Wiley InterScience
DISCOVER SOMETHING GREAT

Citation Data Analysis on Hydrogeology

Frank Schwartz and Y.C. Fang
Department of Geological Sciences, The Ohio State University, Columbus, OH 43210.
E-mail: frank@geology.ohio-state.edu, fang.29@osu.edu

This article explores the status of research in hydrogeology using data mining techniques. First we try to explain what citation analysis is and review some of the previous work on citation analysis. The main idea in this article is to address some common issues about citation numbers and the use of these data. To validate the use of citation numbers, we compare the citation patterns for *Water Resources Research* papers in the 1980s with those in the 1990s. The citation growths for highly cited authors from the 1980s are used to examine whether it is possible to predict the citation patterns for highly-cited authors in the 1990s. If the citation data prove to be steady and stable, these numbers then can be used to explore the evolution of science in hydrogeology. The famous quotation, "If you are not the lead dog, the scenery never changes," attributed to Lee Iacocca, points to the importance of an entrepreneurial spirit in all forms of endeavor. In the case of hydrogeological research, impact analysis makes it clear how important it is to be a pioneer. Statistical correlation coefficients are used to retrieve papers among a collection of 2,847 papers before and after 1991 sharing the same topics with 273 papers in 1991 in *Water Resources Research*. The numbers of papers before and after 1991 are then plotted against various levels of citations for papers in 1991 to compare the distributions of paper population before and after that year. The similarity metrics based on word counts can ensure that the "before" papers are like ancestors and "after" papers are descendants in the same type of research. This exercise gives us an idea of how many papers are populated before and after 1991 (1991 is chosen based on balanced numbers of papers before and after that year). In addition, the impact of papers is measured in terms of citation presented as "percentile," a relative measure based on rankings in one year, in order to minimize the effect of time.

Introduction

Over the last several years, the structure of research in the hydrologic sciences has been examined through citation analysis (Schwartz & Ibaraki, 2001; Schwartz, Fang, & Ibakari, 2002). A tabulation of citation data for papers in journals like *Water Resources Research*, *Journal of Hydrology*, and *Ground Water*, to name a few, determined that only a small fraction of papers were well cited with most papers having less than 15 citations over their entire life time. The premise of our study is in line with conventional thinking, namely that for any given paper, citations serve as a measure of impact of that work (Garfield, 1992). A small number of influential papers attract most of the citations and many "ordinary" papers attract some minimal number. Pioneering work on citation analysis (Garfield, 1998) showed that among the top 1000 most highly cited scientists in the world from 1965 to 1978, 250 are members of the National Academy of Sciences, and 125 are from foreign academies. A list of the 50 most highly cited scientists in 1967 included 12 Nobel laureates and numerous winners of honorary scientific awards. That is, the number of citations to publications can be a good indicator of the impact of the scientific work in that publication (Garfield, 1998).

These ideas remain controversial for a variety of reasons. First, some individuals disagree with the fundamental premise and think that citations are an imperfect measure of research impact. They posit examples of research that has social relevance but is lacking in the way of citations. Also, there are examples of excellent scientific work, in top journals, that apparently does not attract citations. Confusion arises because many equate the term *research impact* with research excellence, a point that will be taken up later. A second general argument against citation analysis stems from a lack of confidence in the reliability of the underlying numbers. Several recent studies (Adam, 2002; Nature, 2002) have pointed out the potential for simply miscounting citations because of spelling errors in names and similar problems. Other issues include, for example, inflation in citation statistics due to self-citation or citation circles.

Citation numbers are commonly (but mistakenly) considered to measure excellence in a piece of scientific research. The argument is compelling—excellent papers in excellent journals often receive many citations, while poor papers in poor journals remain uncited. However, there are apparent inconsistencies demonstrated by large numbers of relatively poorly cited papers in excellent journals like *Water Resources Research*. Authors of papers meeting the high standard of

Received June 23, 2005; revised April 3, 2006; accepted April 3, 2006

 • Published online 11 January 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20526

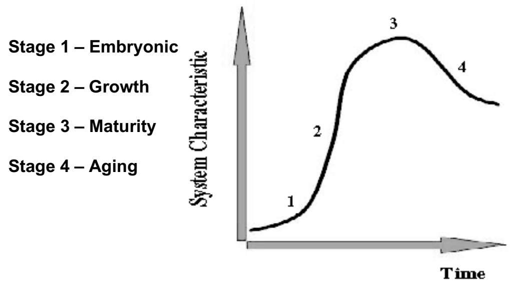

FIG. 1. Stages in a product life-cycle curve.

quality and creativity of the journal conclude that their work is of high quality, irrespective of the citation rate (Schwartz, Fang, & Parthasarathy, 2005). The objective of this study then is to reanalyze citation data from *Water Resources Research* and other journals to further understand the concept of research impact. The hypothesis here is that citations not only are a function of the excellence of the paper, but of other factors that remain to be defined.

Schwartz and Ibaraki (2001) proposed that research strands follow "product life-cycle curves" similar to manufactured products or biological population. The stages of the life-cycle curve (see Figure 1) show the product evolution from inception, through a growth phase, maturity, and old age/decline. The life cycle for most common papers in this field is about 10 years (Schwartz & Ibaraki). The most innovative papers with great new concepts are commonly cited for several decades.

Fang, Schwartz, and Parthasarathy (2001) tabulated citations for several key journals in hydrology. Figure 2 shows the trend lines of median citations for five selected hydrological journals (*Environmental Science and Technology*, *EST*; *Water Resources Research*, *WRR*; *Ground Water*, *GW*; *Journal of Hydrology*, *JH*; and *Journal of Contaminant Hydrogeology*, *JCH*). The results showed that *WRR* is the most cited journal among the five in terms of median number of citations in its early years, approximately before 1980. *Environmental Science and Technology* took over leadership in the early 1980s, with the emergence of new environment-related topics. Interestingly, *JCH* reached its peak one year after its first publication. *GW* and *JH* both bounced around a median number of citations of five for these years. The trend lines decline after 1990 because the more recent papers have accumulated fewer citations.

The key journals in the field of hydrogeology have been around since the mid-1960s. Although other journals have emerged in the water resources field, they are usually sponsored by for-profit publishers and cover much smaller niches. Thus, through time journals like *Water Resources Research* have maintained their leadership role and can be used with confidence to track the progress of this field through time.

The fields of hydrology and hydrogeology encompass all aspects of science dealing with water cycling through the land-based portion of the hydrologic cycle. Although work is published in a number of journals of varying quality, the preeminent journal for this science is *Water Resources Research* (*WRR*), one of the primary journals of the American Geophysical Union (AGU). The analyses presented in this article emphasize this journal because it consistently tracks the evolution of ideas and the best work in this field.

Fang, Schwartz, et al. (2001) and Schwartz et al. (2002) looked specifically at a few major research topics appearing in *WRR* and several subfields in ground-water modeling research. The results suggested that the research topics were an influential factor affecting citations. When a new research strand emerges, the rate of increase in the citation numbers grows rapidly but eventually declines after some time. This behavior again follows a life-cycle curve as described earlier for journals (Figure 1). Their analysis of citations also suggests that at least some traditional areas concerned with ground-water research have lost their research vigor and are beginning to show their age.

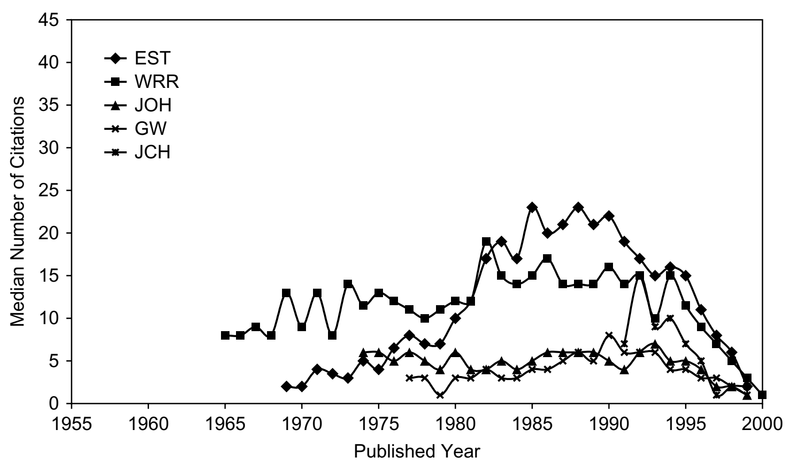

FIG. 2. Year trends of median citation for four selected hydrological journals.

DOI: 10.1002/asi

Methodology

Citation Information and Database Management

Full papers were available in electronic form for 14 years (1972, 1974, 1976, 1980, 1982, 1985, 1987, and 1990–1996) for *WRR*. These papers were published by AGU and available online from 1990 to the present. All of the papers from 1990–1996 were downloaded from AGU's website and converted into simple text files (no graphical content). We supplemented this database with seven additional years (1972, 1974, 1976, 1980, 1982, 1985, and 1987) of textual data from before 1990 by scanning the hardcopy of papers and creating text files using optical character recognition (OCR) software.

Citation counts associated with these 3,120 full-content papers were obtained from ISI Web of Science to the end of 2002. Because citation data are cumulative through time, they were normalized to remove temporal effects. Two modified citation metrics were used in this study: the relative citation index (RCI) and the total citation percentile (TCP). RCI is calculated by dividing the total citations for a paper by the average total citations of the papers published in the same year. An RCI of 1 indicates that the paper receives about the average citation compared with others in that year. TCP is based on the position of a given paper in a sequential ranking of papers in the same year. The effects of time within a year are neglected because the time lag of a paper from being published to being cited is assumed to be a year.

Citation data with the 3,120 full electronic papers are used for data mining and statistical analysis on word counts. This approach lets us find similar topic-related papers for various research strands. Another database consists solely of citation data for every paper published in *WRR* between 1965 and 2000. A total of 7,199 bibliometric records were downloaded through ISI Web of Science to the end of 2002. These data are used for an analysis to define relationships among "citing" and "cited" papers. Here, we find papers with more specific and direct relevance to one another without the requirement of providing full electronic papers.

Other related information (author, title, cited reference, number of citations, publication year and date, journal issue, etc.) for every downloaded paper was stored in our database for management and further information retrieval. These electronic papers were analyzed and tabulated by various types of statistical analyses.

Statistical Correlation Coefficient

This study also relies on correlation analysis based on word abundances to determine how similar papers are to one another. Papers with full contents are compared and analyzed by a correlation coefficient metric. These correlation coefficients are calculated based on counts of words. A set of n selected keywords forms a vector $\boldsymbol{X}$ with elements of counts of words obtained from an individual paper. That is,

$$\boldsymbol{X} = (X_{v1}, X_{v2}, X_{v3} \ldots\ldots X_{vn})$$

where X_{vi} is counts of a word vi.

The correlation between two papers is represented by

$$\mu_{XY} = \frac{\sum_{i=1}^{n} (X_{vi} - \overline{X})(Y_{vi} - \overline{Y})}{n\sigma_X\sigma_Y}$$

where μ_{XY} is the correlation coefficient between two papers X and Y. $\boldsymbol{Y}$ is the vocabulary vector for another paper as $\boldsymbol{X}$. $\overline{X}$ and $\overline{Y}$ are means for the two vectors; σ_X and σ_Y are standard deviations of the two vectors. We consider a coefficient of correlation above 60% to indicate that two articles are related. This value was determined by manually checking test sets for comparability.

Two sets of vocabularies as keywords are used in correlation coefficient analysis. The first set of vocabulary is collected from all the documents from the database. The set consists of more than 13,000 distinct words. The other set of vocabulary is provided by the domain expert with a total of 128 words. Details of word selection with associated weights can be found in Fang, Parthasarathy, and Schwartz (2001).

The criterion of 60% was determined heuristically. Through a series of manual checks, we determined that correlations above 70% yielded very similar papers, but missed obvious choices. Choosing a cutoff at below 50% provided many papers that were not acceptable matches.

Method I: Citation Growth Patterns

In this work, we compare the difference in performance between early papers and recent ones. There is no perfect way to conduct this comparison because the nature of cumulative citations is influenced by time variation. Therefore, we propose a method to compare patterns of citation growth by mapping the growth patterns from the early papers to more recent ones. In order to do this, the time for rate of change in citations of early papers is presented as relative time. For example, cumulative citations for two years after being published of the paper published in 1980 can be compared to the two-year history of a paper published in 2000 (all the citation information was downloaded at least two years after publication).

Highly cited papers from the 1980s are selected in order to compare with the citation patterns for highly cited papers in the 1990s. By doing this, it is possible to compare citations of papers in the 1990s with citations of the same type of excellent paper published a decade earlier. In other words, we set out to determine whether more recent papers have an impact comparable to that of earlier papers.

Because the database only stores total citation counts up to date for all the records, the citation growths for highly cited papers in the 1980s have to be downloaded individually for every year since their publication. Ten top 1% papers in the 1980s are chosen to explore their citation-growth patterns and compare with more recent papers.

In addition to highly cited papers in these two decades, topic-related papers are of interest as well. Two ways of

DOI: 10.1002/asi

finding topic-related papers for the two decades are statistical correlation on words (for the same research strand) and "citing" and "cited" relationships (for follow-on and pioneer). Comparisons of citation behaviors on topic-related papers between top 1% papers in the 1980s and highly cited (at least 97.5% percentile) papers in the 1990s can give an idea of the evolution of a research strand and the innovation in pioneering works on citation numbers.

Method II: Early Papers Versus Follow-on Papers

Another question we set out to answer was whether highly cited papers were "unique" relative to more poorly cited papers. Our hypothesis is that if a highly cited paper was truly seminal, it would be unlike papers that came before it and provide the impetus for similar, follow-on papers. An ordinary paper should be less innovative and be quite comparable with already existing work. As a consequence, this paper would resemble a group of similar papers from before and after. To test this hypothesis, we examined all papers published in *Water Resources Research* in 1991. Because of the way our database of electronic text is constructed, there are equal numbers of papers before and after the year 1991. The collection of 273 papers from 1991 was ordered in terms of the number of citations and divided into 10 bins that were defined as TCPs.

For example, papers falling in the 90% bin represented the top 10% of papers in terms of citations. Every one of these 1991 papers was compared with the entire collection of complete electronic papers in the entire *WRR* database. Those papers that correlated at better than 60%, using the textual comparison tool of similarity described previously, were separated into before and after piles. In other words, given some paper from 1991, we could determine where in time the most similar papers actually fell.

Data Analysis

The first analysis here is designed to show just how variable citations can be in any year of a journal like *WRR*. Figure 3 shows how many citations are required to place a paper in various percentiles from the 95th on down. Notice how nonlinear the patterns of citations actually are for *WRR*. The 95th percentile paper in any one year gets nearly 10 times more citations than the 25th percentile paper in the early 1980s. The bottom 25th percentile papers in their life are able to attract five or less citations, the 50th percentile papers about 15 citations. This plot is like comparable ones from Schwartz and Ibaraki (2001), reinforcing the idea of how rare it is for a paper to attract 50 citations. Also, this result raises the question as to how and why the top papers could attract so many citations and whether the behavior of the citation growth from the 1980s is in any way unique. To answer this question, we selected top 1% papers from the 1980s and used their citation growth in relative time to compare to the behavior of citation growth for papers in the 1990s. The method was described above.

The information collected in this study provides a basis for examining whether papers from the 1980s were more or less successful than the more recent papers from the 1990s in attracting citations. Moreover, one can determine by examining the subject of these important papers whether broad changes in research trends have occurred.

Figure 4 compares the citation growths (in relative time) for ten 99th percentile papers in the 1980s (red lines) and cumulative total citations (up to year 2002) of papers over 97.5th percentile in the 1990s (blue dots). The results are striking. The band formed by the ten 99th percentile papers in the 1980s covers the 97.5th percentile papers in the 1990s. The important messages from the figure are (1) the highly cited papers in the 1990s are following the same growth trajectory as those from the 1980s and (2) seeing the similarity in the patterns of citation implies that one can predict the likely citation success of more recent papers. This also confirms that the citation data are stable, at least starting in the early 1980s.

The data also facilitated an examination of how many follow-on papers related to early papers in the 1980s are among these highly cited papers in the 1990s. We use the statistical correlation coefficient as a means to correlate top papers from the 1980s with all the papers from the 1990s. Figure 5 is similar to Figure 4 except here there are only four

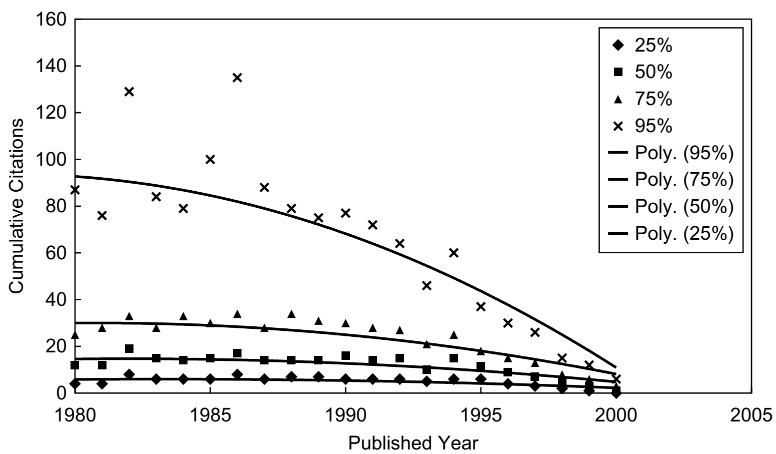

FIG. 3. Distributions of citations for papers in *WRR* for every year.

DOI: 10.1002/asi

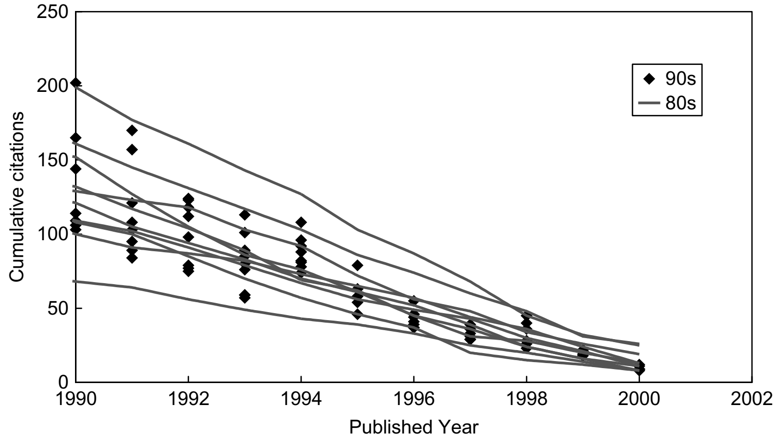

FIG. 4. Ten top 1% papers (red lines) from the 1980s versus highly cited papers from the 1990s (above 97.5th percentile).

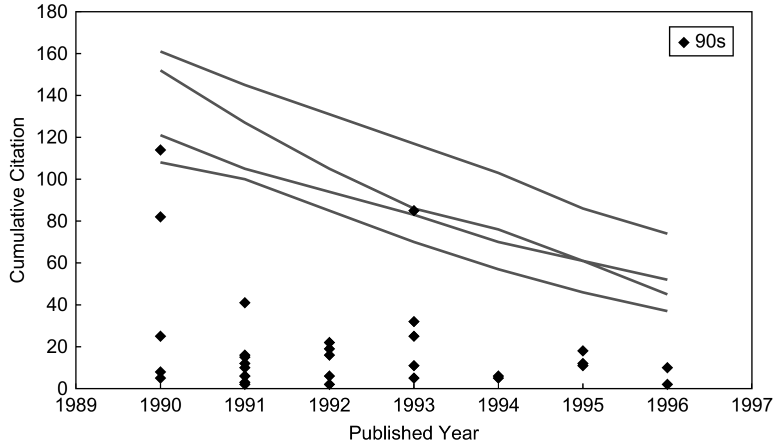

FIG. 5. Four top 1% papers (red lines) in the 1980s versus the papers highly correlated with them in the 1990s (correlation coefficient greater than 0.6).

red lines because only these years in the 1980s have full textual data available for correlation analysis. These lines again represent top papers from four years in the 1980s. There are fewer highly cited papers in the 1990s among all the blue dots (compared with Figure 4), which are related in topic to these four early papers (correlation coefficient greater than 0.6). It is clear that most of the blue dots fall below the red lines, in which the early papers simulate the citation growths for these younger papers. It is obvious that these topically relevant papers in the 1990s did not attract as many citations as similar papers from the 1980s.

In order to explore this interesting discovery in another way, we used reference information in our database to trace follow-on papers. With this alternative approach, we can examine all ten papers from the 1980s and their citers in the 1990s. It is again assumed that by citing a paper from the 1980s a later paper is follow-on. Not surprisingly, Figure 6 yields a similar pattern as Figure 5 in terms of the 10 early papers (red lines) and later follow-on papers (dots). These dots are labeled with various levels of green (from dark to light) to distinguish which references were cited. This figure shows that the follow-on papers from the 1990s that fall in among the collection of red lines (formed by the ten papers from the 1980s) are far fewer than those in Figure 4. The difference between Figures 4 and 6 suggests that most of the dots among the lines in Figure 4 are not the follow-on papers of these ten highly cited papers from the 1980s. They themselves represent new research strands that have formed and become highly cited.

Another noticeable finding from Figure 6 is that most of the dots falling among the lines are related to more recent papers (late 1980s) and were immediate follow-on papers (light green dots in early 1990s). These papers are closely related in time, and still benefit from the discovery of the new research strand. Alternatively, for follow-on papers of early 1980s (dark green dots), the citations for these papers in the 1990s are below the band of highly cited papers from the 1980s, which suggests that they have less impact than the papers they cited, except a few dark ones among the band. They received attention and many citations probably because there are research strand shifts in these papers, which was examined in detail in Schwartz, Fang, and Parthasarathy (2005).

The immediate follow-on papers (light green dots; citers of late 1980s) in the early 1990s are comparable with the highly cited papers in the 1980s. However, by looking at the distributions of light green dots in the figure, the impact that the late-1980s papers make can reach as far as about 10 years (light green dots start falling below the band in the late 1990s).

DOI: 10.1002/asi

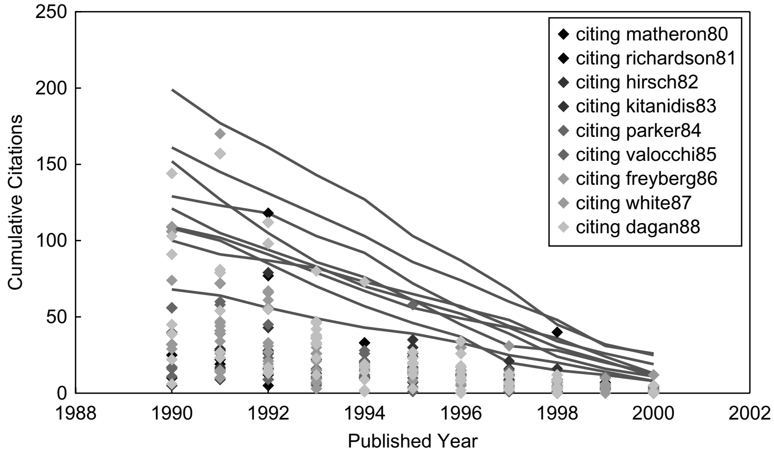

FIG. 6. Top 1% papers (red lines) from the 1980s versus papers citing them from the 1990s.

The objective of this study was to determine what features of a paper might contribute to high citations. This next analysis examines more specifically the idea of uniqueness. In other words, can one say that highly cited papers changed a paradigm and strongly influenced follow-on studies?

As outlined in the methods section, we compared 273 papers in 1991 with 2847 papers before and after 1991 using full-text similarity analysis to determine where in time (before and after 1991) the most similar papers are found. The hypothesis is that poorly cited papers would have topic-related papers before and after the time they are published and strongly cited papers would have few before and many after.

The results of this analysis are shown in Figure 7. The x-axis sorts citations as a function of percentile for every paper in 1991, so that data for highly cited papers (high percentiles) fall to the right and those for poorly cited papers fall to the left. The y-axis represents the number of related papers before and after 1991. There are two curves, showing numbers of papers before and after 1991.

Generally, the patterns indicate that there is a tendency for the highly cited papers to be somewhat unique in terms of what came before (pre-1991), and in terms of influencing the direction of work that followed (post-1991). For example, looking at the papers from 1991 with the highest citations (90–100 bin on the x-axis), we noted seven similar papers on average before 1991. After 1991 there were 14 similar papers on average, a much larger number. With these highly cited papers, the difference in the numbers of pre- and post-1991 papers is the highest. The more poorly cited papers of 1991 (e.g., 20th percentile) were associated with a larger number of papers on this same topic published both before and after. In other words, these 1991 papers appear not to be unique and are simply part of a collection of quite similar papers before and after. The lack of difference in the numbers of papers before and after suggests that these particular 1991 papers did not stimulate later work. The lowest bin of cited papers (0–10 percentile) seems to be special with few papers quite like them before or after. It is likely that these papers are like ducks-out-of-water—perhaps research that is not aligned to the common research strands normally found in *WRR*. Overall, it appears that only papers above the 50th percentile can be seen as successful in stimulating new research in the same topic area.

For every paper in 1991, there exists some number of similar papers that can be divided into a before (pre-1991) and an after (post-1991) group. For all the papers in these

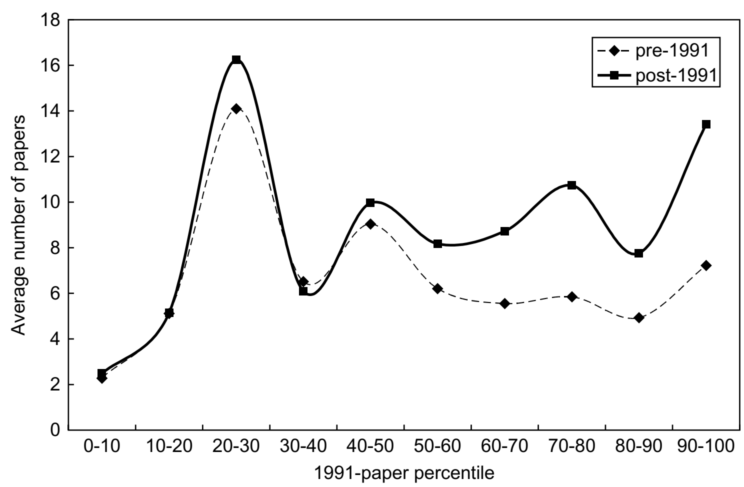

FIG. 7. Average number of papers correlated (at least 60%) with 1991 papers vs. percentile.

DOI: 10.1002/asi

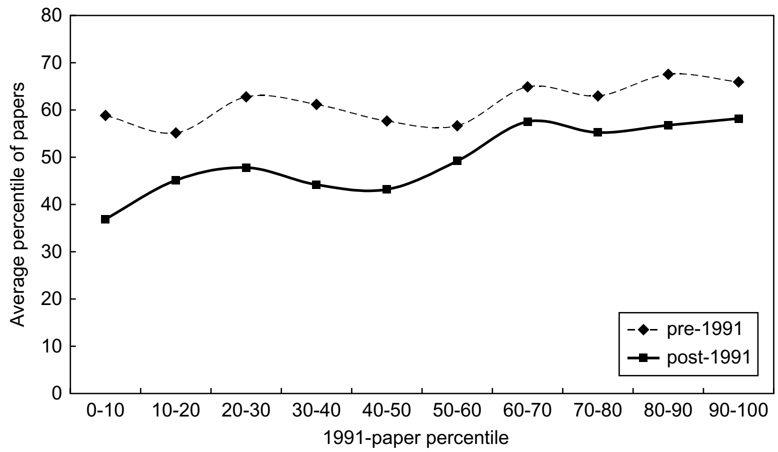

FIG. 8. Average percentile of papers correlated (at least 60%) with 1991 papers vs. percentile.

two groups, there exist citation data. Figure 8 shows the pattern of citations (normalized as average percentile) for the groups of papers as a function of the 1991 collection of binned papers. The results are quite remarkable. Essentially, earlier work (pre-1991) on the same topic is more highly cited than later work (post-1991). For the more poorly cited 1991 papers this difference is largest, up to the 20th percentile. For the more highly cited 1991 papers (80–100 percentiles on x-axis) the difference is small.

The results show clearly that there is a decline in citations for papers published on a similar topic (Figure 8). This decline is most evident for the lower-impact papers. However, the behavior is consistent across the range of cited papers.

Evaluation and Conclusion

An important concern in an analysis such as this one is that wider, Internet distribution of research papers may have impacted recent citation patterns. For the key journals of our study, and especially *Water Resources Research*, the transition to electronic distribution is only now happening and has not influenced the analyses presented here. Even now, the American Geophysical Union for economic reasons has chosen not to substantially broaden the distribution of *Water Resources Research* beyond its traditional subscriber base.

Our analysis provides insight on the pattern of innovation in hydrologic sciences. The growth patterns in citations of highly cited papers from two different time periods (Figure 4) show that these highly cited papers from the two decades have equivalent impact, and that citation data are stable and reliable after the 1980s as well. Authors publishing in the 1990s in *WRR* can expect their papers to attract citations comparable to successful papers of 1980s. It is not clear yet how the recent transition to a fully electronic version of *WRR* might change the patterns of citations because the papers are now more accessible than before and are just one click away from the readers through the internet. However, this similarity in citation behavior of the best papers from the two decades strongly suggests that the mechanism of citation counting is quite consistent over these two decades. Thus, problems of citation counting as described in the introduction that might have existed at the early stages (especially before 1980s) in citation bookkeeping are being dealt with, producing an increased precision of citation information.

The analysis built around Figures 5 and 6 strongly suggests that most highly cited papers in the 1990s are new research strands, with few papers representing immediate follow-on research papers of late 1980s. This is one line of evidence to suggest that the influence of research strands wane with time. Thus, age of a paper in relation to the age of the research strand appears to be important in determining how cited that paper might be. This idea was extended in Schwartz et al. (2005).

The study results (Figures 7 and 8) strongly suggest that innovation is another factor besides research topics influencing citations. The most highly cited papers appear to be unique in that there are relatively few papers like them that were published previously. Moreover, these papers were sufficiently influential to stimulate a relatively large number of similar topic-related papers. This result suggests that the hypothesis (described in Methodology) is correct, namely that the most highly cited work appears to inspire later work.

An in-depth look at citation patterns for *WRR* shows a broad distribution in citations among individual papers. At least 25% of all papers attract less than about five citations in their entire life time, while 5% attract more than 80. We think that concerns about the validity of citation analysis as a tool for analyzing research impact are overexaggerated since our study shows that the citation data are steady and stable for at least two decades.

Although citation analysis remains subject to criticisms of one kind or another (e.g., see Schwartz & Ibaraki, 2001), a variety of studies have identified its role in pinpointing influential research, patterns of connections among disciplines, and downturns in exhausted subfields (Redner, 2005). One of the most convincing early studies on impact was that of Garfield (1998) who showed that a list of the top 1000 most cited scientists in the world (1965–1978) included 250 members of the National Academy of Sciences and 125 members of foreign

DOI: 10.1002/asi

academies. In an extensive study of articles published in Physical Review, Redner (2005) finds many of the same trends that we observe in our analysis of the water-resources literature (Schwartz & Ibaraki, 2001). A few key foundational or discovery papers attract many citations over a long period of time while the greatest proportion of articles (70%) remain virtually uncited (10 cites or less). Our work here differs from more conventional citation analysis in that we are looking at time trends among families of papers related in time.

Citation analysis has an important role to play in understanding the direction of science and the nature of innovation. We think this article begins to lay foundation in the ground water field for development of a model of evolution of research strands.

References

Adam, D. (2002). Citation analysis: The counting house. Nature 415(6873), 726–729.

Fang, Y.C., Schwartz, F. W., & Parthasarathy, S. (2001). Data mining in geoscience research. In M.M. Noga (Ed.), Proceedings of the 36th Meeting of the Geoscience Information Society, Vol.32: Geoscience Information: A Dynamic Odessey (pp. 75 & 80). Boston, MA: GSIS.

Fang, Y.C., Parthasarathy, S., & Schwartz, F.W., (2001, November 29). Using clustering to boost text classification. Paper presented at the International Conference on Data Mining Workshop on Text Mining (TextDM'01). Retrieved from http://www-ai.ijs.si/DunjaMladenic/TextDM01/papers/Srini_TMIR.pdf

Garfield, E. (1992). Citation data: their use as quantitative indicators for science and technology evaluation and policy making, Science and Public Policy, 19(5), 321–327.

Garfield, E. (1998, February). Mapping the world of science. Paper presented at the 150th Anniversary Meeting of the AAAS, Philadelphia, PA.

ISI Web of Science. Retrieved from http://www.webofscience.com/

McCallum, A., (1996). Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering. Retrieved 2002 from http://www.cs.cmu.edu/~mccallum/bow/

Nature, (2002). (Editorial) Errors in citation statistics. Nature, 415(6868), 101.

Redner, S. (2005, June). Citation statistics from 110 years of Physical Review. Physics Today, 49–54.

Schwartz, F.W., & Ibaraki, M. (2001). Hydrogeological research: Beginning of the end or end of the beginning? Ground Water, 39(4), 492–498.

Schwartz, F.W., Fang, Y.C., & Ibaraki, M. (2002). Hydrogeological research redux: Response to critics. Ground Water, 40(3), 317–319.

Schwartz, F.W., Fang, Y.C., & Parthasarathy, S. (2005). Patterns of evolution of research strands in hydrologic sciences. Hydrogeology Journal, 13(1), 25–36.

DOI: 10.1002/asi