Visualization of Soil Erosion Research in a Knowledge Map

A. Zartl[1]***, A. Klik***[2] and ***E. Schiebel***[1]

[1]ARC Seibersdorf Research GmbH, Department of Technology Management, Austria
E-mail: angelika.zartl@arcs.ac.at
[2]University of Agricultural Sciences Vienna (BOKU), Austria, Department of Hydraulics and Rural Water Management

Abstract: Soil erosion research is a popular topic in soil science that appears in conjunction with several other topics like soil and/or water conservation, soil physics, - chemistry and – biology and others. A request for the term "soil erosion" on CAB abstract CD, dated from January 1992 to January 2001, resulted in 15.531 articles. According to the publication year on average 1.639 articles are published per year.

Objective of this paper is to analyze literature of soil erosion research with bibliometric methods and to achieve the following targets: 1) give an overview of the various sub-themes that arise in conjunction with soil erosion research; 2) find potential missing research areas; 3) show co-operations of different science areas and countries. To fulfill these objectives a visualization and analysis tool, called BibTechMon is used. This tool helps to get a quick overview of great amounts of literature, find conceptual interactions and show interactions of content.

Results are various knowledge maps that show the content of soil erosion research and their interactions.

Keywords: soil erosion, bibliometric analysis, visualization, decision support system

1 Introduction

Today's society is confronted with an increasing amount of information and an easier access to knowledge. Internet, online databases and online libraries have dramatically improved the worldwide availability of information. On CAB abstracts on average 1.639 articles concerning soil erosion are published per year. It is impossible to read all those publications and especially for scientists who are getting involved into a new subject it is hard to filter this information flood.

In the information technology tools are developed to structure and filter information and work out the topic specific data. Analysis of information according to importance, temporal distribution and other relevant parameters is necessary to handle the knowledge.

Bibliometric methods were developed for analysis of data by surveying literature and/or patent information from databases. Relations between technological innovation, different methods of application, institutes and experts can be determined using bibliometrics. Upcoming technologies or trend analyses of future developments can be identified. Bibliometric analyses are used for strategic decision support as well as for daily work in R&D (Grupp *et al*., 1990). One of the well known relational bibliometric methods is the co-word analysis.

A combination of bibliometric methods to analyze and information visualization to visualize information can be used to handle all type of information.

2 Objectives

Objective of this paper is to analyze information, especially literature information on soil erosion research with bibliometric methods and to achieve the following target objectives:

(1) give an overview and show the structure of the various sub-themes that arise in conjunction with soil erosion research;

(2) find potential missing research areas.

(3) show co-operations of different science areas and countries

3 Methodologial approach

3.1 Data retrieval

A request for the term “soil erosion” on CAB abstract CD, dated from January 1992 to January 2001, resulted in 15.531 articles. CAB abstracts are organized in an hierarchical thesaurus, which means that also articles that do not contain the term “soil erosion” showed up, because also broader and narrow terms were included in the search. Bibliographic information collected from CAB abstracts showed title, author and related department, source in terms of publication, descriptor, abstract, publication year and a geographic descriptor.

The descriptors were used for further analysis.

3.2 Bibliometric analysis

One of the well known relational bibliometric methods is the co-word analysis. Co-word analysis originally was applied to make target oriented retrieves and later on it was used to evaluate and present research outputs (Callon *et al*., 1983). Today co-word analysis is often found in connection with information visualization. The method allows the relational analysis of documents based on terms and term-groups.

Frequency of co-occurrence of terms are the basis of a linear classification using cluster analysis. The co-occurrence is illustrated in a matrix. For statistical reasons the Jaccard Index (J) was used to normalize the elements of the respective co-occurrence matrix. The Jaccard Index gives information about the correlation of keyword and varies between 0 and 1 where 0 means no co-occurrence and 1 highest co-occurrence.

$$J_{ij} = \frac{c_{ij}}{c_i + c_j - c_{ij}}$$

J Jaccard Index; $0 \leqslant J_{ij} \leqslant 1$

c_{ij} co-occurrence of keywords i and j

c_i total number of occurrence of keyword i

c_j total number of occurrence of keyword j

By the means of various mathematical methods like cluster analysis, multidimensional scaling, classical statistics and data analysis this matrix can be illustrated in a two-dimensional science-map that represents a well structured graphic rendition of information.

3.3 Information visualization

Although it is obvious to use co-word analysis for visualization of text data, it is quite hard to handle these data and their illustration in co-word networks. To process and visualize such big amount of literature citations or patents a method was developed that uses an automated process to illustrate so called co-word maps with a spring model approach (Kopcsa and Schiebel, 1998). Keywords in this model are mass points with a mass and size proportional to the total frequency J_{ii}. The mass points are connected to each other by forces (springs) correlated to J_{ij}. The mass points are positioned randomly as a starting position and then move driven by forces defined above. This is done by iteration of the respective n-dimensional differential equation system. Though the mass points will be positioned on a 2-d map in correlation to their co-operational relation to each other as defined by the Jaccard matrix. Keywords are positioned according to their correlation; high correlation is expressed by close proximity of these points in the map.

The co-word maps can show a network structure of more than thousand points and more than hundred thousand lines. Co-word maps allow the representation of networks on the basis of different objects. In general different networks are calculated and separately used. A cluster analysis on the basis of the Jaccard matrix leads to keyword groups called themes.

3.4 BibTechMon

The program BibTechMon was developed by ARC Seibersdorf Research GmbH (Schiebel and Kopcsa, 1995). The program analyzes information based on bibliometric analysis, especially co-word analysis and visualizes information using a spring model approach as described above.

BibTechMon calculates relations between data records and visualizes connections in information pools. Objects like words, articles, websites and many more and its connections and interactions can be illustrated. The tool helps to realize the knowledge of an information pool. Basis for the bibliometric analysis are databases like project documentation database, fellow employee data, company databases or press releases. Using the concept of co-occurrence, content oriented relationships of objects can be identified. A cluster analysis proposes grouping of regarded objects. Objects can be keywords, documents, institutions, persons, qualifications and many more. For cluster identification a color code is used and themes can be defined.

Output of a BibTechMon analysis is the illustration of contents and its interaction in form of networks on technologies, experts, institutions and so on. Each object is represented by a circle. The position of an object in the network is relative and depends on high relation to its neighbors. The viewer of BibTechMon is interactive and databases can be loaded and listed for each object.

4 Results

4.1 Creation of a knowledge map

First a table of citation was created in Microsoft Access including all available information gained from CAB abstracts (Table 1). This table is the basis of the following analysis.

Table 1 Table of citation[1]

quan	title	source	descriptor	publye	autho
1 of 8816 - cab	nitrogen requirements of broccoli in cc	journal-of-vegetable-crop-	broccoli-; mulches-; mulching-;	1997	abdul-baki
2 of 8816 - cab	salvadora persica - a boon to wastelan	journal-of-scientific-and-ir	production-possibilities; uses-; n	1997	zodape-st
3 of 8816 - cab	the role of banana in the functioning of	infomusa. 1997, 6: 1, 27-3	surveys-; cultural-methods; bana	1997	rishirumuh
4 of 8816 - cab	the effect of topographical modificatior	proceedings of the sixth ir	apples-; slopes-; planting-; terra	1997	iancu-m; r
5 of 8816 - cab	the chronosequence of rehabilitating sl	south-african-journal-of-s	plant-succession; rehabilitation-;	1997	ferreira-sr
6 of 8816 - cab	ecological restoration of degraded fore	revue-forestiere-francaise	forest-plantations; afforestation-;	1997	vallauri-d;
7 of 8816 - cab	the ancient landscapes of the gerardm	revue-forestiere-francaise	forest-management; planning-; n	1997	husson-jp
8 of 8816 - cab	fertilizer response to of pithecellobium	environment-and-ecology.	soil-degradation; waste-land; affe	1997	singh-ak;
9 of 8816 - cab	hydrology of montane forests in the sie	mountain-research-and-de	hydrology-; forests-; deforestatic	1997	hunzinger-
10 of 8816 - ca	forest management in tribal areas: fore	1997, 189 pp.; refs at end	tribes-; forest-policy; ethnic-grou	1997	patnaik-bc
11 of 8816 - ca	social forestry in india: a study of ecor	1997, 250 pp.; + refs at e	social-forestry; development-pro	1997	mishra-nk
12 of 8816 - ca	the hydrology of three high-altitude for	hydrological-processes. 1	hydrology-; rain-; interception-;	1998	negi-gcs;
13 of 8816 - ca	guide for the cultivation of dodonaea vi	guia-tecnologica -instituto-	seeds-; seed-collection; hedges-	1993	comacho-

Quan means quest accession number and is the identification number of an abstract. Several other columns showing information on the article like title, author, descriptor, journal published (source), home country of author, publication year, abstract a.s.o. are included in the table of citation.

Second a table of keywords was created, containing quan and descriptors; it is called table of affiliation (Table 2).

Data are not completed in table 1 and table 2.

Table 2 Table of affiliation for descriptors

	quan	wort
	10 of 8816 - cab abstrac	erosion
	10 of 8816 - cab abstrac	forest management
	10 of 8816 - cab abstrac	participation
	10 of 8816 - cab abstrac	rural development
	10 of 8816 - cab abstrac	shifting cultivation
	10 of 8816 - cab abstrac	socioeconomics
	10 of 8816 - cab abstrac	deforestation
	1000 of 4764 - cab absti	grasses
	1000 of 4764 - cab absti	wind erosion
	1000 of 4764 - cab absti	soil conservation
	1000 of 4764 - cab absti	erosion
	1000 of 4764 - cab absti	water erosion
	1001 of 1951 - cab absti	maize
▶	1001 of 1951 - cab absti	nutrients

In the first column Table 2 shows the identification number of each article and in the second column the descriptors. Descriptors are now listed by row though for the same quan several keywords are existing.

4.2 Discussion of knowledge map

4.2.1 Keyword map

Using the method described above a map consisting of circles (representing the descriptor) and lines (representing their connections) can be created. Only keywords occurring in more than 15 out of 15.531 – though in more than 0.1 % of the articles – were used. The size of the circle depends on the occurrence of the keyword in different articles Overall the map is based on 15.497 articles.

The most frequent term is “erosion” which occurs in 5464 articles (Fig. 1). The position of the circle is relative which means there is no coordinate plane and it does not matter whether the objects are in the northern or southern part of the map. Only the relative position of objects to each other is important and also the position of the circle on the map whether it is in the center or on the edge. Objects that are close together have a close relation to each other represented by co-occurrence in different articles. Such a close relation can also be represented by a line, which shows the connections between objects. Objects that occur together in different articles but in different context with unequal keywords will not be placed next to each other because they have a different surrounding area pictured by their keywords.

Navigating through the map shows that sections representing research topics. But as this way of grouping objects would be subjective, a cluster analysis was done.

A cluster analysis using Pearson correlation and Ward linkage was performed to group objects in a objective way and to check the clusters. In the beginning five main clusters were determined(Fig. 1) shows the results, color code is according to clusters. The biggest cluster, concerning numbers of objects in the cluster, is “soil conservation”, including 394 keywords and mostly situated from the center to whole left part of the map. Names of the clusters are given by the most frequent word in the cluster. The “erosion in general” cluster, including 382 keywords is mostly situated from the center to the north of the map. Cluster number 3 is “reclamation”, including 317 descriptors, situated in the center and the right part of the map. The “tillage” cluster is mostly in the center, spreading out to the southern and southeast parts

of the map. Finally a small but very strong cluster only including 36 objects is “pollution”. This cluster is situated in the east part of the map and except the word pollution which is situated in the center those objects are situated next to each other and additionally have strong connections – though a very separated, defined cluster and therefore also research topic.

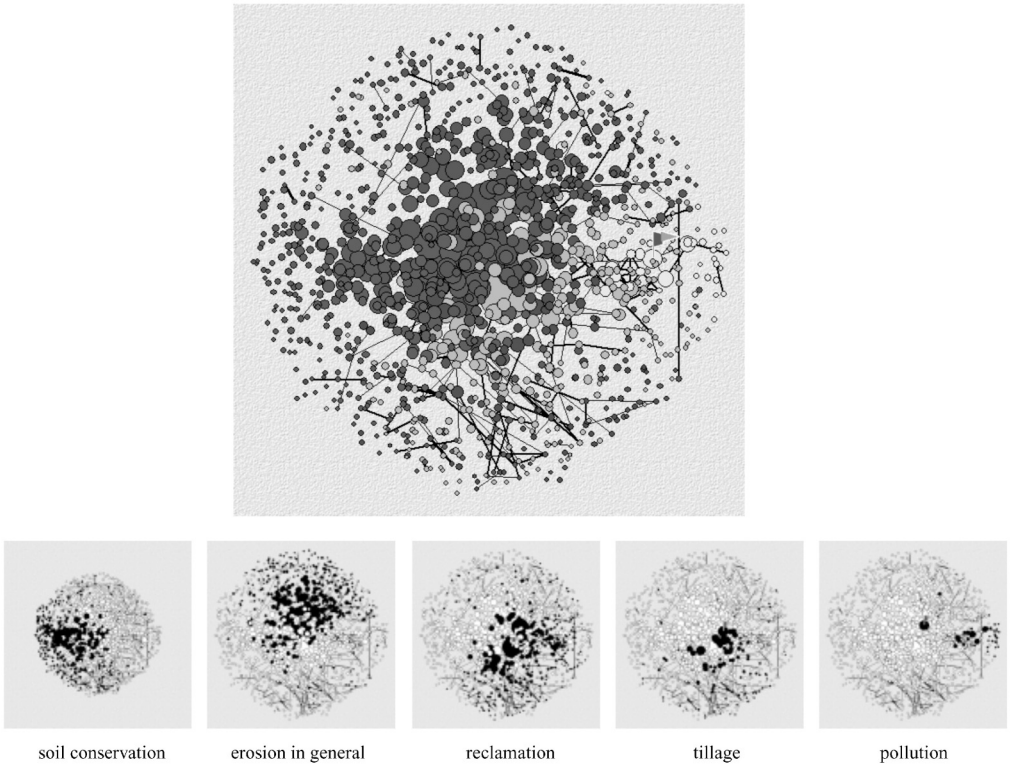


Fig.1 Knowledge map showing network of descriptors/keywords and a detailed look at the position of the five sub-themes

It is interesting that soil conservation, on the left side of the map, and reclamation, on the right side of the map, are opposite to each other. Though we focused more into detail of each cluster and made a cluster analysis with the same parameter as described above but limiting it to 25 clusters.(Fig.2)

Table 3 Split-up of word groups according to the basic five clusters

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
soil conservation	erosion in general	reclamation	tillage	Pollution
soil conservation	erosion	afforestation	soil management	Pollution
soil fertility	erosion control	soil	tillage	heavy metals
forests	hydrology	reclamation	grasslands	
control	sediment	revegetation	weeds	
land use	soil water	water quality		
forest management		fertilizer		
environm impact				
water use				

In the map can be seen that topics erosion and erosion control are not merging like erosion control and soil conservation. Erosion control and soil conservation are more applied research subjects and erosion is more a scientific one. Between them there is a gap, though from the publications can be seen that not so many scientists working with the basics and fundamentals of soil erosion are also publishing on erosion control. The same can be found for soil fertility and tillage, but not for topics fertilizer and tillage. In the erosion general cluster hydrology and soil water are close together but do not overlap. Hydrology deals with GIS and remote sensing as well as overland flow and soil water consists more of soil physical terms. The connection of both will be necessary in the future to understand the whole hydrological circle.

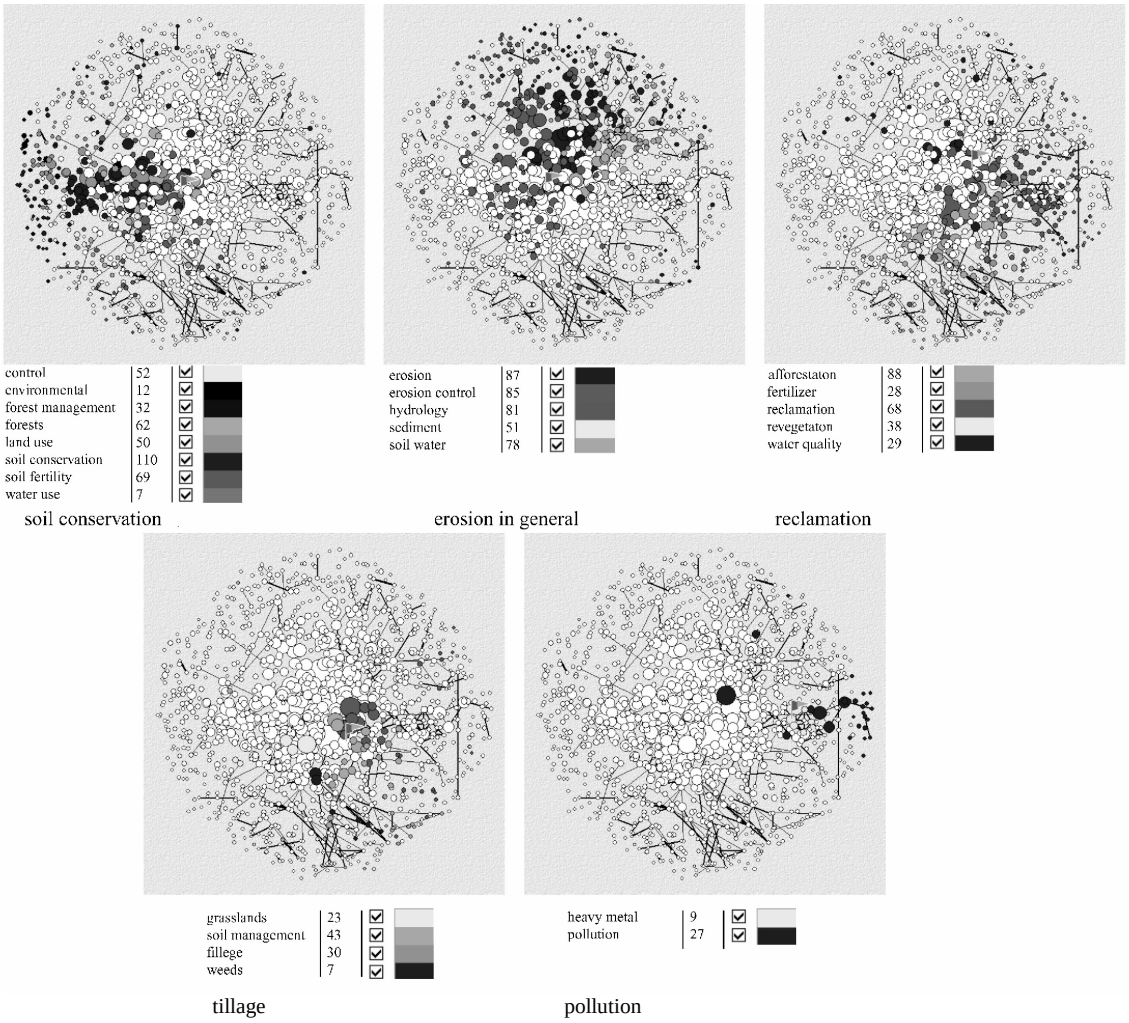
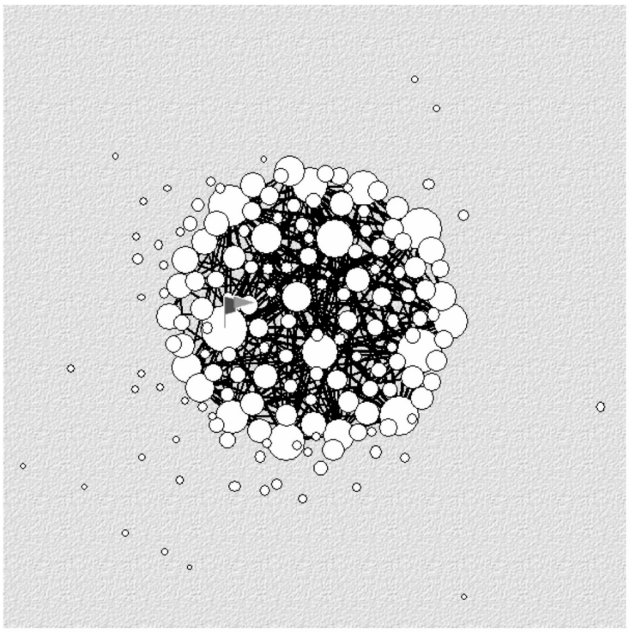


Fig.2 Tillage detailed networks of keywords, showing positions of 25 clusters

4.2.2 Network of countries

From the literature and publications it can be seen that there is not too much cooperation going on concerning joint publications from different countries. Though a network based on the content was developed (Fig. 1). This means that countries that use the same keywords appear close to each other.



	country	Anzahl von quan
	usa	1091
	india	995
	australia	891
	china	889
	uk	770
	germany	718
	canada	706
	spain	683
	italy	619
	brazil	616
	russia	611
	japan	550
	indonesia	501
	france	471
	kenya	450
	poland	448
	new-zealand	444
	philippines	421
	south-africa	409
▶		
	austria	211

Fig.3 Network of co-operation of countries based on the content and ranging of countries according to their number of publications

The network is very homogenous, there is a strongly connected circle in the middle and only a few spread out countries. This means that most countries work on the same topics. To get more detailed information it would be necessary to focus on interesting countries. Finally the number of publications according to each country is given.

The top three countries concerning numbers of publications are USA, India and China. First European country, which appears in the list is UK followed by Germany.

5 Summary and conclusions

In this paper the bibliometric tool BibTechMon was used to structure the overall theme soil erosion and cluster the topic in several sub-themes found objectively. Literature of soil erosion on CAB abstract CD, dated from January 1992 to January 2001 was analyzed and five main sub-themes could be identified: soil conservation, erosion in general, reclamation, tillage and pollution. Pollution is a very isolated topic in conjunction with soil erosion and should get more attention in the future. All themes were analyzed more detailed. The knowledge map shows that there are potential missing research areas represented by sub-themes that are not merging in the map, which means that there is no comprehensive research going on, like between erosion and erosion control or soil fertility and tillage.

The map of the countries based on their content is very homogenous, which means that most countries are working on all or most of the sub-themes and not focusing on one special topic.

The knowledge maps help to get an overview of the various sub-themes arising in conjunction with soil erosion, structure the topic and find potential missing research areas. Nevertheless this paper could just give a quick and short overview of the complex topic soil erosion and the ability to analyze it with a bibliometric tool.

References

Callon M., J.P. Courtial, W.A. Turner, and S. Bauin, 1983. From Translations to Problematic Networks: an Introduction to Co-Word Analysis. Social Sciences Information, 22, 191-235.

Grupp H., T. Reiss, and U. Schmoch, 1990. Knowledge interface of technology and science: developing tools for strategic R&D management. FhG-ISI report to the Volkswagen Foundation, Karlsruhe.

Kopcsa A. and E. Schiebel, 1998. Science and Technology Mapping: A New Iteration Model for Representing Multidimensional Relationships. Journal of the American Society for Information Science. JASIS Jan 1998, No.1/Vol49, p 7ff.

Kopcsa A. and E. Schiebel, 1995. Entwicklung eines Modells zur Unterstützung der Forschungs- und Entwicklungstätigkeiten bei Unternehmen unter Verwendung bibliometrischer Methoden. Fallbeispiel: Analyse für Semperit Reifen AG, Seibersdorf, Austria. Austrian Research Centers.