Mapping science through bibliometric triangulation: an experimental approach applied to water research

Journal:	*Journal of the Association for Information Science and Technology*
Manuscript ID	JASIST-2014-08-0558.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Wen, Bei; Rathenau Institute, Science System Assessment; Delft University of Technology, Water management; KWR Watercycle Research Institute, Knowledge Networks and Futures Studies Horlings, Edwin; Rathenau Institute, Science System Assessment Van der Zouwen, Marielle; KWR Watercycle Research Institute, Knowledge Networks and Futures Studies Van den Besselaar, Peter; VU University Amsterdam, Organization Sciences & Network Institute
Keywords:	knowledge discovery < automatic extracting < computer applications < computer operations < (activities and operations), content analysis < text processing < information processing < computer operations < (activities and operations), analytical bibliography < bibliography (descriptive activity) < library operations < (activities and operations)

Mapping science through bibliometric triangulation: an experimental approach applied to water research

Bei Wen[1,2,3], Edwin Horlings[1], Mariëlle van der Zouwen[2], Peter van den Besselaar[4,*]

[1)] Rathenau Institute, Department of Science System Assessment, The Hague, The Netherlands,

[2)] KWR Watercycle Research Institute, Knowledge Management Group, Nieuwegein, The Netherlands

[3)] Delft University of Technology, Faculty of Civil Engineering & Geosciences, Department of Water management, Delft, The Netherlands

[4)] VU University Amsterdam, Department of Organization Science & Network Institute, Amsterdam, The Netherlands,

* Corresponding author: p.a.a.vanden.besselaar@vu.nl

Abstract:

The idea of constructing science maps based on bibliographic data has intrigued researchers for decades, and various techniques have been developed to map the structure of research disciplines. Most science mapping studies use a single method. However, as research fields have various properties, a valid map of a field should actually be composed of a set of maps derived from a series of investigations using different methods. That leads to the question what can be learned from a combination – triangulation – of these different science maps. In this paper, we propose a method for triangulation, using the example of water science. We combine three different mapping approaches: journal-journal citation relations (JJCR), shared author keywords (SAK), and title word-cited reference co-occurrence (TWRC). Our results demonstrate that triangulation of JJCR, SAK, and TWRC produces a more comprehensive picture than each method does individually. The outcomes from the three different approaches are associated with each other and can be systematically interpreted, and provide insights into the complex multidisciplinary structure of the field of water research.

Keywords:
Hybrid mapping; methodological triangulation; science maps; water research.

Introduction

Bibliometrics provides a set of methods to describe quantitatively various attributes of a corpus of literature, such as patterns in journal, paper, or author relations. In this way, bibliometrics aims to provide insight into knowledge dynamics: the development of knowledge in a given area, in relation to larger knowledge landscape. A variety of techniques has been developed to *map* the structure and dynamics of disciplines and its research fronts. Co-word analysis and citation analysis are the most commonly used bibliometric mapping methods, based on the content of and the relations between publications in a field. Comparing word-based and citation-based maps leads to different clustering outcomes, suggesting that one probably needs multiple maps showing different insights (Borner et al. 2003).

The last decade, attention for science maps has increased, and different kinds of science maps have been proposed, to reveal relations among, for example, authors, documents, journals, or keywords, and they are usually constructed based on citation, co-citation, bibliographic coupling, or co-occurrence of words in documents. Theses mapping methods may be broadly grouped according to different levels of scope based on their targeted units of analysis, for example, analysis of journal citations may present the broadest scope at the journal level; co-authorship or title word co-occurrence analyses may present a highly condensed summary content level; and keyword, abstract, or content words co-occurrence analyses may represent more detailed content level. Many of these methods have been successfully applied to various scientific fields. Most of the empirical studies map science using a single method, depending on the purpose of the study (for an overview: Morris & Van der Veer Martens 2008). Apart from the analytical tools, the recent decade also has shown an increased emphasis on the visualization of the results. More recently, mapping studies have started to *compare* the cluster solutions resulting from various similarity approaches or classification algorithms (Ahlgren & Colliander 2009; Börner et al. 2003; Janssens et al. 2009; Jarneving 2005; Lu & Wolfram 2012; Shibata et al. 2009), in order to find "the most accurate representation" (Boyack & Klavans 2010; Boyak et al 2011).

However, as is also the case of geographical maps, different maps highlight different aspects of the phenomenon under study. For science maps to be reliable and useful, they should be based on a combination – a 'triangulation' - of mapping approaches using a variety of data. In social science, triangulation is defined as the mixing of data or methods so that diverse viewpoints cast light upon a topic (Olsen 2004). Cohen and Manion (2000) viewed triangulation as an "attempt to map out, or explain more fully, the richness and complexity of human behavior by studying it from more than one standpoint." Altrichter et al. (1993) contended that triangulation "gives a more detailed and balanced picture of the situation."

Although scholars have started realizing that using multiple maps provides a broader picture of a scientific field (Borner et al. 2003), it is not yet common practice. One reason is that even mapping outcomes from individual approaches are already often complex to interpret. Another reason is that we lack systematic methods for comparing maps. The *key challenge* in this paper is developing such a systematic 'triangulation' method for linking multiple mapping approaches and interpreting them in a meaningful manner. In order to do this, we first select three often-used approaches for mapping scientific fields. On top of that we propose a method for *cross-tabulation* of clusters revealed by the science maps.

As a proof of concept, we apply this bibliometric triangulation method to *water research*. This field was selected because the scope and volume of water research are growing rapidly, making it increasingly difficult to understand the complex relationships between the involved scientific specialties. This poses challenges for effective research planning in this societal important field of science (and technology). Consequently, there is a need to better characterize, define and understand the field of water research.

This paper contributes to the development of bibliometric metrics that helps to formulate a (practical relevant) theory of knowledge dynamics. We explore how different approaches to mapping a scientific field provide different insights, and how they can be related. We demonstrate how triangulation results, in conjunction with visualization, can lead to a better understanding of the – disciplinary or interdisciplinary – structure of scientific fields.

The paper is organized in five sections. The section *Research framework and related work* introduces a triangulation model based on a multi-method approach used to delineate knowledge domains and identify research specialties, including a brief review of the related bibliometric techniques. The *Data and methods* section details data collection and the application of our model. In the *Results* section we show the cognitive maps derived from each individual methodology for delineating the field of water research. In the section *Triangulation*, we combine the three methods and illustrate the added value of triangulation. Finally, we draw conclusions and end with a general discussion.

Research framework and related work

In this section, we discuss several commonly used bibliometric methods and provide the reasoning behind our triangulation framework. Boyack et al. (2005, p352) stated that a correctly constructed science map can help to understand the inputs, associations, flows, and outputs of science and technology: “Just like in a physical world, maps help us to understand our environment – where we are, what is around us, and the relationships between neighboring things”. Of course, the issue in this paper is not so much whether the map is ‘correct’, but what different insights come from differently constructed maps. Morris and Yen (2004) have identified a variety of entities that can be the object of mapping exercises, such as papers, authors, references, journals, and terms (e.g., keywords). These basic entities of science are interlinked through papers. For example, papers are written by authors, published in journals, characterized by title words and keywords, and linked to other literature through cited references.

Papers contain new knowledge in detail as well as changes in the interests and concerns of the discipline or the author constituencies. They embody the evolution and dynamic of science over time. They are one of the most common units used to map a knowledge domain (Börner et al. 2003). We can construct maps by clustering papers based on title words, keywords, references, journals or combinations thereof. The clusters of (in one or another way similar) papers are the entities the map consists of. A *keyword* is an index term representing the core of a documents’ *content*, such as the method and the specific objects. Authors as well as editors assign keywords to papers. We use the *author* keywords as that can be considered a carefully positioning of the paper (Whittaker 1989). If papers share more author keywords, they have a more similar technical content. *Title words* have a different role. Although they may lead to similar maps in small and homogeneous fields, in larger fields title words refer more to the newness and the topic of the study (Whittaker 1989). By selecting *references*, authors link their work to previous work. This better represents the disciplinary (or multidisciplinary) identity of the work than the technical content or the topic. The emergence of field-specific academic *journals* is a sign of the growth and maturity of a discipline. Journals are scholarly media, normally long-lived, and journals belonging to the same field have similar aggregated citation patterns. Delineating a journal communication network may offer more definitions of disciplines and specializations. Journal maps can also provide the relative relationships between major disciplines at a macro view of science (Börner et al. 2003). This makes it possible to define how different fields of knowledge interact, and provides effective ways to evaluating research performance and predicting interdisciplinary impact (Garfield 1972; Van den Besselaar & Leydesdorff 1996).

The two entities (papers and journals) and the three attributes (keywords, title words, and references) can be analyzed using different bibliometric procedures, leading to different maps, each of which may clarify specific aspects of a field. We want to study whether triangulating (combining) these methods tells more about a field than the different maps together. We start with three different analytical procedures (analyzing *journal structures* using references; analyzing *papers structures* using keywords; analyzing *papers structures* using title words and references). The resulting maps give a first analysis of the field. As journal clusters can also be handled as clusters of papers, we do have three paper-cluster structures. *Cross-tabulation* of the three paper-cluster structures provides interesting additional information about the (in this paper: water research) field.

Journal networks based on journal-journal citation

In most disciplines, journals are the dominant channel of scholarly communication. Journals provide entries into topical specializations (research fronts) within a field or a discipline, and they provide researchers with means to find relevant information. Journals belonging to a field are expected to have a shared knowledge base, which forms a major source for references in papers published in those field specific journals. Consequently, journals belonging to a (sub)field are expected to have similar aggregated citation patterns. Analyzing patterns of citation between journals allows us to delineate scientific fields, as well as to determine knowledge flows between fields. Journal-journal citation analysis has been widely accepted as a powerful method for mapping the intellectual structure and dynamics of science at the macro level, and for the analysis of scientific specialties and the disciplinary organization of the sciences in terms of networks of journals (Van den Besselaar & Leydesdorff 1996). Delineating research fields on the higher aggregation level is an essential step to investigate the (inter)disciplinary identity of research fields (Van den Besselaar et al. 2001; Vugteveen et al. 2014). Journal-journal citation analysis enables us to delineate knowledge domains and to sketch out the boundaries between research specialties.

Publication networks based on keywords

Co-word analysis, introduced by Callon et al. (1983; 1986), makes use of the patterns of co-occurrence of words or phrases in a corpus of texts to cluster the texts thematically. Mostly it is based on title words, abstract words, or keywords, although increasingly full text is used. Co-word analysis has been used to map the cognitive structure and the development of research fields (Bauin, 1986; He, 1999) and of science as a whole (Boyack & Klavans 2013). Quite a few researchers have used co-keyword analysis to reveal patterns and trends in a specific discipline. Some examples concern technology foresight (Su & Lee 2010), research policy (Lee & Su 2010), ethics and dementia (Baldwin et al. 2003), library and information science (Åström 2002), and information retrieval (Ding et al. 2001).

Whittaker (1989) mapped the structure of scientific fields by using co-word analysis of both the keywords and the titles of a set of papers. His study suggested that the keyword-derived results provide substantially greater detail than title or abstract words, as the former tends to show the relationship with other papers, whereas the latter often emphasizes the supposed originality of a paper. In this paper, we use papers, rather than words, as the units of analysis in order to compare the resultant network with the other paper network. We use the author keywords-based approach drawing upon Whittaker's argument that authors choose technical terms carefully to constitute an adequate description of the content in terms of problem/method combinations. Consequently, the more co-occurring keywords two papers share, the more similar they are. This creates a paper-network that provides a good entry point for understanding the set of problems and related methods structure of a scientific field.

Publication networks based on word-reference combinations

Science maps of articles are not only based on keywords, as in the above section, but also on title-words, and on citation relations (co-citation analysis, or bibliographic coupling). In all these cases only one attribute is used. An increasing set of 'hybrid' approaches at article level has been developed, using combinations of attributes of papers (Zitt 2015). For instance, Braam et al. (1991) investigated the structural and dynamical aspects of science maps based on a sequential combination of author co-citation and co-word analysis. Åström (2002) constructed maps for delineating library and information science using the co-occurrence of keywords and cited authors. Zitt and Bassecoulard (2006) developed hybrid approaches associating lexical and citation-based analysis which they believe can be efficiently applied for clustering and mapping research specialties. Boyack and Klavans (2010) tested the accuracy of cluster solutions resulting from different similarity approaches, including a hybrid text-citation approach – and suggest that the hybrid approach has more potential.

Our analysis is based on shared co-occurring title word-cited reference combinations (Van den Besselaar and Heimeriks 2006). Word-reference analysis differs from the well-known concept of bibliographic coupling because it includes not only references but also title words for the coupling of texts. The idea behind this hybrid approach is that title words point at the topic of research, while cited references represent the relations of the paper with previous research, indicating the paradigmatic identity of a paper. Scholars select title words to describe the supposed originality of their research and cite specific literature to indicate the tradition to which their work is related. Title word-cited reference combinations measure these dimensions simultaneously, providing a fine-grained topical structure of a scientific field or specialization. Other advantages exist too, such as avoiding threshold (and through this coverage) problems (Van den Besselaar & Heimeriks 2006).

Data and methods

There are many ways to collect the bibliographic data and map the scientific fields, either globally (those based on the entire bibliometric databases like WoS or Scopus) or locally (those based on a subset of data). In this study we focus on *water related science and technology* for which the related bibliographic data represent only a subset of the entire database. The mapping solution from this subset is also expected to be different from a global solution using the entire database. To achieve a solution that is specific and fitting the targeted water domain, we consulted content experts to judge if our selected sets of publications and the resultant clusters make sense rather than relying on parameters from un-supervised algorithms. Further, the idea of triangulation is to address the complex interrelationships between clustering solutions at three distinct levels of scope, i.e. journal level, content level, and paper level. Therefore we chose some of the established analytical methods at each level, i.e. journal mapping (Leydesdorff & Cozzens 1993; Van den Besselaar & Leydesdorff 1996), co-word (here keyword) mapping (Callon 1986; Klavans & Boyak 2006), and hybrid citation/lexical mapping (Van den Besselaar & Heimeriks 2006).

We avoid discussions of what constitutes ‘the best method’ for doing journal citation based mapping, hybrid mapping (Boyack and Klavans 2010) or for co-word analysis based mapping (Boyack et al. 2011). Firstly, because the ‘best’ is in our view often not the issue - different mapping methods result in maps offering different perspectives of the field under study - and secondly because we focus in this paper on how those differences can be used productively.

We use the following analytical steps.

1. Selecting an appropriate dataset using relevant search terms. This leads to the delineation of a corpus of relevant papers – in interaction with field experts.
2. Organizing the data relating to a unit of analysis, i.e. constructing title word-cited reference combinations, a list of shared keywords, and journal-journal citations counts.
3. Calculating correlations or similarities for each unit of analysis (papers and journals), applying a clustering algorithm to identify research communities, and visualizing the structure of the data using social network analysis software.
4. Assigning higher-level denominations to the communities identified in each map in order to interpret the three structures – in interaction with field experts. This leads to a conclusion about what the different maps show.
5. Triangulation: ‘cross-tabulation’ of the three maps to deepen the understanding of the structure of the field.

Data acquisition and preparation

The analysis is based on 2010 publications downloaded from the five citation databases of Thompson Reuters Web of Science (WoS) in April 2013. (footnote 1) We used “topic search” which retrieves documents based on the appearance of selected search terms in the title, the abstract, or the keywords. We restricted the search to citable items: articles, reviews, proceedings papers, notes, and letters. For processing and analyzing the bibliographic data we used the Science Assessment Integrated Network Toolkit (SAINT). (footnote 2)

The first exploratory task is to define a set of search terms that properly delineates the field of water (related to the production, distribution, and use of drinking water, wastewater treatment, hydrology, etc.). A set of search terms is considered perfect if all resultant papers belong to the water field (precision) and, simultaneously, if all of the papers belonging to the water field are found by this set of search terms (recall). To achieve this goal, the identification of search terms was done through an iterative process of asking experts, retrieving papers and validating the retrieved set with the experts, then updating the search terms and starting a next cycle. After the experts validated the cognitive maps, a sound balance between recall and precision had been reached, and the process was stopped (Van Den Besselaar & Gurney 2009). The experts involved did cover the following water research fields: hydrology, wastewater treatment, waster re-use, drinking water and desalination, etc., and most of them were employed at a water cycle research institute.

An initial experiment was done with a non-restricted set of search terms: water treat*, water quality, drinking water, waste water OR wastewater, desalinate* and hydrolog* using topic search in the WoS. For the period 2008-2009, there were 60,162 documents in total (Table 1).

The sum of individual search term searches was 76,375, while the sum of unique hits of each individual search term was 46,275. This implies that 77% of the documents were found using only one search term. In the document set, the initial search terms turned out to be rather disjunctive (Figure 1).

The initial document set was used to sketch out preliminary cognitive maps based on keywords, journals, and papers. These maps provided an overview of research communities and their interrelationships. The results were complemented and validated with expert insights. Experts from KWR Watercycle Research Institute (footnote 3) were asked to reflect on the obtained set of publications. Their suggestions were included by adding a number of additional search terms, such as “water re-use OR water reuse”, “water cycle*”, “fluid dynamics”, “water system*”, “water management”, “sewer* OR sewage”, “water distribution”, “water suppl*”, “water safety”, “water sanitation”, “water resource*”, “water quantity”, “water demand”, “water policy”, “water sustainab*”, “climate change”, “global warming” and “water energy”.

Is the selection of search terms correct? There is a trade-off between precision and recall: the more precise we try to be, the higher the risk that relevant papers are excluded, and the better recall we try to get, the lower the precision generally is. We assessed the list suggested by experts in terms of recall and precision. Adding search terms such as “fluid dynamics”, “water safety”, and “water energy” increases recall. For instance, 98% of the papers found with the search term “fluid dynamics”, 76% with “water energy”, and 59% with “water safety” were not included in the initial dataset. However, the precision of the dataset decreases, because these words contain a high proportion of papers belonging to non-relevant research topics. This becomes clear when we take a closer look at what is behind the unique hits. For example, “fluid dynamics” has a wide range of applications focused much more on mechanical engineering than on water issues. Conversely, search terms such as “water safety”, “water energy” and “water re-use OR water reuse” are eliminated from the list because they do not bring in new relevant documents. With the updated search terms, the document set (Table 2) was improved in terms of recall without seriously compromising precision.

We applied the new set of search terms to the WoS 2010 corpus, resulting in 23,406 documents (among which 22,929 are articles, reviews, proceedings papers, notes, and letters). The sum of the results of all individual search terms is 31,685. The sum of the unique hits of each search term is 17,271, which means that 74% of the papers are found through only one search term.

Relationship matrices

For each of the three approaches, we constructed a relationship matrix. The cells of the journal-journal matrix represent the citation frequency, and the cells in the two paper-paper matrices represent the number of shared keywords and the number of shared word-reference combinations. The constructed matrices provide the strength of the relation between the entities (journals, publications). Applying clustering techniques, groups of similar entities (papers, journals) were identified.

Journal-journal citation analysis: From our dataset we calculated the number of papers that was published in each journal. The distribution of a total of 3,317 journals in the dataset is very skewed, and the top 150 journals covers 50% of all papers. Core journals were selected using two criteria: they include more than 1% of total papers in our dataset (meaning that the journal is important for the field), or a journals' papers in the dataset account for more than 20% of all papers published in that journal (meaning that the journal is enough focused on the field, and that water S&T is not a marginal topic in the journal). A total of 24 core journals satisfied these criteria, which accounted for about 25% of all papers in our dataset. These 24 journals were used as seeds to further select their inter-citing journals to construct a journal-journal citation environment. A citing or cited journal was selected if it covers more than 0.6% of all cites to of from one of the seed journals. The threshold of 0.6% was chosen due to our computational capacity, and it restricts the journal network to 254 journals. The journal-journal citation information was obtained from the Journal Citation Report of the Science Citation Index (SCI), published by Thomson Reuters. The final relationship matrix with journal citing and cited counts contained 254 inter-citing journals (including the 24 seed journals, covering 11,598 papers in our database) as rows (i) and columns (j), in which the cell (i,j) represents the total number of times journal i is being cited by journal j. This (non-symmetrical) relationship matrix was normalized using cosine as the similarity measure (Leydesdorff & Probst 2009).The cosine value was calculated based on the cited dimension, which was used as the input for the subsequent clustering analysis.

Shared key-word analysis: We extracted m ($m=43,745$) keywords from n ($n=18,816$) papers with at least one keywords available. For each pair from the n papers, we counted the number of shared keywords pairs, resulting in a large, symmetric, and pair-wise $n*n$ shared keywords matrix. We only considered publications ($N=9,410$) that share two or more keywords with at least one other publications to construct a shared author keyword (SAK) matrix as input for the subsequent clustering analysis.

Word-reference analysis: Consider a paper with a title of x meaningful (footnote 4) title words and with y cited references. This paper is characterized by $x*y$ *title word-cited references combinations* (TWCR). For each pair of retrieved papers n ($n=20,890$), we count the number of shared TWCR combinations. The degree of similarity between each pair of papers is represented by a Jaccard coefficient, which is equal to the intersection of two papers divided by the union of two papers (the sum minus the intersection). This results in an $n*n$ symmetric relationship matrix that was used as input for subsequent clustering analysis.

Clustering algorithms

The three methods that we compared in this paper share one significant property: they organize the data in the form of large and complex networks. Studying such networks demands efficient methods to retrieve comprehensive information about their structure. We applied clustering algorithms to classify bibliometric units into mutually exclusive communities with high homogeneity within clusters and low similarity between clusters. Furthermore, for identifying and characterizing the important concepts of a scientific field, we label the clusters of these research interests in a structured content-map according to the detailed information extracted from each cluster.

Two clustering techniques are applied in our study to find communities of similar papers: Blondel et al.'s method (2008) and factor analysis. The clustering algorithm of Blondel et al. is used to identify coherent clusters in the keyword based paper network and in the word-reference based publication network. The method is a heuristic method based on the modularity optimization of the partition as an objective function, which is highly suited to find structure in our dataset, because the number of keyword pairs and title word-cited reference combinations can be extremely large.

The combined citation environments of the selected journals are consolidated. The resulting journal-journal citation matrix is factor-analyzed, not using the pearson correlations between the variables but using the cosine. The processed journals are the independent variables. Factor analysis enables multivariate exploration for either variable reduction or structure detection. It results in clusters of journals with similar citation patterns – which is different from strongly citing each other. When applied to the main journals in our dataset, the resulting map shows the different fields and disciplines that constitute water research, as well as fields that provide relevant knowledge to these fields and disciplines (Van den Besselaar & Leydesdorff 1996).

Visualization

Similarity, occurrence, and correlation between publications, keywords or journals can be expressed in the form of network data. The networks were visualized in Gephi (Bastian et al. 2009), a software tool for large network analysis, using the OpenOrd layout algorithm (Martin et al. 2011).

Triangulation through cross-tabulation

As explained above, we mapped the field of water research using three different approaches: as journal clusters (JJCR), as publications clusters through shared author keywords (SAK), and as publications clusters through shared word-reference combinations (TWCR). Each approach provides its own insight into the structure of a field, which can be visualized in the form of a network. The objective of triangulation is to relate the results of the different bibliometric approaches. We analyze the results using three cross-table comparisons.

Each of the clustering methods assigns a paper to a set: a journal set representing a field; a paper set representing a topic and a paper set representing a problem/method combination. We now can do the following cross-tabulations:

(i) Topics by field. This shows how topics are embedded in one of more fields – and informs us about the disciplinarity of the topics. If most of the papers of a topic belong to only one field, the topic is clearly mono-disciplinary, if the papers are in more fields, the topic is multidisciplinary. As there are always some papers of a topic in any field, we only count those fields that cover at least 5% of the papers of the topic. The number (N) of fields is similar to the variety (Porter & Rafols 2009). Not only is the number of fields (N) per topic relevant, also the distribution of the papers of a topic over the fields. If a topic is mostly in one field and only slightly in others, it is less multidisciplinary than in case the papers of a topic are evenly distributed over fields. We use the coefficient of variation (CoV) for this: the ratio of the standard deviation and the mean, and this is similar to *balance* as defined by Porter & Rafols (2009). The lower the CoV, it is, the more evenly the papers are distributed over the involved fields, the higher the level of multi-disciplinarity.

(ii) Methods by field. This shows the relation between the methods and the fields: what methods are used in the water research fields, and in what other fields are they developed and used. Here again, there may be a 1 to 1 relation between method and field, or a method may be used in more fields. In the same way as with the topics we can define the multi-disciplinarity of the methods: the higher the N and the lower the CoV, the more multidisciplinary the method is.

(iii) Topics by methods: We can describe the use of methods in the various topics, and calculate the degree of 'multi-method' of the topics. And, we get insight in which topics are covered by similar methods – and which methods.

Coverage

Table 3 summarizes the coverage of the papers clustered by these three methods. In total, there were 1,823 papers out of 22,929 papers were not covered by any of these three methods.

We compared the papers covered by a method with the papers uncovered by that method. This was done by looking at how these papers were clustered through the two other methods, i.e. if the covered and uncovered papers in one method show different clustering solutions in another method. Figure 2 shows that the covered and uncovered papers by one method followed roughly the same distribution through other clustering methods, and therefore the included papers seem a reasonable sample from the complete papers set, with only a few exceptions – exceptions that may need further exploration in follow-up studies. This indicates that although a substantial proportion of papers was excluded by our filtering criteria, it unlikely led to a biased figure about the interrelationships between the clustering results from different methods.

Results

In this section we applied each individual method to delineate the field of water research. The results are described in the following subsections.

Clustering using journal citation relations

What are the dominant journals and citation environment in water research and technology? From the document set, we selected the most frequently occurring 24 journals (footnote 5). These journals were used as entrance (or seed) journals for building the citation environment of water research. The citation environment contains 254 journals. Factor analysis was used to cluster these journals according to similarities in the way they cite other journals in the environment, resulting in 28 factors, each representing a different research field. Figure 3 illustrates the structure of the journal citation environment. In the center of the map we find two factors (clusters) "water science & technology" and "environmental science". These two factors contain a large part of the entrance journals (25% each) and form therefore the core of water research. Next to it, we have a third smaller cluster belonging to water research, which is desalination. Also this contains several (8%) entrance journals, but these form half of the desalination factor. The fourth water research related cluster is hydrology that contains 21% entrance journals. The distance in the map between hydrology and the three other clusters indicates that water research in our definition may consist of two weakly linked parts.

The network furthermore shows the relative position of different scientific fields that are related to water research, and some may be a knowledge source for water research, and others may use water research results. For example, close to "water science and technology" we find "biotechnology", suggesting that in water research quite some biotechnology knowledge and methods may be used.

The value of the journal map is that it shows some main characteristics of water research, and it shows the relative positions of different scientific fields that are related to water research. That clarifies some aspects of the nature of water research, but leaves unknown other crucial characteristics, such as the research fronts and their methodological and (multi) disciplinary nature. For that one needs, as we will see, the combination of maps.

Clustering using author keywords

The SAK method clusters 9,410 papers into 765 clusters. We only include the top 20 communities in the analysis (those containing at least 1% of total publications), which account for 82% of the total or 7,717 publications. The keyword communities appear to reflect various *methods* used in the water research field in relation to the specific research aim or problem: such as filtering (the method) for organic matter removal (the problem to be solved). The keyword network offers a *content* map. Figure 4 presents the resulting map and reveals the relationships among publications in terms of shared keywords.

The dominant community in the map is "modeling the relationship between land use and water quality", with about 10.5% of the papers. The map reveals a dichotomy between the models used in water management in the left side of the map and the techniques used in water treatment in the right side.

The advantage of the keyword-based map is that we show what method/problem combinations do dominate water research. But the map at the same time only reveals partial insight, as the relationship between these methods and the dominant research fields and the main research topics remains unclear.

Clustering using title-words/cited-reference combinations

The TWCR method clusters 20,890 publications into 192 communities (Figure 5). Community names have been assigned on the basis of author keywords, assigned keywords, title word combinations, journal titles, journal categories, most cited articles, and authors. To do so, we also examined sub-communities in each community (communities aggregated at a lower level). The results provide insight into the cognitive structure of water research topics – the focus of this paper. The largest community is "influence of climate change on hydrology cycle and water resource management" with about 19.0% of the papers. Quite a few other research topics relate to waste water treatment in different ways. Also this map also reveals a dichotomy – here between water management in the left side of the map and water science and technology in the right side.

One of the questions that immediately comes up here – but cannot be answered with this map alone – is about the disciplinary nature of the research topics. Also here we assume that triangulation may help.

Triangulation

We have used three bibliometric approaches for studying the cognitive structure of scientific fields, using the water research field as an example. The clustering results from these approaches (figure 3-5) obviously highlight different perspectives. The analysis of the journal citation networks shows the disciplinary environment of water research. For example, JJCR analysis shows that "environmental science" and "water science & technology" are the central disciplines of the field (figure 3). The SAK focuses more on the content in terms of approaches, methodologies and techniques used to solve specific problems. For example, various modeling techniques are used support management of water quality and nutrient pollution, as shown in the largest cluster (figure 4). On the other hand, the TWCR outcomes shed light on the dominant research topics. For example, quite some research addresses challenges related to climate change and water resource management, as well as to wastewater treatment (figure 5). Although the three bibliometric approaches are all valuable, they also lead to rather different maps, each providing only a partial image of the complex structure of the field. Understanding the relations and differences between the three clustering outcomes may therefore provide a more comprehensive picture. In the next section we will answer the question how the three maps can be combined, using cross tabulation.

Cross-tabular analysis

The main clusters derived from three different mapping approaches, i.e. the 12 largest JJCR clusters (J1-12; see annex 3), the 20 largest SAK clusters (C1-20; see annex 2), and the 16 largest TWCR clusters (T1-16; see annex 1), were cross-tabulated in a pair-wise manner (Table 4-6). The shading of the cells represents the distribution of a topic across fields (Table 4), of a method across fields (Table 5), and of a topic across methods (Table 6). The darker the cell, the more papers are in it. The number of fields (N) per topic shows the variety of the fields a topic is embedded in. The same holds for the number of methods by field. The lower the *variety* (N), the more disciplinary a topic is. The CoV shows whether the topics are or aren't evenly distributed over the fields: a high or low *balance* (Porter & Rafols 2009). A topic is less multidisciplinary if the CoV is high: then one or two fields are dominant, and the role of the other fields is small. When the CoV is low, the papers of a topic are evenly distributed over the relevant fields, indicating a high balance and therefore a higher level of multi-disciplinarity. In other words, these two indicators can be used to measure the level of multidisciplinary of a topic. High variety (High N) and high balance (low CoV) suggest a high level of multidisciplinary.

Apart from these more formal indicators of multi-disciplinarity, the three matrices also give substantive insight in the nature of the field. More specifically, triangulation by cross-tabulation shows in what fields the core topics are embedded, and what methods are deployed in these fields and topics, and additionally what fields these methods come from. Together with the findings of the three mapping methods, this cross-tabulation provides a detailed and multi-perspective set of maps of the field of water S&T. A few examples may illustrate the specific benefits of the triangulation through cross-tabulation.

(i) Topics by fields: most topics in the water research field are highly multidisciplinary, as they are embedded within the fields of water S&T, environmental science, and in some cases also in hydrology or desalination. The multidisciplinary topic of "eutrophication as a threat to water quality" (Topic 2, N=4, CoV=0.40, Table 4) is more or less equally distributed across four research fields including "environmental pollution", "water S&T", "marine biology", and "hydrology". Moderate multidisciplinary topics are water desalination (Topic 9, N=4, CoV=0.93) which is mainly in the field "desalination", and less strongly within "water S&T", and "stable isotopes in Holocene hydrological cycles" (Topic 14, N=3, CoV=0.85) which is in hydrology and geophysics. Another moderate to weak multidisciplinary topic is "the influence of climate change on hydrology cycle and water resource management" (Topic1, N=5, CoV=1.16, Table 4). This topic is mainly but not completely concentrated within the discipline of "hydrology", next to four other marginally involved fields. At the other extreme we find the mono-disciplinary topic "Water management aspects of PEM fuel cells" (Topic 16, N=1, CoV=0, Table 4), which is almost completely within the electrochemistry field

(ii) Methods by fields: The cross-tabulation shows what methods are used were, but also the disciplinary background of methods. "Anaerobic digestion of wastewater sludges" (Method 11, N=5, CoV=0.57, Table 5) is one of the broader deployed methods, as it is used by three water research fields ("water S&T", "environment and pollution", "desalination"), with a firm basis in "applied biotechnology/ biochemistry" and "chemical engineering". On the other hand, "water resource system management models" (Method 20, N=3, CoV=1.2, Table 5) shows a monodisciplinary pattern, as it is mainly used in hydrology research.

(iii) Methods used in topics. Also, a different pattern can be identified. The topic of "aerobic wastewater treatment" (Topic 3, N=8, CoV=0.41, Table 6) involves a most diverse set of methods, such as "nitrification and denitrification for nitrogen removal", "activated-sludge process", "anaerobic digestion of wastewater sludges", "advanced oxidation technologies for removal of pharmaceuticals", "nitrogen management to uphold water quality", "degradation by oxidation kinetics", "filtration/ ultrafiltration/ microfiltration for removal of natural organic matter" and "modeling for the management of water quality and nutrient pollution". In the same way, the triangulation shows the multi-method nature of quite some of the other topics. Conversely, "advanced oxidation technologies" is mainly deployed in the topic of "pharmaceuticals (antibiotics) extraction in water treatment processes" (Topic 5, N=3, CoV=1.26, Table 6).

Summing up, our triangulation approach reveals the multidisciplinary and complex web of fields, methods/problems, and the topical focus of water research.

Conclusion and discussion

In this paper, we have presented a triangulation approach to mapping research fields. In contrast to other methods, we do not restrict the map to show the structure of a field at various levels of granularity, or its degree of interdisciplinarity. Our mapping exercise also shows the content of the field and the methods and approaches selected. Furthermore, the method indicates the complex relations between fields, topics and methods and therefore the measurement of multi-disciplinarity remains not only a formal characteristic ("the level of multi-disciplinarity"), but it gets substantial meaning.

- Using Web of Science data, we have mapped the field of water research using three commonly used bibliometric methods: journal clustering through journal-journal citation relations (JJCR), paper clustering using shared author keyword (SAK), and paper clustering through shared title word-cited reference combinations (TWCR). We showed that the three resulting networks represent different phenomena at different levels. Journals are channels of scholarly communication, and show the disciplinary environment of water research. JJCR networks map the range of disciplines involved in and relevant for a scientific field. Paper maps based on keyword similarity show the methods, techniques and materials used in research. Paper networks based on TWCR closely represent the topics that researchers actually study.
- In both paper-based maps, we found a distinction between water management and water technology 'regions', with specific research fronts and specific method-problem clusters within each of the two regions.
- By translating the journal map into a paper map, we could identify the three dimensional overlap of the papers clusters. This underlies the proposed triangulation approach. The three mapping approaches and their integration tell us various interesting properties of the field of water science and technology: (i) In that field water S&T is embedded, and what neighboring field the knowledge sources for water S&T are: the journal map; (ii) What topics are studied within water S&T: the TWCR map, showing the divide between water management and waste water treatment (iii) What the dominant methods are: the SAK map, showing the dominance of modeling, and a large amount of water cleaning techniques; (iv) Which of the topics are multidisciplinary and which not, and more importantly, what disciplines they are embedded in (topics by field cross table); (v) What methods are deployed in the various topics (the methods by topic cross table); and finally (vi) From which fields the methods are imported (the methods by field cross table). Here we see the dominant role of biotech and several chemistry fields.
- In this paper we have used the method to develop a three dimensional map of water science and technology *at one moment in time*. However, as shown elsewhere (Vugteveen et al., 2014), it would be relevant to add a dynamic perspective by relating the journal, keyword, and paper networks at moment T to the same networks at moment T-1. Within the dynamics of science, the convergence and divergence over time of keywords can be used as indicators of scientific specialization. A changing topic structure gives another perspective on the dynamics of research fronts. Finally, changing journal clusters indicate the changing disciplinary structure of science. Then we may be able to investigate how dynamics at the three levels influence dynamics at the other levels. From this perspective, the current paper not only offers a method for making

multidimensional *static maps* of research fields, but it may also contribute to the study of *knowledge dynamics*.

Overall, our triangulation method successfully integrated the results from multiple maps and enabled a meaningful interpretation that adds to the findings from the individual maps at the different levels. The proposed triangulation method results in a better understanding of the complicated network of water research.

Acknowledgements

The authors wish to thank the ASIS&T SIIGMET student paper contest 2011 committee, and the three anonymous reviewers for their constructive feedback on earlier versions. The authors are also grateful to professor Wim van Vierssen (KWR Water Research Institute and Delft University) for valuable discussions and advice, and to the researchers who participated in the expert meeting at KWR Water Research Institute.

References

Ahlgren, Per, & Colliander, Cristian. (2009). Document–document similarity approaches and science mapping: experimental comparison of five approaches. *Journal of Informetrics, 3*(1), 49-63.

Altrichter, Herbert, Posch, Peter, & Somekh, Bridget. (1993). *Teachers investigate their work: An introduction to the methods of action research*: Psychology Press.

Åström, Fredrik. (2002). Visualizing library and information science concept spaces through keyword and citation based maps and clusters. *Bruce, F.; Ingwersen, P. et al*, 185-197.

Baldwin, Clive, Hughes, Julian, Hope, Tony, Jacoby, Robin, & Ziebland, Sue. (2003). Ethics and dementia: mapping the literature by bibliometric analysis. *International journal of geriatric psychiatry, 18*(1), 41-54.

Bastian, Mathieu, Heymann, Sebastien, & Jacomy, Mathieu. (2009). Gephi: an open source software for exploring and manipulating networks. Paper presented at the ICWSM.

Bauin, Serge. (1986). Aquaculture: A field by bureaucratic fiat. *M. Callon, J. Law, & A. Rip (Eds.), Mapping the dynamics of sriencr and technology: Sociology ofscience in the real world*, 124-141.

Blondel, Vincent D, Guillaume, Jean-Loup, Lambiotte, Renaud, & Lefebvre, Etienne. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008.

Börner, Katy, Chen, Chaomei, & Boyack, Kevin W. (2003). Visualizing knowledge domains. *Annual review of information science and technology, 37*(1), 179-255.

Börner, Katy, Klavans Richard, Patek Michael, Zoss, Angela M, Biberstine, Joseph R, Light, Robert P, Larivière, Vincent, & Boyack, Kevin W. (2012). Design and Update of a Classification System: The UCSD Map of Science. *PLoS ONE, 7*(7), e39464.

Boyack, Kevin W, & Klavans, Richard. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389-2404.

Boyack, Kevin W, Klavans, Richard, & Börner, Katy. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351-374.

Boyack, Kevin W, Newman, David, Duhon, Russell J, Klavans Richard, Patek Michael, Biberstine, Joseph R, Schijvenaars, Bob, Skupin Andre, Ma, Nianli, & Börner, Katy. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. PLoS One *6*, e18029

Braam, Robert R, Moed, Henk F, & Van Raan, Anthony FJ. (1991). Mapping of science by combined co-citation and word analysis, I. Structural aspects. *JASIS, 42*(4), 233-251.

Callon, Michel, Courtial, Jean-Pierre, Turner, William A, & Bauin, Serge. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social science information, 22*(2), 191-235.

Callon, Michel, Law, John, & Rip, Arie. (1986). *Mapping the dynamics of science and technology*: Springer.

Cohen, L , & Manion, L. (2000). Research methods in education (5th Edition). *London: Rouledge-Palmer*.

Ding, Ying, Chowdhury, Gobinda G, & Foo, Schubert. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management, 37*(6), 817-842.

Garfield, Eugene. (1972). *Citation analysis as a tool in journal evaluation*.

He, Qin. (1999). Knowledge Discovery through Co-Word Analysis. *Library trends, 48*(1), 133-159.

Janssens, Frizo, Zhang, Lin, Moor, Bart De, & Glänzel, Wolfgang. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management, 45*(6), 683-702.

Jarneving, Bo. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics, 65*(2), 245-263.

Klavans, Richard, & Boyack, Kevin W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology, 57*(2), 251-263.

Lee, Pei-Chun, & Su, Hsin-Ning. (2010). Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. *Innovation: Management, Policy & Practice, 12*(1), 26-40.

Leydesdorff, Loet, & Cozzens, Susan. (1993). The Delineation of Specialties in Terms of Journals Using the Dynamic Journal Set of the Science Citation Index. *Scientometrics, 26*, 133-154.

Leydesdorff, Loet, & Probst, Carole. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. *Journal of the American Society for Information Science and Technology, 60*(8), 1709-1718.

Leydesdorff, Loet & Rafols, Ismael. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348-362.

Lu, Kun, & Wolfram, Dietmar. (2012). Measuring author research relatedness: A comparison of word☐based, topic☐based, and author cocitation approaches. *Journal of the American Society for Information Science and Technology, 63*(10), 1973-1986.

Martin, Shawn, Brown, W Michael, Klavans, Richard, & Boyack, Kevin W. (2011). *OpenOrd: an open-source toolbox for large graph layout.* Paper presented at the IS&T/SPIE Electronic Imaging.

Morris, Steven A, & Van der Veer Martens, Betsy. (2008). Mapping research specialties. *Annual review of information science and technology, 42*(1), 213-295.

Morris, Steven A, & Yen, Gary G. (2004). Crossmaps: Visualization of overlapping relationships in collections of journal papers. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl 1), 5291-5296.

Olsen, Wendy. (2004). Triangulation in social research: qualitative and quantitative methods can really be mixed. *Developments in sociology, 20*, 103-118.

Porter, Alan L, & Rafols, Ismael. (2009). Is Science Becoming more Interdisciplinary? Measuring and Mapping Six Research Fields over Time. *Scientometrics,* 81(3), 719-45.

Shibata, Naoki, Kajikawa, Yuya, Takeda, Yoshiyuki, & Matsushima, Katsumori. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology, 60*(3), 571-580.

Su, Hsin-Ning, & Lee, Pei-Chun. (2010). Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics, 85*(1), 65-79.

Van den Besselaar, Peter. (2000). Communication between science and technology studies journals: A case study in differentiation and integration in scientific fields. *Scientometrics, 47*(2), 169-193.

Van Den Besselaar, Peter, & Gurney, Thomas. (2009). Regenerative medicine: an emerging field. *Den Haag: Rathenau Instituut.*

Van Den Besselaar, Peter, & Heimeriks, Gaston. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics, 68*(3), 377-393.

Van den Besselaar, Peter, & Heimeriks, Gaston. (2001). Disciplinary, Multidisciplinary, Interdisciplinary: Concepts and Indicators. *M. Davis & C.S. Wilson (Eds.), ISSI 2001, 8th international conference of the Society for Scientometrics and Informetrics, Sydney: UNSW 2001*, 705-716.

Van den Besselaar, Peter, & Leydesdorff, Loet. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. *Journal of the American Society for Information Science, 47*(6), 415-436.

Vugteveen, Pim, Lenders, Rob, & Van den Besselaar, Peter. (2014). The dynamics of interdisciplinary research fields – the case of river. *Scientometrics, 100*, 73-96.

Whittaker, John. (1989). Creativity and conformity in science: titles, keywords and co-word analysis. *Social Studies of Science, 19*(3), 473-496.

Zitt, Michel, & Bassecoulard, Elise. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information processing & management, 42*(6), 1513-1531.

Footnotes

1. We have used the five citation databases of Thomson Reuters Web of Science: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index-Science, and Conference Proceedings Citation Index-Social Science & Humanities.

2. SAINT Toolkit is a set of data processing tools for bibliometric research, developed by Rathenau Institute. http://www.rathenau.nl/themas/project/bibliometrische-softwaretools/saint.html (Andre Somers, Thomas Gurney, Edwin Horlings, and Peter van den Besselaar, “A bibliometric toolbox for analyzing knowledge dynamics”, Rathenau Instituut, 2009).

3. KWR Watercycle Research Institute (http://www.kwrwater.nl/) is a private R&D firm whose shareholders are publicly owned Dutch drinking water companies.

4. Stop words were removed.

5. Water Science and Technology, Journal of Hazardous Materials, Water Research, Desalination, Bioresource Technology, Environmental Science & Technology, Journal of Hydrology, Water Resources Research, Chemosphere, Hydrological Processes, Chemical Engineering Journal, Environmental Monitoring and Assessment, Water Resources Management, Science of the Total Environment, Agricultural Water Management, Water Air and Soil Pollution, Hydrology and Earth System Sciences, Water Environment Research, Fresenius Environmental Bulletin, Journal of Environmental Sciences-China, Journal of Membrane Science, Industrial & Engineering Chemistry Research, International Journal of Hydrogen Energy, and Ecological Engineering.

Table 1. Initially selected search terms

Initial search terms PY=2008-2009 (October, 2010)	Numb. of hits	Numb. of unique hits	% of unique hits	Impact on initial dataset
TS=water treat*	28991	17133	60%	48%
TS=water quality	15545	9703	62%	26%
TS=drinking water	6836	3239	47%	11%
TS=(waste water OR wastewater)	16919	9098	54%	28%
TS=desalinat*	1029	637	62%	2%
TS=hydrolog*	7555	6465	86%	13%
Sum	76375	46275		128%
Combination (total number of documents)	60162		77%	

Table 2. Selected search term combinations for extracting the final dataset.

Revised search terms PY=2010 (April, 2013)	Number of hits	Number of unique hits	% of unique hit	Impact on initial dataset
"water treat*"	2,279	694	30%	10%
"water quality"	3,235	1,766	55%	14%
"drinking water"	3,202	1,901	59%	14%
("waste water" OR wastewater)	7,163	4,560	64%	31%
desalinat*	608	414	68%	3%
hydrolog*	3,868	2,682	69%	17%
"water cycle*"	241	115	47%	1%
"water system*"	954	671	70%	4%
"water management"	1,003	446	44%	4%
(sewer* OR sewage)	2,493	1,147	46%	11%
"water distribution"	491	249	51%	2%
"water suppl*"	1,352	532	39%	6%
"water sanitation"	19	10	55%	0%
"water resource*"	1,741	615	35%	7%
"water quantity"	106	20	19%	0%
"water demand"	256	65	25%	1%
"water policy"	87	25	29%	0%
"water sustainab*"	20	4	18%	0%
("climate change" AND water)	2,217	1,205	54%	9%
("global warming" AND water)	350	150	43%	1%
Sum	31,685	17,271		135%
Combination	23,406		74%	

Table 3. The number of papers studied by the three bibliometric methods, and the overlap.

Method			Covered by JJCR (28 clusters)	Covered by SAK (765 clusters)	Covered by TWCR (192 clusters)
JJCR	Covered	10864	-	4856	10457
	Not covered	12065	-	4284	10433
AWKC	Covered	9140	4856	-	8772
	Not covered	13789	6008	-	12118
TWCR	Covered	20890	10457	8772	-
	Not covered	2039	407	368	-
Total		22929	10864	9140	20890

Table 4. Topics by field

	T1	T2	T3	T4	T5	T6	TT	T8	T9	T10	T11	T12	T13	T14	T15	T16	N
Marine biology		0.2															1
Electrochemistry								0.1							0.5	0.9	3
Chemistry				0.1													1
Irrigation	0.1																1
Soil agriculture							0.2							0.1			2
Geophysics	0.2													0.3			2
Chemical engineering				0.1				0.1	0.1				0.1				4
Applied biotechnology/ biochemistry			0.3	0.1		0.2	0.1	0.1	0.1	0.1		0.1			0.4		9
Desalination				0.2				0.1	0.5	0.1			0.1				5
Hydrology	0.6	0.1				0.1				0.1	0.6			0.6			6
Environmental pollution	0.1	0.3	0.1	0.1	0.5	0.3	0.4	0.2		0.3	0.1	0.6	0.3				12
Water Science & Technology	0.1	0.2	0.5	0.4	0.3	0.4	0.2	0.4	0.3	0.4	0.2	0.3	0.4		0.1		14
Coefficient of Variation = balance	1.16	0.40	0.68	0.65	0.41	0.47	0.57	0.77	0.93	0.85	0.99	0.87	0.80	0.85	0.51	n.a.	
N(umber of fields) = variety	5	4	3	6	2	4	4	6	4	5	3	3	4	3	3	1	

Values below 5% were omitted - rest rounded up of decimals

Table 5. Methods by field

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	N
Marine biology	0.1									0.1											2
Electrochemistry														0.2				0.1			2
Chemistry							0.1		0.1												2
Irrigation	0.1		0.2													0.1	0.2			0.1	5
Soil agriculture																	0.1				1
Geophysics																	0.1				1
Chemical engineering							0.1	0.1	0.1		0.1		0.1					0.1	0.1		7
Desalination		0.1		0.4	0.1	0.1	0.2	0.1	0.1		0.1				0.1				0.1		10
Appl biotech / biochemi		0.1		0.1	0.1		0.1	0.2	0.1	0.1	0.3	0.1	0.2	0.3	0.2			0.3	0.2		14
Hydrology	0.2		0.6			0.2				0.1			0.1			0.6	0.4			0.7	8
Environmental pollution	0.3	0.3	0.1	0.1	0.3	0.3	0.1	0.1	0.1	0.3	0.2	0.3	0.1	0.1	0.3	0.2	0.1	0.1	0.3	0.1	20
Water S&T	0.2	0.4	0.1	0.3	0.3	0.4	0.5	0.5	0.4	0.3	0.3	0.5	0.6	0.3	0.5	0.1	0.1	0.3	0.4		19
Coeff of Variation = balance	0.56	0.76	1.04	0.85	0.52	0.59	0.91	0.85	0.87	0.76	0.57	0.74	1.14	0.47	0.67	0.97	0.82	0.68	0.70	1.21	
N(umber of fields) = variety	5	4	4	4	4	4	6	5	6	5	5	3	5	4	4	4	6	5	5	3	

Table 6. Topics by method

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	N
Water resources system management models											0.1					1
Decolorization and biodegradation of dye wastewaters								0.1								1
Electricity generation from wastewater								0.1								1
Hydrological and vegetation modeling	0.1															1
Optimization models concerning system uncertainty											0.5					2
Antioxidants against metal toxicity										0.2		0.6				2
Anaerobic digestion of wastewater sludges			0.1				0.1								0.6	3
Nitrification and denitrification for nitrogen removal			0.2													1
Drinking water disinfection and chlorination process						0.2							0.2			3
Degradation by oxidation kinetics			0.1		0.1		0.1	0.2							0.1	5
Nitrogen management to uphold water quality		0.1	0.1				0.1									4
Adsorption to remove heavy metal ions from aqueous solutions				0.2				0.1							0.1	3
Activated-sludge process			0.2						0.2						0.1	4
Activated carbon adsorption (biosorption) to remove heavy metal ions from aqueous solutions				0.4						0.1						2
Adsorption for contaminants in drinking water	0.1	0.1				0.1		0.1		0.4	0.1	0.1	0.1	0.2		9
Adsorption for removal of heavy metal				0.1		0.1	0.4			0.2		0.1				5
Filtration/ Ultrafiltration/ Microfiltration for removal of natural organic matter			0.1	0.1	0.1	0.1		0.1	0.6				0.2	0.1	0.1	10
Modeling the impact of climate variability and climate change in water management	0.5	0.1				0.1	0.1				0.1			0.4		6
Advanced Oxidation Technologies (solid-phase extraction, tandem mass-spectrometry, TiO2, etc.) for removal of pharmaceuticals		0.1	0.1	0.1	0.5	0.2	0.1	0.4			0.1		0.1	0.1	0.1	11
The management of water quality and nutrient pollution (modeling the relationship between land use and water quality)	0.2	0.5	0.1			0.1	0.1			0.1	0.1		0.1	0.2		9
Coefficient of Variation = balance	0.99	1.1	0.41	0.81	1.26	0.47	0.89	0.98	0.74	0.69	1.02	0.98	0.51	0.84	1.35	1.22
N(umber of fields) = variety	4	5	8	5	3	7	7	7	2	5	6	3	5	5	6	5

Annex 1. Method clusters (SAK):

M1	Modeling for the management of water quality and nutrient pollution
M2	Advanced Oxidation Technologies for removal of pharmaceuticals
M3	Modeling the impact of climate variability and climate change in water management
M4	Filtration/ Ultrafiltration/ Microfiltration for removal of natural organic matter
M5	Adsorption for removal of heavy metal
M6	Adsorption for contaminants in drinking water
M7	Activated carbon adsorption/ biosorption to remove heavy metal ions
M8	Activated-sludge process
M9	Adsorption to remove heavy metal ions from aqueous solutions
M10	Nitrogen management to uphold water quality
M11	Degradation by oxidation kinetics
M12	Drinking water disinfection and chlorination process
M13	Nitrification and denitrification for nitrogen removal
M14	Anaerobic digestion of wastewater sludges
M15	Antioxidants against metal toxicity
M16	Optimization models concerning system uncertainty
M17	Hydrological and vegetation modelling
M18	Electricity generation from wastewater
M19	Decolorization and biodegradation of dye wastewaters
M20	Water resources system management models

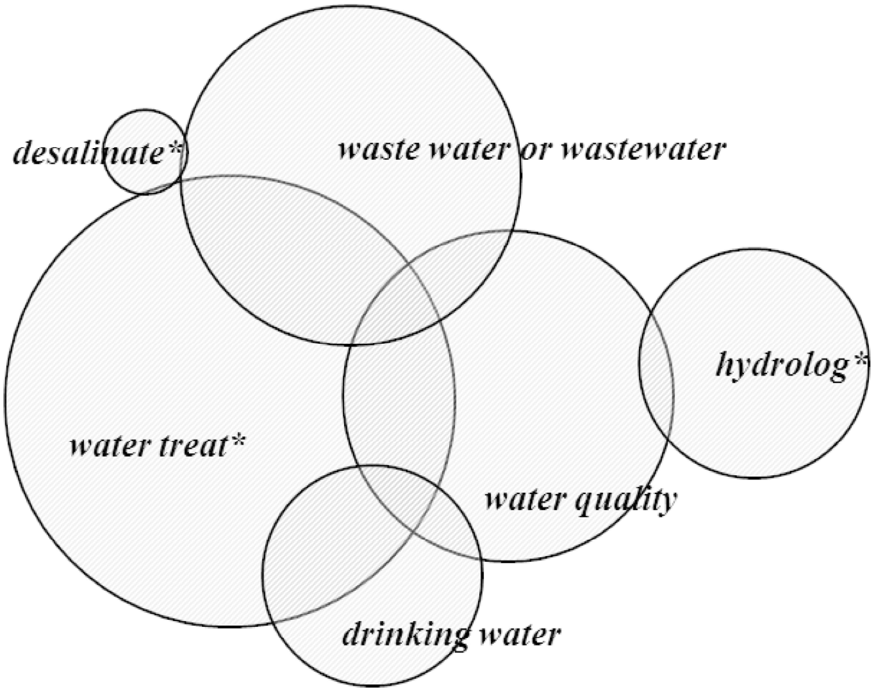
Annex 2. Topic clusters (TWCR):

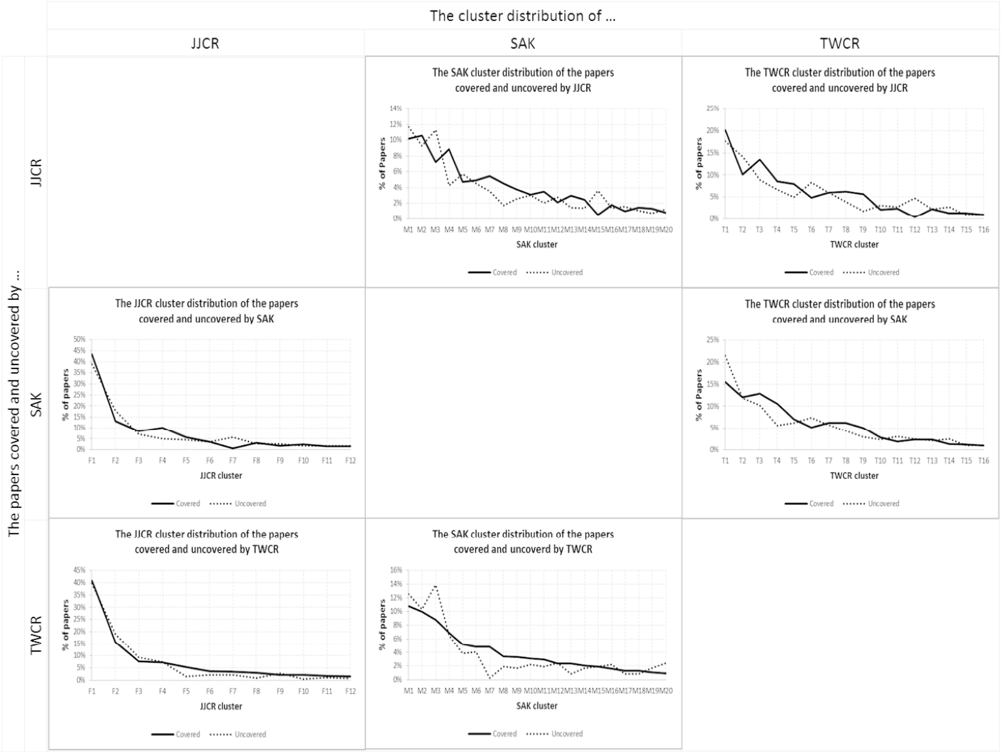
T1	The influence of climate change on hydrology cycle and water resource management
T2	Eutrophication as a threat to water quality
T3	Aerobic wastewater treatment (nitrification & denitrification, constructed wetlands, activated sludge, and anaerobic digestion)
T4	Adsorption kinetics and adsorption isotherms in wastewater treatment
T5	Pharmaceuticals (antibiotics) extraction in water treatment processes
T6	Biofilms and bacterial drinking water quality
T7	Heavy metals in sewage sludge composting
T8	Electrochemical (photocatalysis and electrocoagulation) treatment in wastewater
T9	Water desalination (membrane, reverse-osmosis, nanofiltration, etc.)
T10	Arsenic adsorption & removal from groundwater & drinking water
T11	Optimization in water distribution systems using genetic algorithms
T12	Oxidative stress and toxic metals
T13	Disinfection by-products and natural organic matter in drinking water

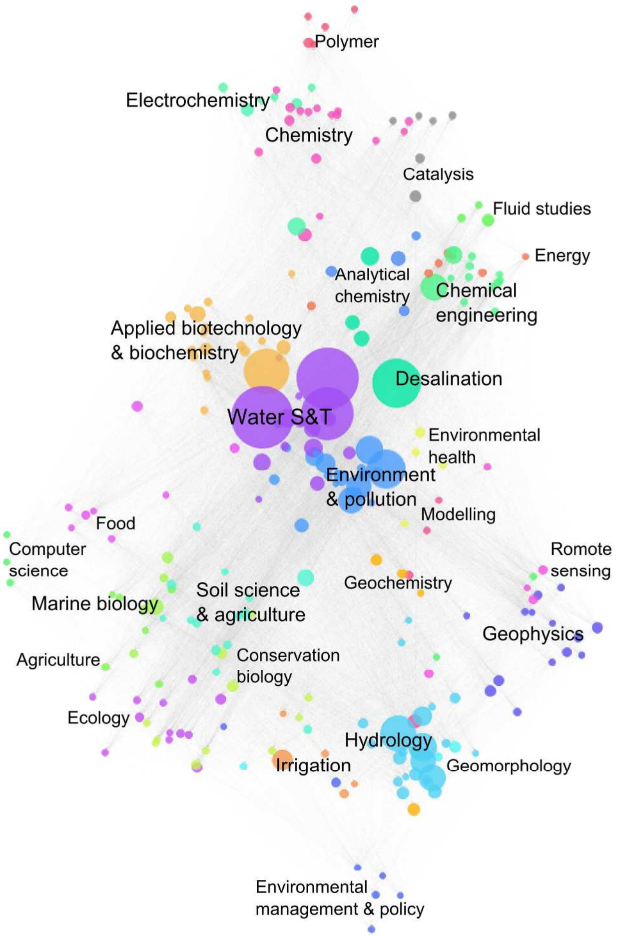
T14	Stable isotopes in Holocene hydrological cycles
T15	Hydrogen production
T16	Water management studies in PEM fuel cells

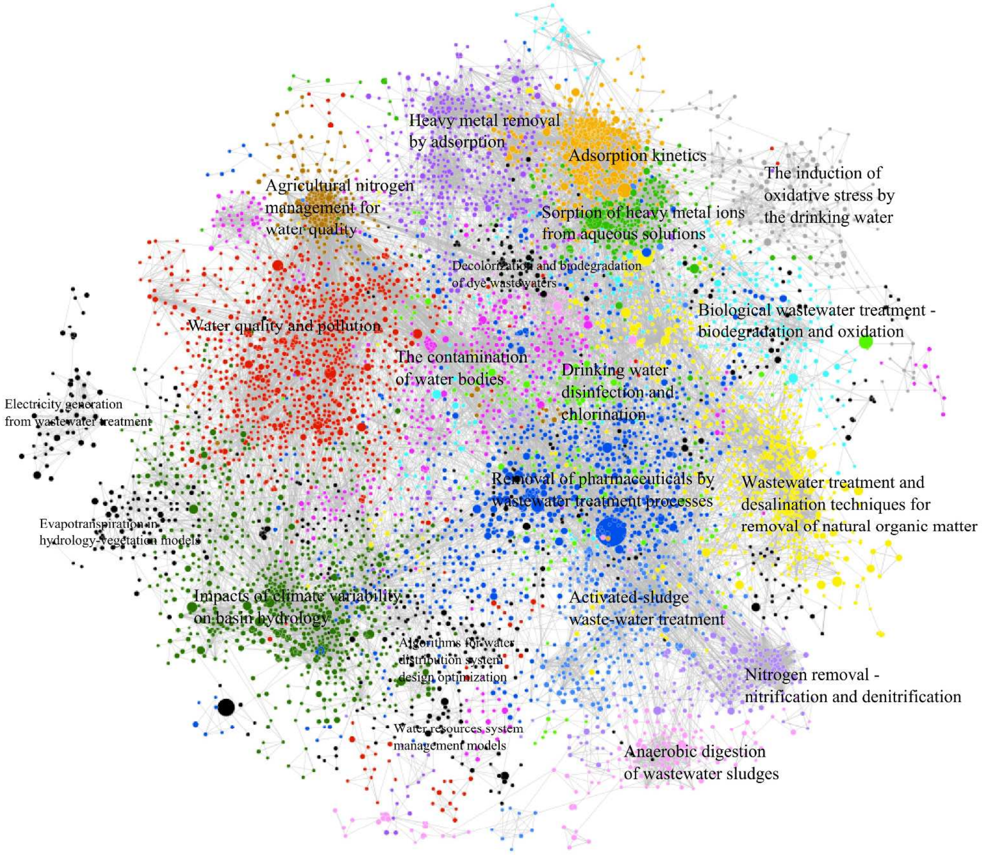
Annex 3. Journal clusters (JJCR)

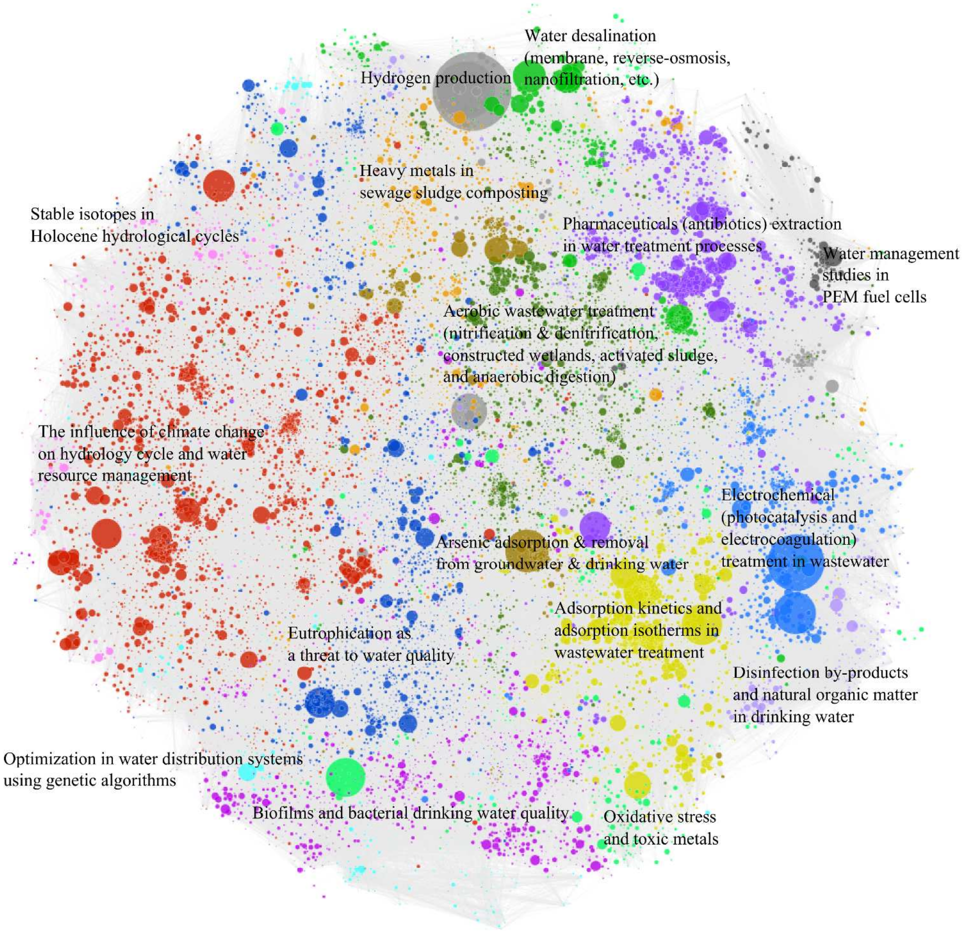
F1	Water Science and technology
F2	Environment and pollution
F3	Hydrology
F4	Desalination
F5	Applied biotechnology & biochemistry
F6	Chemical engineering
F7	Geophysics
F8	Soil science & agriculture
F9	Irrigation
F10	Chemistry
F11	Electrochemistry
F12	Marine biology


desalinate*
waste water or wastewater
hydrolog*
water treat*
water quality
drinking water


The cluster distribution of ...
JJCR
SAK
TWCR
The papers covered and uncovered by ...
JJCR
SAK
TWCR
The SAK cluster distribution of the papers covered and uncovered by JJCR
% of Papers
SAK cluster
The TWCR cluster distribution of the papers covered and uncovered by JJCR
% of papers
TWCR cluster
The JJCR cluster distribution of the papers covered and uncovered by SAK
% of papers
JJCR cluster
The TWCR cluster distribution of the papers covered and uncovered by SAK
% of Papers
TWCR cluster
The JJCR cluster distribution of the papers covered and uncovered by TWCR
% of papers
JJCR cluster
The SAK cluster distribution of the papers covered and uncoverd by TWCR
% of papers
SAK cluster
Covered
Uncovered


Polymer
Electrochemistry
Chemistry
Catalysis
Fluid studies
Energy
Analytical chemistry
Chemical engineering
Applied biotechnology & biochemistry
Desalination
Water S&T
Environmental health
Environment & pollution
Modelling
Food
Computer science
Romote sensing
Geochemistry
Soil science & agriculture
Marine biology
Geophysics
Agriculture
Conservation biology
Ecology
Hydrology
Irrigation
Geomorphology
Environmental management & policy


Heavy metal removal by adsorption
Adsorption kinetics
The induction of oxidative stress by the drinking water
Agricultural nitrogen management for water quality
Sorption of heavy metal ions from aqueous solutions
Decolorization and biodegradation of dye wastewaters
Biological wastewater treatment - biodegradation and oxidation
Water quality and pollution
The contamination of water bodies
Drinking water disinfection and chlorination
Electricity generation from wastewater treatment
Removal of pharmaceuticals by wastewater treatment processes
Wastewater treatment and desalination techniques for removal of natural organic matter
Evapotranspiration in hydrology-vegetation models
Impacts of climate variability on basin hydrology
Activated-sludge waste-water treatment
Algorithms for water distribution system design optimization
Nitrogen removal - nitrification and denitrification
Water resources system management models
Anaerobic digestion of wastewater sludges


Hydrogen production
Water desalination
(membrane, reverse-osmosis,
nanofiltration, etc.)
Heavy metals in
sewage sludge composting
Stable isotopes in
Holocene hydrological cycles
Pharmaceuticals (antibiotics) extraction
in water treatment processes
Water management
studies in
PEM fuel cells
Aerobic wastewater treatment
(nitrification & denitrification,
constructed wetlands, activated sludge,
and anaerobic digestion)
The influence of climate change
on hydrology cycle and water
resource management
Electrochemical
(photocatalysis and
electrocoagulation)
treatment in wastewater
Arsenic adsorption & removal
from groundwater & drinking water
Adsorption kinetics and
adsorption isotherms in
wastewater treatment
Eutrophication as
a threat to water quality
Disinfection by-products
and natural organic matter
in drinking water
Optimization in water distribution systems
using genetic algorithms
Biofilms and bacterial drinking water quality
Oxidative stress
and toxic metals