENERAL	documents including general approaches, overviews and reviews
B BIBLIOGRAPHY	indicates documents with a bibliography or long list of references
N NEW DEVELOPMENTS	a claim of novelty (in the patent sense) is made

Table 3 : Some Productive Sources for Canadian geophysics publications

Bedford Inst. of Oceanogr., Dartmouth, NS
Geol. Survey of Canada, Ottawa, Ont.
Toronto Univ., Ont.
Atmos. Environ. Service, Downsview, Ont.
Dalhousie Univ., Halifax, NS
British Columbia Univ., Vancouver, BC
Calgary Univ., Alta.,
Inst. of Ocean Sci., Sidney, BC
McMaster Univ., Hamilton, Ont.
Alberta Univ., Edmonton, Alta.
McGill Univ., Montreal, Que.
Memorial Univ. of Newfoundland, St. John's, NFLD.
Saskatchewan Univ., Saskatoon, Sask.
Canadian Climate Centre, Downsview, Ont.
Canada Centre for Inland Waters, Burlington, Ont.
Canada Centre for Remote Sensing, Ottawa, Ont.

Table 4 : Distribution of Canadian geophysics articles over journal titles

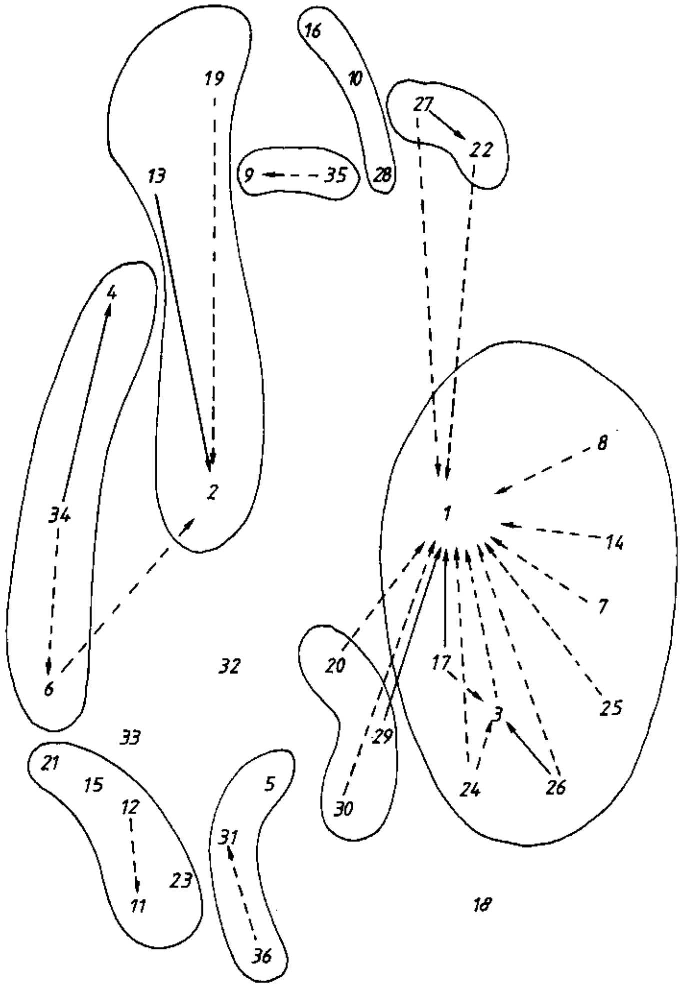
N	Journal Title (Country)
60	Can. J. Earth Sci. (Canada)
56	J. Geophys. Res. (USA)
17	J. Phys. Oceanogr. (USA)
13	Nature (UK)
8	Geophysics (USA)
7	Mon. Weather Rev. (USA)
6	Geology (USA)
6	Geophys. Res. Lett. (USA)
6	Journal of Applied Meteorology
5	Geophys. J. (Blackwell) (UK)
5	Journal of Atmospheric and Oceanic Technology
5	Photogramm. Eng. Remote Sens. (USA)
4	Applied Optics
4	Atmos. Environ. (UK)
4	Bull. Seismol. Soc. Am. (USA)
4	Earth Surf. Process. Landf. (UK)
4	Geophys. Astrophys. Fluid Dyn. (UK)
4	Geophys. Prospect. (Netherlands)
4	J. Gt. Lakes Res. (USA)
4	Photogrammetria (Netherlands)
4	Phys. Earth Planet. Inter. (Netherlands)
4	Tectonics
4	Water Resour. Res. (USA)
19	Journals with 3 articles
30	Journals with 2 articles
65	Journals with 1 article

Table 5 : Dependence of the final display of Canadian geophysics (Fig.1) on the threshold value C_t

C_t	Information lost* (%)	Number of headings to be displayed
1	0	272
2	5	170
3	8	136
4	12	112
5	15	96
6	18	85
7	21	74
8	24	65
9	25	62
10	28	56
11	29	54
12	32	47
13	35	43
14	39	38
15	40	36
16	42	34
22	52	21

* Number of headings not considered (as percentage of the total number of headings used).

Figure 1 : Inclusion Map for 1988 Canadian Geophysics

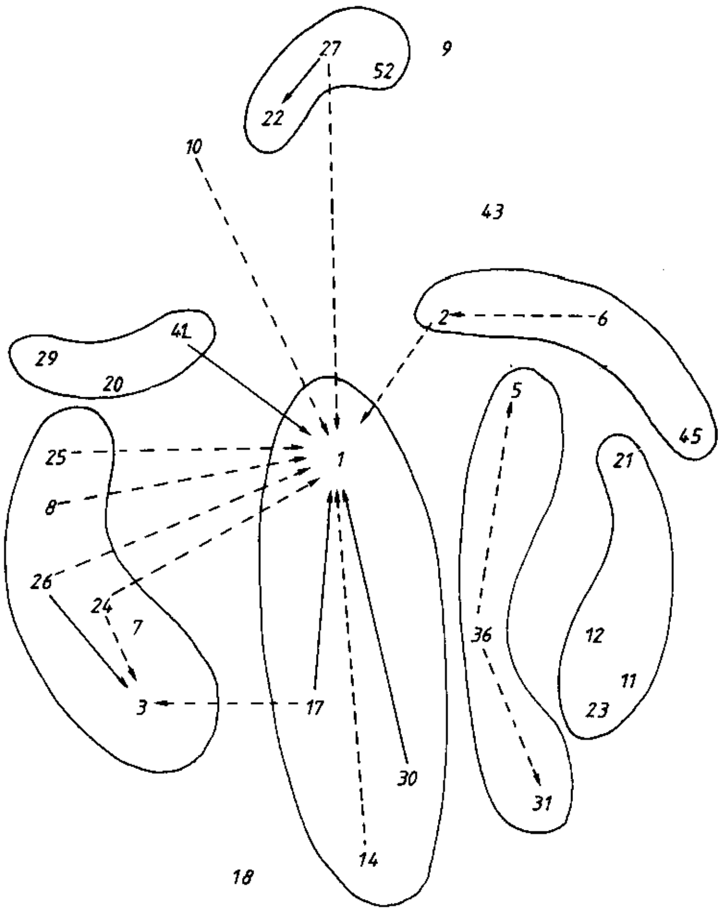


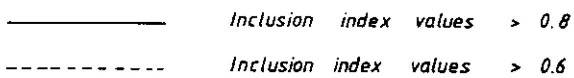
———————— Inclusion index values > 0.8

- - - - - - - - - - Inclusion index values > 0.6

Plot symbols are given in Table 1.

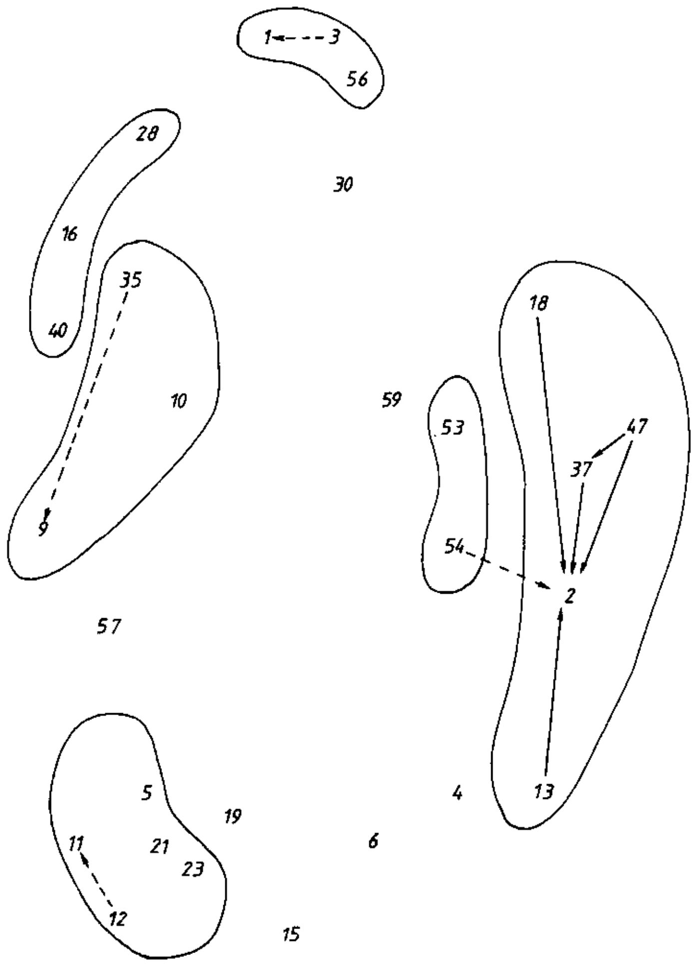
Figure 2 : Inclusion Map for 1988 Canadian Experimental Geophysics





Plot symbols are given in Table 1.

Figure 3 : Inclusion Map for 1988 Canadian Theoretical Geophysics

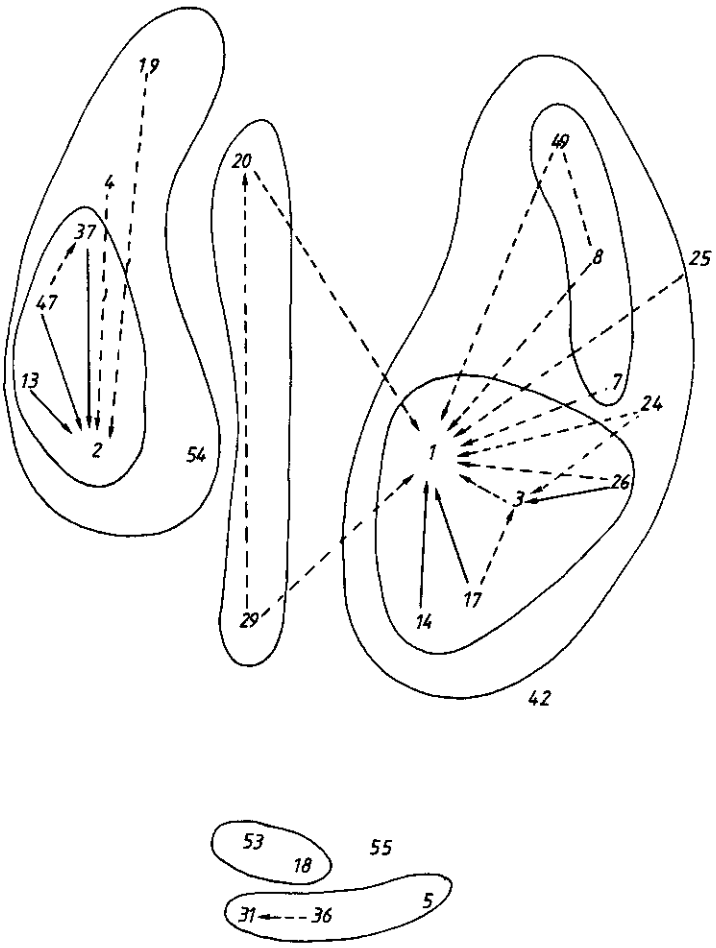


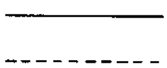
Inclusion index values > 0.8

Inclusion index values > 0.6

Plot symbols are given in Table 1.

Figure 4 : Inclusion Map for 1988 Canadian Solid Earth Geophysics



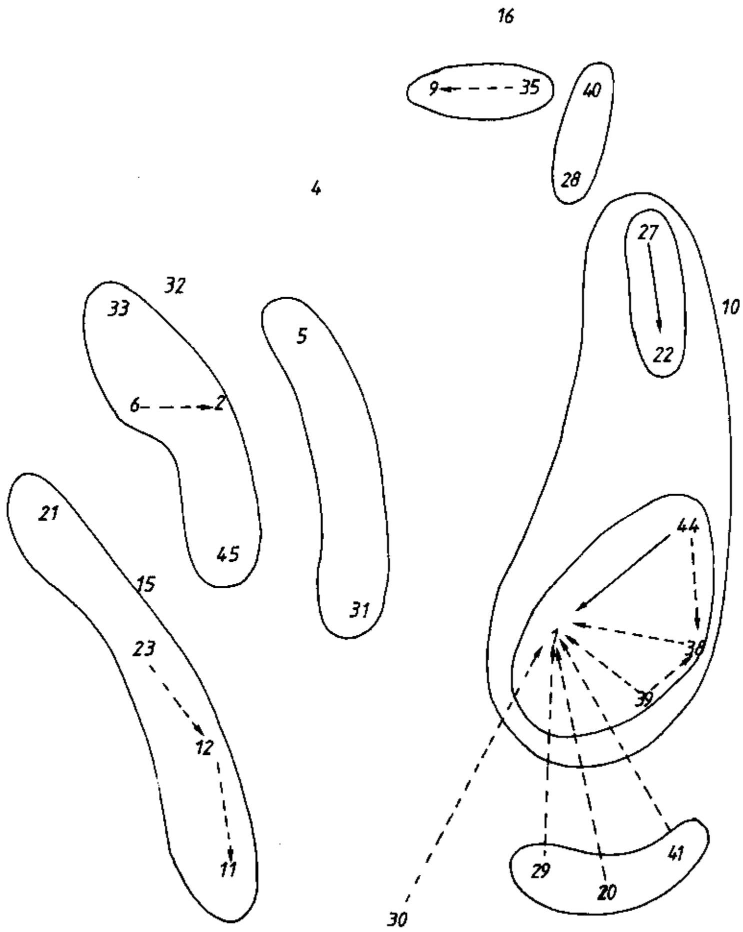


Inclusion index values > 0.8

Inclusion index values > 0.6

Plot symbols are given in Table 1.

Figure 5 : Inclusion Map for 1988 Canadian Hydrographic and Atmospheric Geophysics

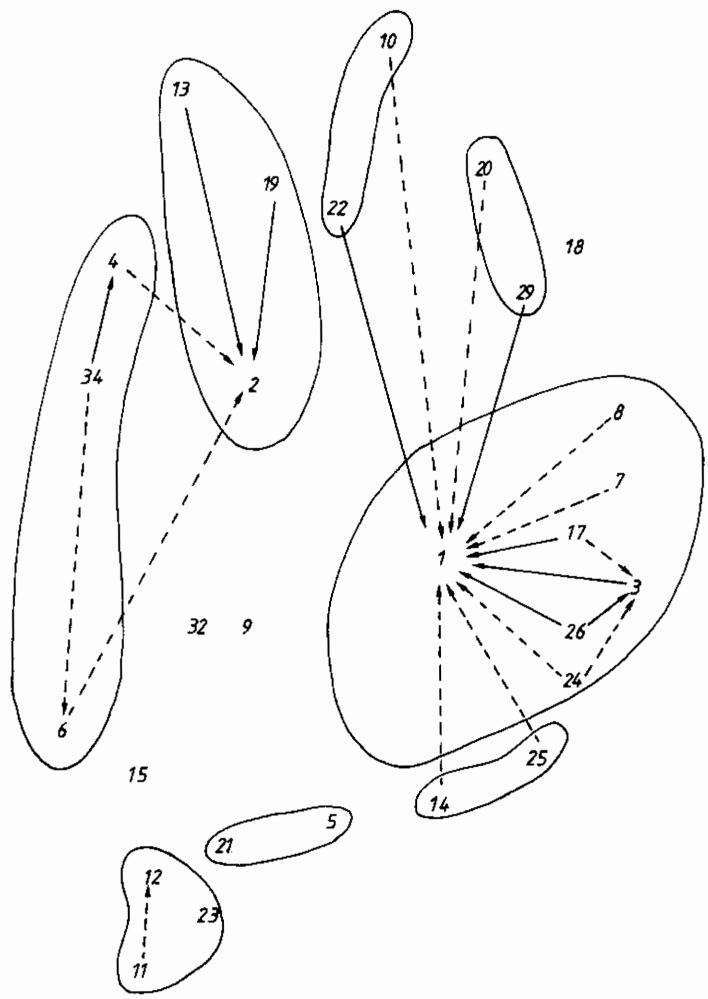


Inclusion index values > 0.8

Inclusion index values > 0.6

Plot symbols are given in Table 1.

Figure 6 : Inclusion Map for 1988 Canadian Geophysical Observation and Techniques



Inclusion index values > 0.8

Inclusion index values > 0.6

Plot symbols are given in Table 1.