Long-term patterns in the aging of the scientific literature, 1900–2004[1]

Vincent Larivière*, Éric Archambault**, Yves Gingras***

*lariviere.vincent@uqam.ca
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succursale Centre-ville, Montréal, Québec, H3C 3P8 (Canada) and
Graduate School of Library and Information Studies, McGill University, Montréal, Québec (Canada)

**eric.archambault@science-metrix.com
Science-Metrix, 4572 avenue de Lorimier, Montréal, Québec H2H 2B5 (Canada) and
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal, Québec (Canada)

***gingras.yves@uqam.ca
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succursale Centre-ville, Montréal, Québec, H3C 3P8 (Canada)

Abstract
Despite a very large number of studies on the aging and obsolescence of scientific literature, no study has yet measured, over a very long time period, the changes in the rates at which scientific literature becomes obsolete. This paper aims at studying the evolution of the aging phenomenon and, in particular, how citation half-lives are changing over more than 100 years of scientific activity. It shows that the average and median age of cited literature has undergone several changes over the period. Specifically, the two World Wars had the effect of raising the average and median age of the cited literature significantly. Moreover, and contrary to a widely-held belief, the age of cited material has risen continuously since the mid-60s. Among the possible explanations for this counter-intuitive phenomenon, the most probable is the levelling off of the growth of scientific literature related to the steady-state dynamics of modern science that follows its exponential growth.

Keywords
Aging; half-life; obsolescence; Price index; diasynchronous analysis; scientific publications.

Introduction
The typical citation *life-cycle* of scientific papers starts with a fast increase during their initial years on the scientific scene, followed by a peak, and then a slow but steady fall into oblivion or are incorporated in the canon of normal science. This short and intense life and its subsequent aging process have always fascinated information scientists and bibliometricians, going as far as the seminal paper by Gross and Gross (1927). Since then, there has been a very large number of studies on aging and obsolescence (see the extensive reviews by Line and Sandison, 1974 and Line, 1993), most of them made using library loans and citation indexes and, more recently, with document download data (Nicholas et al., 2006). Despite the important body of literature on the topic, no study has yet measured on a global scale how the aging process of scientific literature has changed with time.

It has been suggested that, given the accelerated pace of scientific development, the scientific literature becomes more rapidly obsolete (Line 1970, 1993; Price, 1963, 1965). Knowledge is more rapidly disseminated with electronic means and, thus, one might expect that the useful life of scientific literature gets shorter. On the other hand, others (e.g., Odlyzko, 2002) have suggested that these electronic means and online bibliographic databases would have exactly the opposite effect—that is, authors would increasingly refer to older material. This paper examines these diverging hypotheses in

[1] The authors wish to thank François Vallières and Gilles Renaud for the construction of the bibliometric databases and Jean-Pierre Robitaille, Professor Jamshid Beheshti and the two anonymous reviewers for their valuable comments and suggestions.

order to determine whether the scientific literature is becoming more rapidly obsolete or if, on the contrary, it is increasingly being referred to for longer periods of time. In order to measure this phenomenon, data on the average and median of cited literature are compiled over the 1900 to 2004 period. Overall, this paper aims at providing a better understanding of the aging process of scientific literature and of the changes it has undergone over the last 100 years.

Methods

In their 1974 review paper, Line and Sandison have characterized three types of obsolescence studies: diachronous, synchronous, and diasynchronous. While diachronous studies follow the citation of specific documents through time, synchronous studies analyse the age distribution of cited documents at a given time. Finally, diasynchronous studies compare the age distribution of cited documents at different time periods, thus allowing for the measurement of changes in the aging process of literature (Line and Sandison, 1974). This paper belongs to the diasynchronous type of study, since it aims at analysing the evolution of yearly synchronous scores computed over the 1900–2004 period.

This paper uses data from Thomson Scientific, which is the only organization having indexed citations from scientific sources over more than 100 years. For each document indexed in Thomson's databases (source items), a list of references is included. Data between 1900 and 1944 are drawn from the Century of Science, which indexes 266 distinct journal titles covering most natural sciences and medical fields. From 1945 to 1979, data are from the Web of Science, Thomson Scientific's online bibliographic database. Finally, from 1980 to 2004, data are from the CD-ROM version of the Science Citation Index, Social Sciences Citation Index, and Arts and Humanities Citation Index.

In order to mitigate the effect of errors in the data, a 100-year and a 20-year citation window were used. The 100-year citation window proved to be the best equilibrium between minimizing errors in cited document years and maximizing the number of references. However, taking into account the potential effects that a 100-year citation window might have on the computation of some indicators (for instance, the average life is highly influenced by citations to older documents), we also used a 20-year citation window, which is the citation window used by Thomson Scientific in compiling their half-life measures. Finally, all statistics in this paper are based on the three standard types of documents: Articles (including notes), Reviews, and Meeting Abstracts.

Results

Figure 1 presents the evolution over a century of the number of scientific papers included in Thomson Scientific's databases together with the number of references made in these source papers. One can immediately see two salient features of this dataset: the publication of scientific research slowed considerably during each of the two world wars. The other important feature, although less salient, is the progressive slowing down in the growth of scientific production in this dataset after 1980. However, for this same dataset, the number of references has not been levelling off, although there is a slight decrease in the growth rate of the number of references starting around 1985. As predicted by Price (1963)—and as common sense would expect—the exponential growth could not last forever, and this Figure shows that the growth has indeed started to level off around 1980.

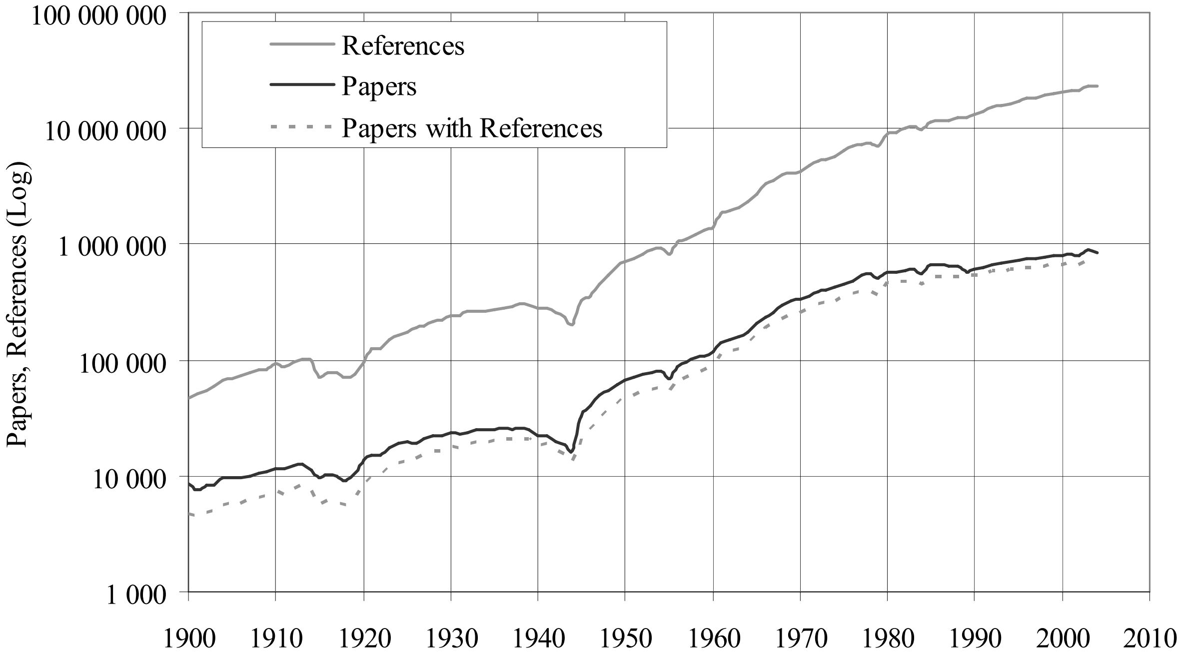

Figure 1. Number of papers, number of papers with at least one reference, and number of references, 1900-2004

Figure 2 presents the evolution for the 1900–2004 period of the average age and median age of cited literature—the latter being equivalent to the citing half-life—for citation windows of 20 and 100 years. One readily notices that there is an important difference between the average age and the median age—especially for the 100-year citation window. All curves show that the two world wars had a significant effect on the age of cited literature, increasing both the average and the median age of about 1.5 to 2 years. The cause of this effect is quite obvious: as the number of papers published decreased, for obvious reasons, during the two wars (Figure 1), researchers relied more on papers published before the wars, which, in turn, increased the age of cited literature.

The data contained in Figure 2 also show that, between 1945 and 1975, the average age of cited literature (100-year window) has been steadily decreasing, from an average of almost 12 years to 9 years. With the 20-year citation window, this decline in the age is also observed, albeit for a shorter period (1950–1960). This decrease is not observed, however, in the median age curves, which increased by about 2 years since the early fifties. All in all, the four curves show that the age of literature (average and median) has been increasing steadily since the mid-seventies[2].

[2] The huge upward jump observed in the mid-seventies for the average age calculated with a 100-year citation window is caused by the creation of the AHCI by Thomson Scientific.

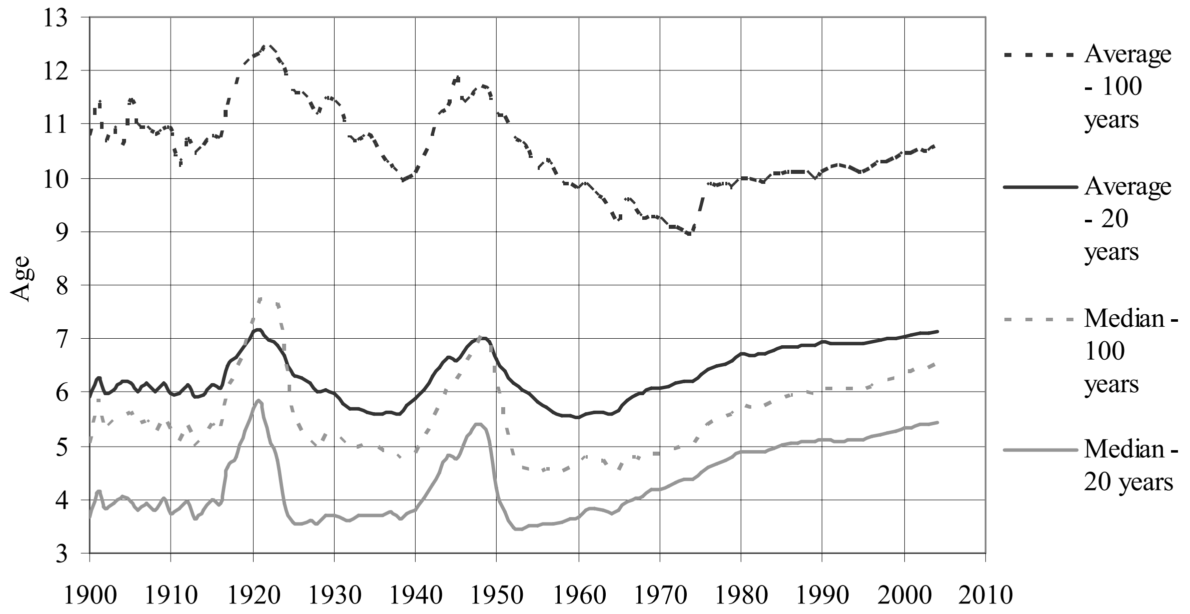

Figure 2. Average and median age of cited literature, 1900-2004

In order to see whether the phenomena observed is also valid for a fixed journal set, we compared the trend for two journals over the 1946–2005 period. As shown in Figure 3, the tendency towards a greater age of cited literature is also observed in *Science* and *Nature*[3]. For the sake of clarity, data for the 1900–1945 were removed because the small numbers of articles generated very large fluctuations. While the evolution of both curves is not identical to that of all ISI indexed journals presented in Figure 2, they are quite similar. Though the increase in the median age is less pronounced, one can observe that it has increased by about one year since the beginning of the fifties. The curves on the average age are also consistent with those presented in Figure 2.

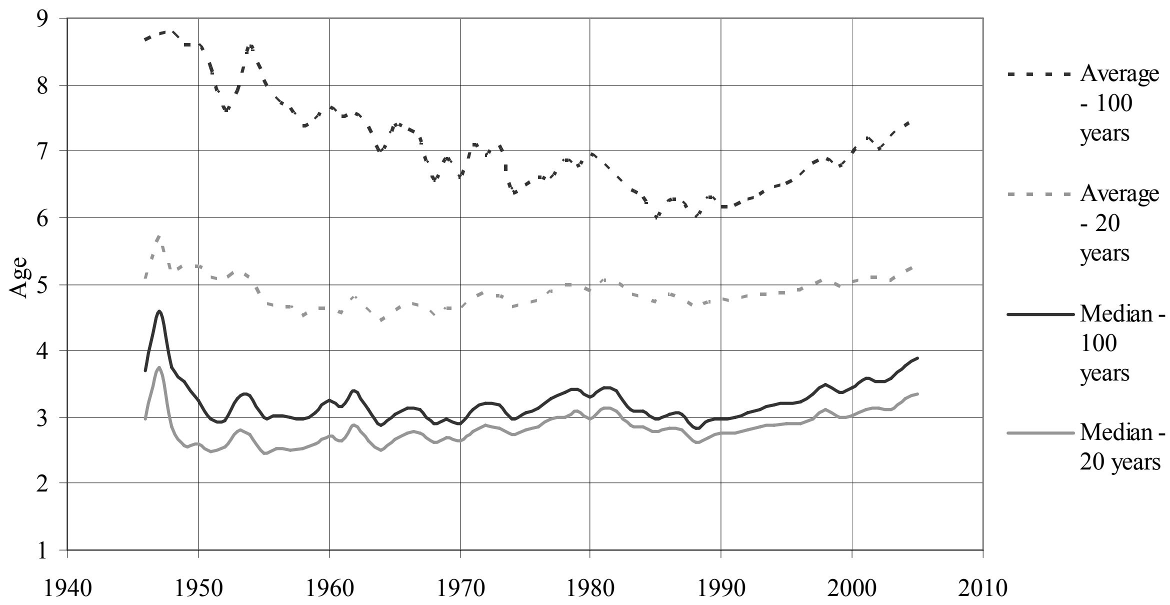

Figure 3. Average and median age of cited literature for *Science* and *Nature*, 1946-2005

[3] Note that the upward jump is absent from Figure 3, confirming the role of the AHCI in accounting for it in Figure 2.

Another measure of the aging of literature is the percentage of references, for a given year, to material that is five years old or younger. This measure—the Price Index—was developed by Price (1986) to distinguish fields having fast growth and an intense research front from less research-intense fields. Given the observed rise in the median age, we should expect to see a decline of the Price Index over time. Figure 4 shows that, since the mid-fifties, the share of references made to very recent literature has declined. Indeed, in addition to the huge effect of the two world wars, both the 20-year and the 100-year citation window curves show that the relative importance of recent literature among all cited material has been steadily decreasing, from 63% to 47% with a 20-year citation window, and from 53% to 41% with a 100-year citation window.

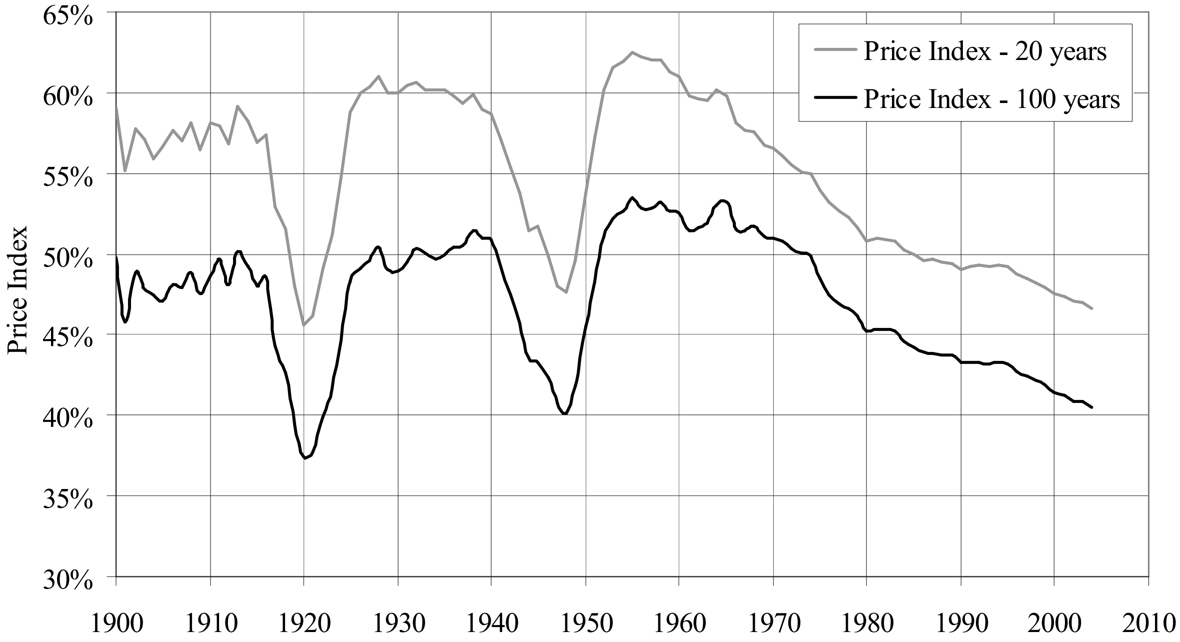

Figure 4. Price index, 1900-2004

Given the fact that the number of references per paper has increased tremendously since the mid-sixties, another way of measuring changes in the aging pattern of cited literature is to analyse its evolution by *citing unit* instead of amongst all cited material. Figure 5 presents the average number of references per paper, broken down for nine discrete age classes. One can see that the increase in the number of cited documents per paper is, by and large, caused by the absolute increase of *mid-aged* or *mature* literature—that is, papers having 3 to 5, 6 to 10, and 11 to 20 years of age. On the other hand, papers aged 1 year have seen their absolute importance decrease, while these aged 2 years stagnate after the end of the Second World War, and it only started increasing (at a much smaller rate than older papers) at the beginning of the eighties. Another point worth mentioning is the fact that documents aged 0–cited documents for which the publication and the cited year are the same–did not increase at all, and even decreased over the period studied. Taken together, these findings are contrary to what is to be expected, if one accepts the widespread idea that growing competition creates a push towards the use of younger literature and that the Web facilitates access to preprints and newly published research.

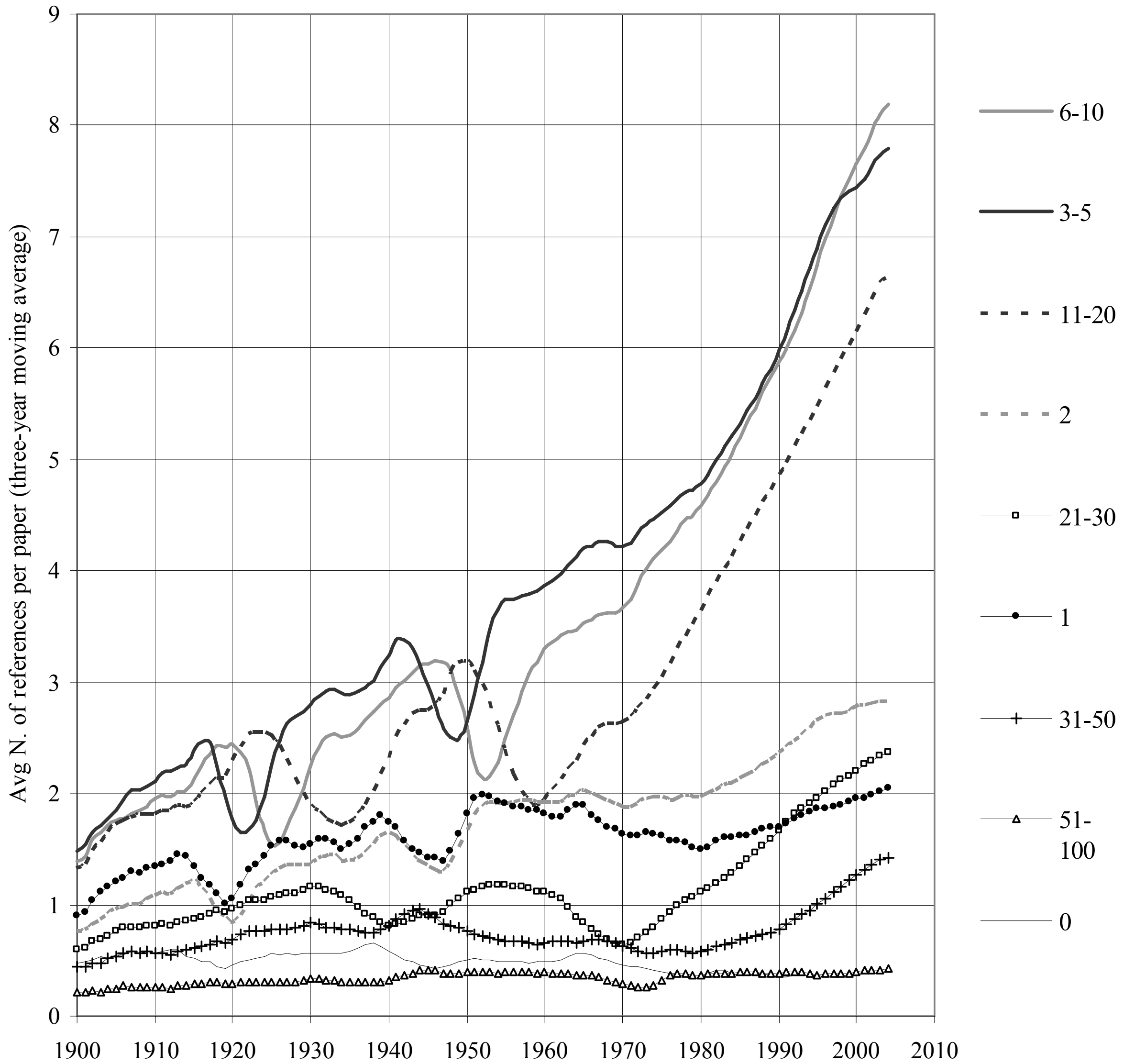

Figure 5. Average number of references per paper, by age of cited document, 1900-2004

Discussion and conclusion

The Figures presented thus far reveal a striking feature of the cited literature. Indeed, in contrast to a widely-held belief, scientific literature does not become obsolete faster nowadays and, actually, quite the opposite is observed. The useful life of scientific publications has been increasing steadily since the mid-seventies.

A first explanation for this rise in the age of cited literature can be inferred from the effects the two world wars had on the age of cited documents. As shown on Figures 1 and 2, because of the lower number of papers published during the wars (Figure 1), a significant increase in the average and median age of the documents cited by these papers can be observed (Figure 2). Given that a small decrease in the number of papers published has had a significant effect in the age of what is cited, a stabilization of the number of published papers will have a similar, albeit less pronounced, effect and increase the average and median age of cited documents.

These findings are consistent with Egghe's (1993: 199) models suggesting that, in synchronous analyses, "the higher the growth rate of the literature is, the faster it becomes obsolete." Conversely, if the rate of growth of the literature slows down, one could therefore expect a lengthening of the life of documents. Even though some countries—such as China—do currently have an exponential growth rate, in most countries, especially in North America and Europe—who account for most papers in the database—the growth has either turned to a fairly low exponential growth rate or even into linear mode, thus suggesting that these systems are in a steady state, if not slowing down (Ziman, 1994). All of this evidence points to the suggestion that the phenomena we are observing are internal characteristics of the scientific system. However, that is not to say that there are no factors at play that would be reflections of the changing behaviour of scientists. After all, there might be, underneath this trend, multiple sociological factors shaping the evolution of such a complex system.

In order to better understand the surprising growth of median age over the last quarter of the 20th century, we have distinguished, using the method developed by Larivière *et al* (2006), citations made to serials from those made to non-serials. As Figure 6 shows, the median age of cited non-serials in the natural sciences and engineering (NSE) and in medical fields (MF) is increasing more rapidly than for serials. However, for the social sciences and humanities (Figure 7), the median age of serials and non-serials is quite similar. This suggest that—at least for the NSE and the MF—part of the rise in the median age observed in Figure 2 is in fact due to the use of older non-serials. We cannot at this point suggest why non-serials show this intriguing pattern, and only further research could provide an explanation.

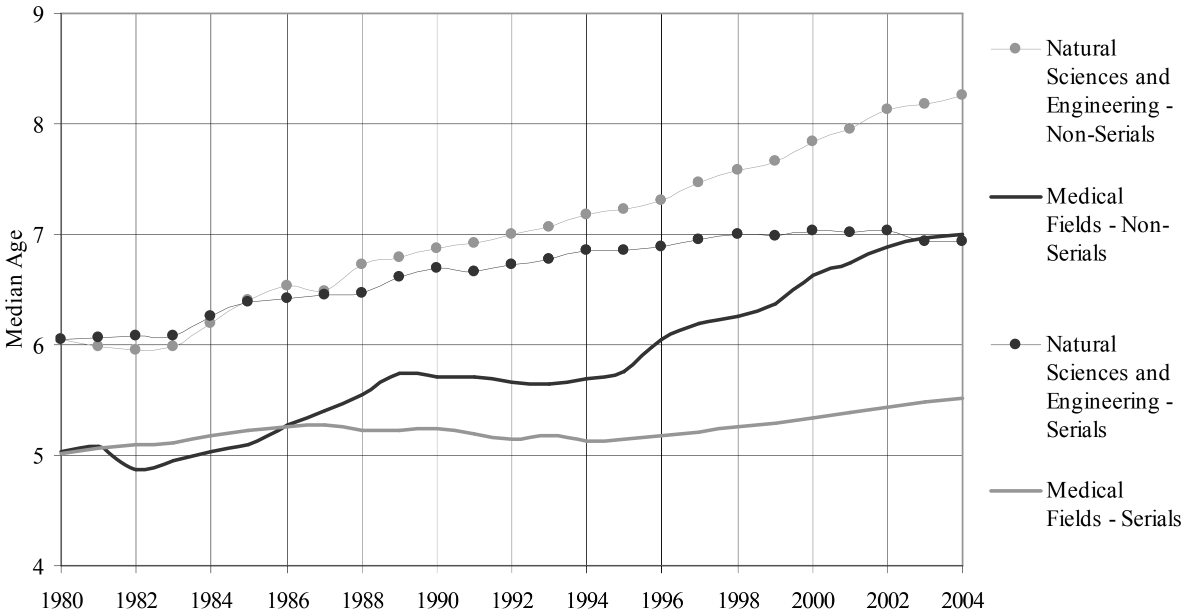

Figure 6. Median age of cited literature, by cited document, natural sciences and engineering and medical fields, 1980-2004 (100 years citation window)

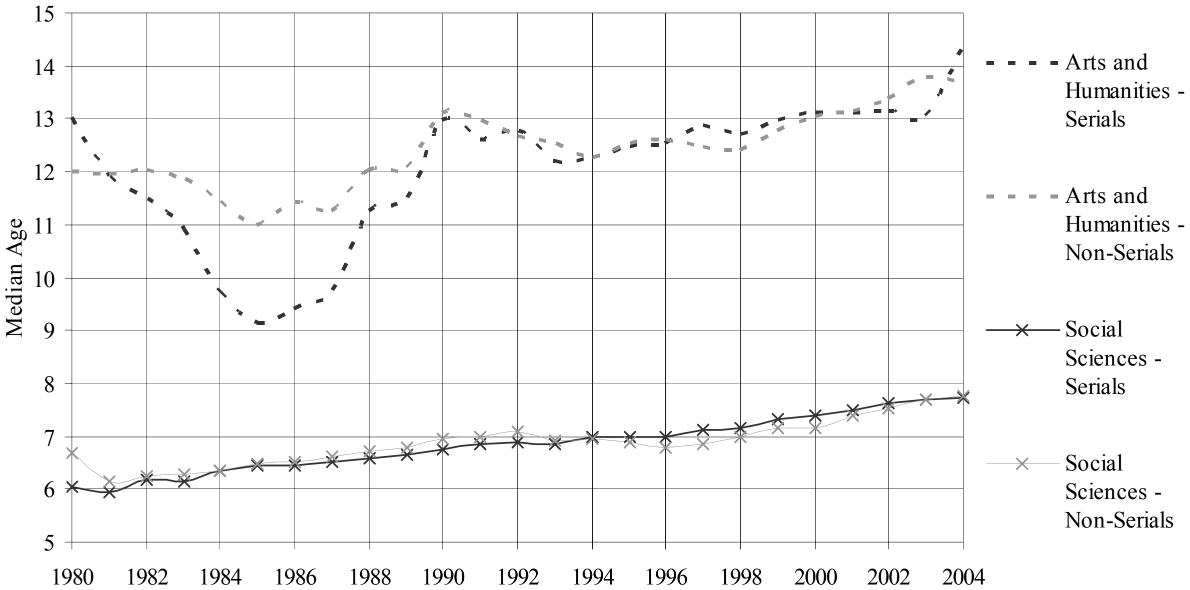

Figure 7. Median age of cited literature, by cited document, social sciences and humanities, 1980-2004 (100 years citation window)

Another explanation suggested by a reviewer is the effect of the increasing importance of review articles in Thomson's databases. Indeed, from the sixties to this day, the share of review articles increased from about 1% to 4%. Taking into account that review articles have a higher number of references and may cite older items, their growing importance might affect the median age of cited literature presented in Figure 2. Surprisingly, Figure 8 shows that, since the mid-forties, the median age of cited literature is higher for articles than for review articles or meeting abstracts. Consequently, the presence of review articles does not explain the trend observed, since the effect of an increase of reviews article would be to decrease the global median age of cited literature. This Figure also shows that, as one could expect, meeting abstract cite literature that is, on average, significantly younger than for articles or reviews.

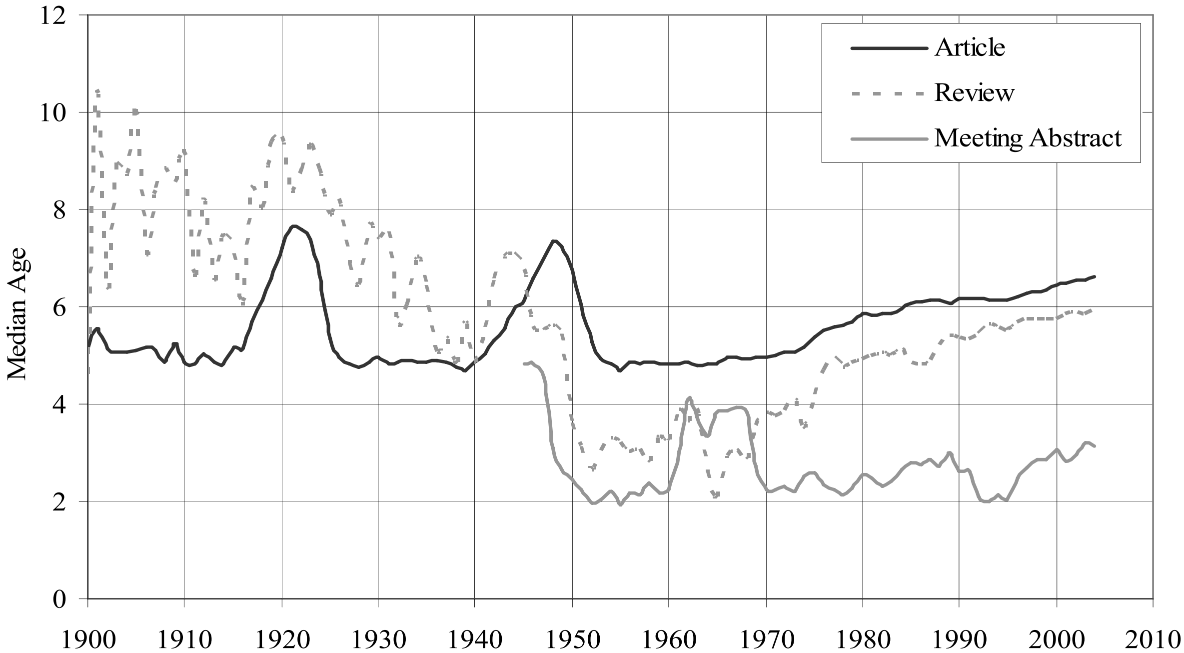

Figure 8. Median age of cited literature, by type of citing document, 1980-2004 (100 years citation window)

Other phenomena which could contribute to the longer life of scientific literature are the explosion of online bibliographical tools containing retrospective collections of serials. Such online tools certainly help researchers access increasingly old material, which they then could cite more frequently (see Boyce et al., 2004). Although this might have contributed to the variation in the aging process in the most recent years, Figure 5 clearly shows that citing older material more frequently started as early as 1960 (e.g., citing material that is 11 to 20 years old) and grew steadily afterwards. One cannot deny that the growing availability of computerized search tools since the mid-sixties (Neufeld and Cornog, 1986) and their subsequent wider availability contributed to this change. This presents indirect evidence that online databases of historical archives—such as JSTOR—should be highly encouraged and supported. Moreover, another hypothesis that can be explored, following Luwel and Moed (1997), is the impact that increased publication delay might have on the median and average age of cited literature.

As shown by the Price Index, science as a whole has been less and less intense at the research front. This is another argument pointing at a slower *renewal* of classics and of research fields. After the golden age of science (1946–1975), scientists had solved many of the important bottlenecks they faced, and no major "scientific revolutions" have appeared since. We would, thus, now be in a period of steady-state science (Ziman, 1994). Also, it may be that major contributions were more frequent before 1975 and that scientists increasingly cite material from that last innovative period. Although these explanations are worth further research, we suggest that the principal cause of the increased age of the cited scientific information is a fairly mechanistic response to the phenomenal growth in the quantity of published material after the war and to the current slowing in the growth of science.

References

Boyce, P., King, D.W., Montgomery, C. & Tenopir, C. (2004). How electronic journals are changing patterns of use, *The Serials Librarian*, 46, 121-141.

Egghe, L. (1993). On the influence of growth on obsolescence. *Scientometrics*, 27, 195-214.

Gross, P.L.K. & Gross, E.M. (1927). College libraries and chemical education. *Science*, 66, 385-389.

Larivière, V., Archambault, É., Gingras, Y. & Vignola-Gagné, É. (2006) The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities, *Journal of the American Society for Information Science and Technology*, 57, 997-1004.

Line, M.B. (1970) The half-life of periodical literature: apparent and real obsolescence. *Journal of documentation*, 26, 46-54.

Line, M.B. (1993) Changes in the use of literature with time: obsolescence revisited. *Library Trends*, 41, 665-683.

Line, M.B., Sandison, A. (1974) "Obsolescence" and changes in the use of literature with time, *Journal of Documentation*, 30, 283-350.

Luwel, M. & Moed, H.F. (1998) Publication delays in the science field and their relationship with the ageing of scientific literature. *Scientometrics,* 41, 29-40

Neufeld, M.L. & M. Cornog (1986). Database History: From Dinosaurs to Compact Discs. *Journal of the American society for information science*, 37, 183-190.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali M., H.R. & Polydoratou, P. (2005). Revisiting obsolescence and journal article decay. *Information Processing & Management*, 41, 1441-1461.

Odlyzko, A. (2002). The rapid evolution of scholarly communication. *Learned Publishing*, 15, 7-19.

Price D.J.D. (1963). *Little science, big science*. New York: Columbia University Press.

Price D.J.D. (1965). Networks of scientific papers. *Science*, 149, 510–515.

Price D.J.D. (1986). Citation measures of hard science, soft science, technology, and nonscience. In Nelson, C.E. & Pollack, D.K., (Eds.), *Communication among scientists and engineers* (pp. 155-179). New York: Columbia University Press.

Ziman, John M. (1994). *Prometeus bound: science in a dynamic steady state*. Cambridge: Cambridge University Press.