Long-Term Variations in the Aging of Scientific Literature: From Exponential Growth to Steady-State Science (1900–2004)

Vincent Larivière
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. E-mail: lariviere.vincent@uqam.ca

Éric Archambault
Science-Metrix, 1335A avenue du Mont-Royal E, Montréal, Québec, H2J 1Y6, 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. E-mail: eric.archambault@science-metrix.com

Yves Gingras
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. E-mail: gingras.yves@uqam.ca

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 article studies the evolution of the aging phenomenon and, in particular, how the age of cited literature has changed over more than 100 years of scientific activity. It shows that the average and median ages of cited literature have undergone several changes over the period. Specifically, both World War I and World War II had the effect of significantly increasing the age of the cited literature. The major finding of this article is that contrary to a widely held belief, the age of cited material has risen continuously since the mid-1960s. In other words, during that period, researchers were relying on an increasingly old body of literature. Our data suggest that this phenomenon is a direct response to the steady-state dynamics of modern science that followed its exponential growth; however, we also have observed that online preprint archives such as arXiv have had the opposite effect in some subfields.

Received May 8, 2007; revised August 1, 2007; accepted August 1, 2007

 • Published online 18 October 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20744

Introduction

The typical citation *life cycle* of scientific articles starts with a rapid increase during their initial years on the scientific scene, followed by a peak, and then a slow but steady fall into oblivion. Alternatively, they become incorporated in the canon of normal science (Merton, 1968). This short and intense life and the subsequent aging process have always fascinated information scientists and bibliometricians, as far back as the seminal article by Gross and Gross in 1927. Since then, many studies have been done on aging and obsolescence (for extensive reviews, see Line, 1993; Line & Sandison, 1974), most of them made using library loans and citation indexes—having in mind shelf-management issues—and more recently, with document download data (Nicholas et al., 2005). Aging studies also have been done using bibliometric methods, either for specific citing years or for small journal sets. For instance, Glänzel and Schoepflin (1999) studied the age distribution of references made in articles published in 1993 while van Raan (2000) used material cited in 1998. In addition, Moed, van Leeuwen, and Reedijk (1998) analyzed references made in 1995 to articles published between 1981 and 1995. Finally, Luwel and Moed (1998) and Glänzel and Schoepflin (1995) examined the aging pattern of articles published in small samples of journals. Despite this important body of literature on the topic, no study has yet measured how the aging process of scientific literature has changed over a very long period of time.

Many years ago, it was suggested that given the accelerated pace of scientific development, scientific literature becomes more rapidly obsolete (Line 1970, 1993; Price, 1963, 1965). Along those lines, the widespread use in some fields of open-access and e-print servers such as arXiv provides scientists with more rapid access to new research. Thus, one might expect the cited literature to be younger today than before the advent of these electronic means and that the useful life of scientific literature would shorten. On the other hand, others such as Odlyzko (2002) suggested that these electronic means and, especially, online bibliographic databases would have exactly the opposite effect—that is, authors would increasingly refer to older material because of their increased electronic indexation as well as online availability. This article examines these diverging hypotheses to determine whether scientific literature is becoming more rapidly obsolete or if, on the contrary, it is increasingly being referred to for longer periods of time. To measure this phenomenon, data on the median and average age of cited literature were compiled, covering more than a century of scientific activities (1900–2004). Though this article is mostly concerned with macro-level patterns, data for specific fields also are analyzed. Overall, this article provides a better understanding of the aging process of scientific literature and measures the changes of the last 100 years.

Methods

In their 1974 review article, Line and Sandison characterized three types of obsolescence studies: diachronous, synchronous, and diasynchronous. While diachronous studies follow the citation of specific documents through time, synchronous studies analyze 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 & Sandison, 1974). This article belongs to the diasynchronous type, as it analyzes the evolution of yearly synchronous scores computed over the 1900–2004 period.

This article uses data from Thomson Scientific, which is the only organization having indexed references made in scientific articles since 1900. For each document indexed in Thomson's databases (source items), a list of references is included. Data between 1900 and 1944 are drawn from Thomson Scientific's Century of Science, which indexes 266 distinct journal titles covering most natural sciences and medical fields. From 1945 to 1979, data are from the natural sciences, engineering, and medical journals covered in the *Web of Science*, and from 1980 to 2004, data are from the *Science Citation Index*. The data do not include articles in the arts and humanities or the social sciences. To each journal included in these databases, we assigned a field and subfield classification based on the scheme used in the Science and Engineering Indicators of the National Science Foundation. Note that the number of journals included in the study increases over the studied period, which is a reflection of the expanding scientific community. For instance, 77 distinct journal titles are covered from 1900, 269 from 1945, 3,590 from 1979, and 3,718 from 2004.

To mitigate the effect of errors in the data, 100-year and 20-year citation windows were used. The 100-year citation window proved to be the best compromise between minimizing errors in cited document years, such as counting references made to articles published in the Year "15" or "1397" for example, 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, one may consider that the average life is exceedingly influenced by citations to older documents—we also used a 20-year citation window. It is worth mentioning that there is no "true," "absolute," or even "best" citation window. Though a fixed citation window is needed to produce comparable measures from year to year, the choice of the length of that window for an obsolescence study is necessarily arbitrary. One even could argue that the best citation window would be to keep all old references—because it means something to cite articles that are 100 or 200 years old—but since there are some errors in the publication year of the cited documents, we preferred to keep cited references within the 100-year and 20-year windows. Moreover, the percentage (~1%) of cited articles older than 100 years is quite low over the period.

It is well documented that the distribution of cited references is highly skewed (i.e., that the vast majority of references are made to recent material) (e.g., van Raan, 2000). Though the median can be considered a more adequate measure of skewed distributions than can the arithmetic mean, we provide both indicators for comparison. Finally, all statistics in this article are based on references made by three types of documents that are widely considered as representing original contributions to knowledge: articles (including notes), review articles, and meeting abstracts.

Results

Figure 1 presents the evolution over a century of the annual number of published journal articles covered by Thomson Scientific's databases together with the number of references made in these source articles, for both medical fields (MED) and natural sciences and engineering (NSE). One can immediately see two salient features of this dataset: (a) The publication of scientific research slowed considerably during each world war, and (b) 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 leveling 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 Figure 1 shows that the growth indeed started to level off at the end of the 1970s.

DOI: 10.1002/asi

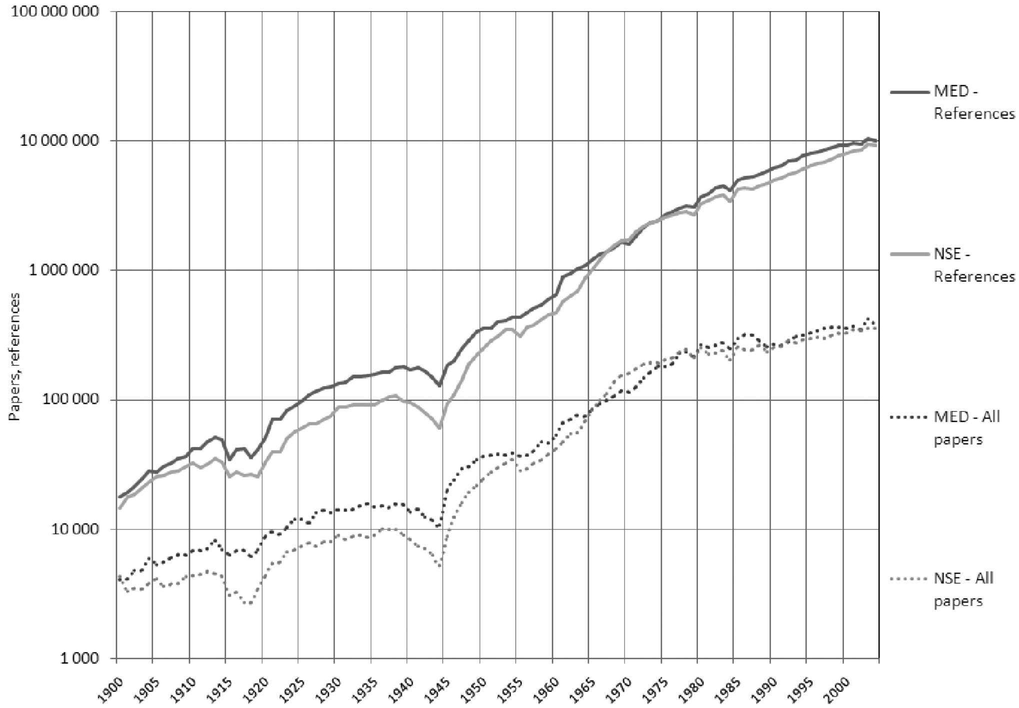

FIG. 1. Annual number of articles and references for medical fields (MED) and natural sciences and engineering (NSE), 1900–2004.

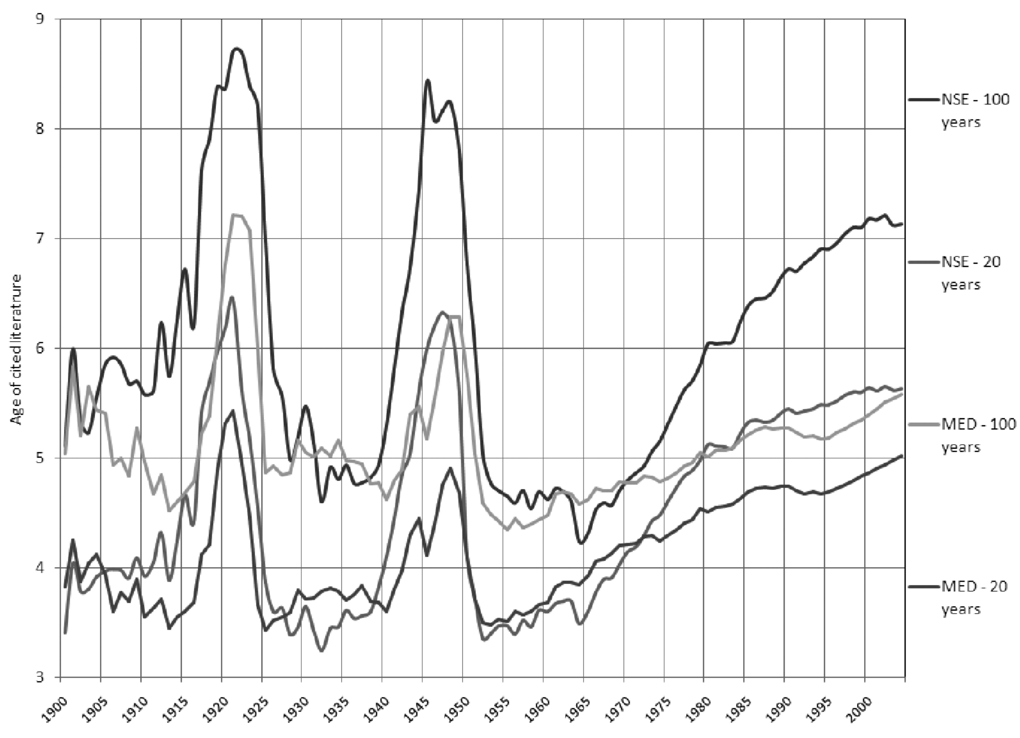

FIG. 2. Median age of cited literature, natural sciences and engineering (NSE), and medical fields (MED), 1900–2004.

Figure 2 presents the evolution of the median and Figure 3 that of the average of cited literature for MED and NSE, using 20- and 100-year citation windows for the 1900–2004 period. In both figures, one can readily notice the effect both world wars had on the age of cited literature—that is, an increase in age in both MED and NSE and for the two citation windows. The cause of this effect is quite obvious: As the number of published articles decreased during the two wars (see Figure 1), researchers relied more on articles published before the wars, which increased the age of cited literature.

DOI: 10.1002/asi

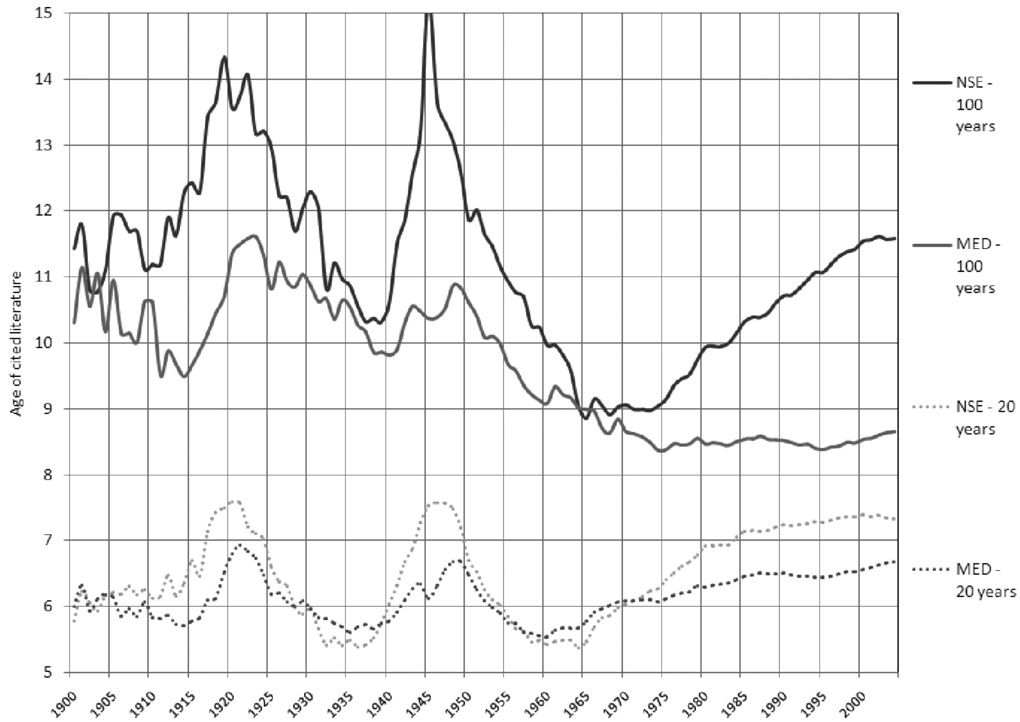

FIG. 3. Average age of cited literature, natural sciences and engineering (NSE), and medical fields (MED), 1900–2004.

Figure 2 shows that after being stable between ~1955 and 1970, the median age of cited literature has been increasing steadily since 1970. Note that between 1955 and 1970, the median age of cited literature for both NSE and MED was quite similar; however, starting in 1970, the age of references in the NSE literature started to increase significantly faster than that of MED: The median age of cited literature (which is often referred to as the citation's half-life) in NSE grew from about 4.5 in 1955 to more than 7 in 2004; for MED, it grew only from 4.5 to 5.5 (100-year window).

The data contained in Figure 3 show that between 1945 and 1975, the average age of cited literature decreased steadily in both MED and NSE (100-year window); however, while the average of NSE references increased steadily from 1975 on, that of MED references was more or less stable over the same period. With the 20-year citation window, the decline in age is also observed, albeit for a shorter period (~1945–1964). Overall, Figures 2 and 3 show that the age of literature (average and median) has been increasing steadily since the mid-1970s. In that respect, the age of references in NSE articles has been significantly higher than that of references in MED articles. Also note that the average measures are influenced more by extremes than is the median—by very recent articles or by very old articles—which is why they show a sharper decline in the age of cited documents between 1945 and 1970 than with median measures. As expected, the difference between the average and the median is quite high—especially for the 100-year citation window—which is an effect of the skewness of the time distribution of references.

Another useful measure of the aging of literature is the percentage of references, for a given year, to material that is 5 years old or younger. This measure, called 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 and average age, one should expect to see a decline of the Price Index over time. Figure 4 shows that since the mid-1950s, the share of references made to very recent literature has declined. Indeed, in addition to the huge effect of World Wars I and II, 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, though more importantly in NSE than in MED. That being said, for all citation windows and for both fields, less than half of the references were made to material published in the previous 5 years.

To see whether the phenomenon also can be observed at the level of a single subfield, we measured the median and average ages for astronomy and astrophysics. This subfield was chosen because (a) it is relatively small, (b) it is well covered in Thomson Scientific's databases over the 20th century, and (c) it is one of the first subfields to adopt a parallel system of knowledge dissemination (i.e., e-print servers such as arXiv). As shown in Figure 5, the trend toward a greater age of cited literature is also observed for that subfield, starting around the beginning of the 1970s. What is quite interesting, however, is that starting in 1991—the year arXiv was founded—the average and median ages started to decrease, which could be an indication that the use of e-print servers provides faster access to new research. Figure 6 presents the same data for astronomy and astrophysics and nuclear and particle physics—another subfield that relies

DOI: 10.1002/asi

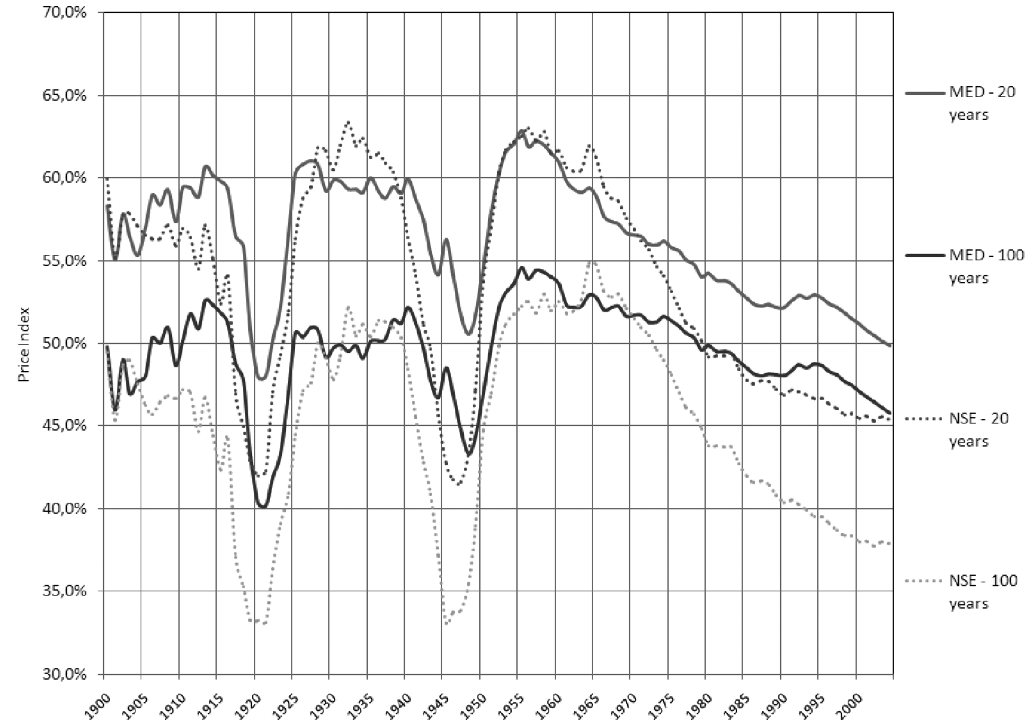

FIG. 4. Price Index for medical fields (MED) and natural sciences and engineering (NSE), 1900–2004.

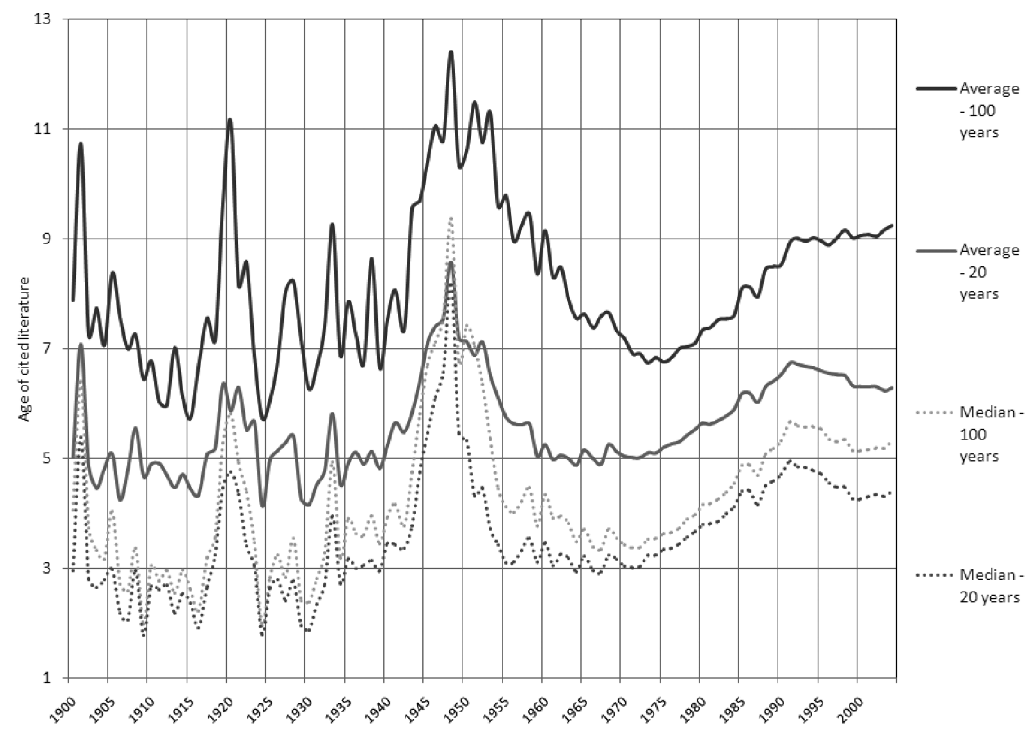

FIG. 5. Average and median age of cited literature in the field of astronomy and astrophysics, 1900–2004.

heavily on e-print archives—over the course of the last 25 years. It shows that the trend toward a younger age of cited literature also can be observed for nuclear and particle physics, starting in the mid-1990s. On the whole, these two figures are consistent with the macro-level data presented in Figures 2 and 3, with the exception that in these two subfields, we can observe the effect of e-print archives on the age of cited literature since the beginning of the 1990s.

Given the fact that the number of references per article has increased tremendously in both MED and NSE since the

DOI: 10.1002/asi

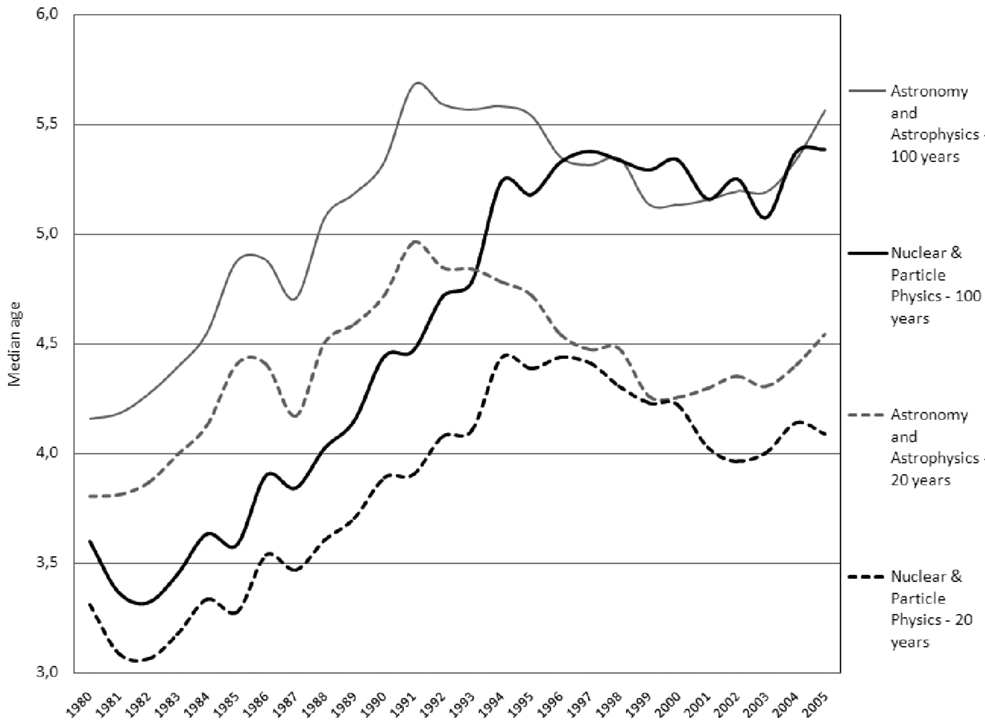

FIG. 6. Median age of cited literature in the fields of astronomy and astrophysics, and nuclear and particle physics, 1980–2005.

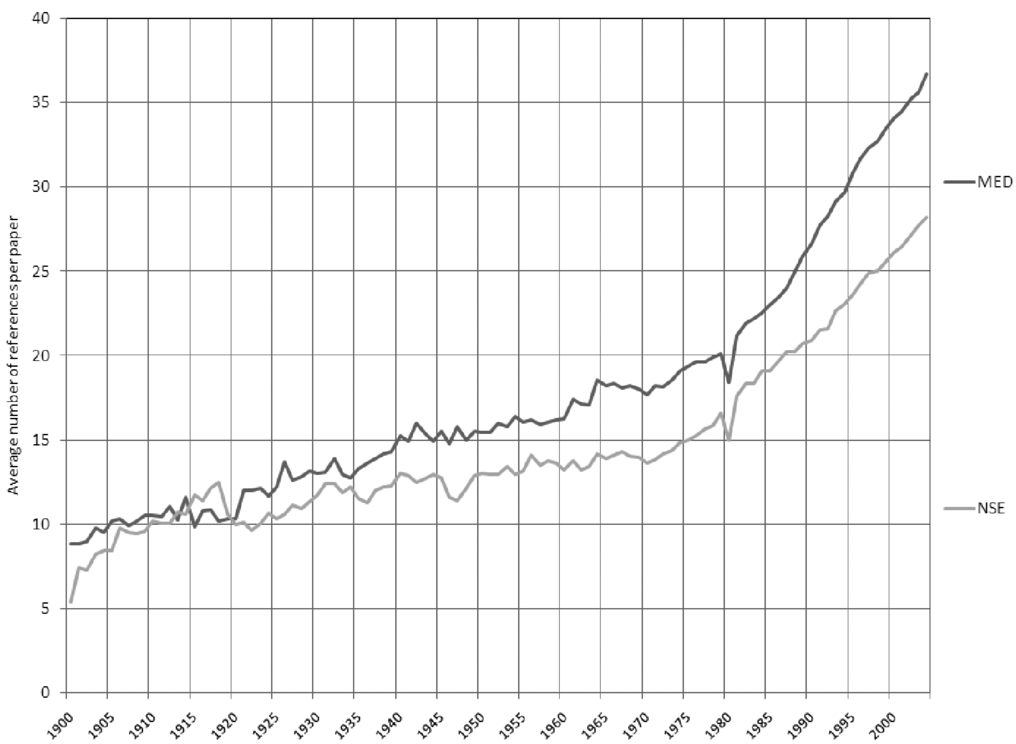

FIG. 7. Number of references per article, medical fields (MED), and natural sciences and engineering (NSE), 1900–2004.

mid-1960s (Figure 7), another way of measuring changes in the aging pattern of cited literature is to analyze its evolution by *citing unit* instead of among all cited material.

Figures 8 and 9 present the number of references per article, for eight classes of age of cited document, in both MED (Figure 8) and NSE (Figure 9). To minimize the effect of the increase in the number of references per article and thus enhance the clarity, the data are presented as the *share of total references*. One can see that in both fields, articles aged 1 year have seen their relative importance decrease

DOI: 10.1002/asi

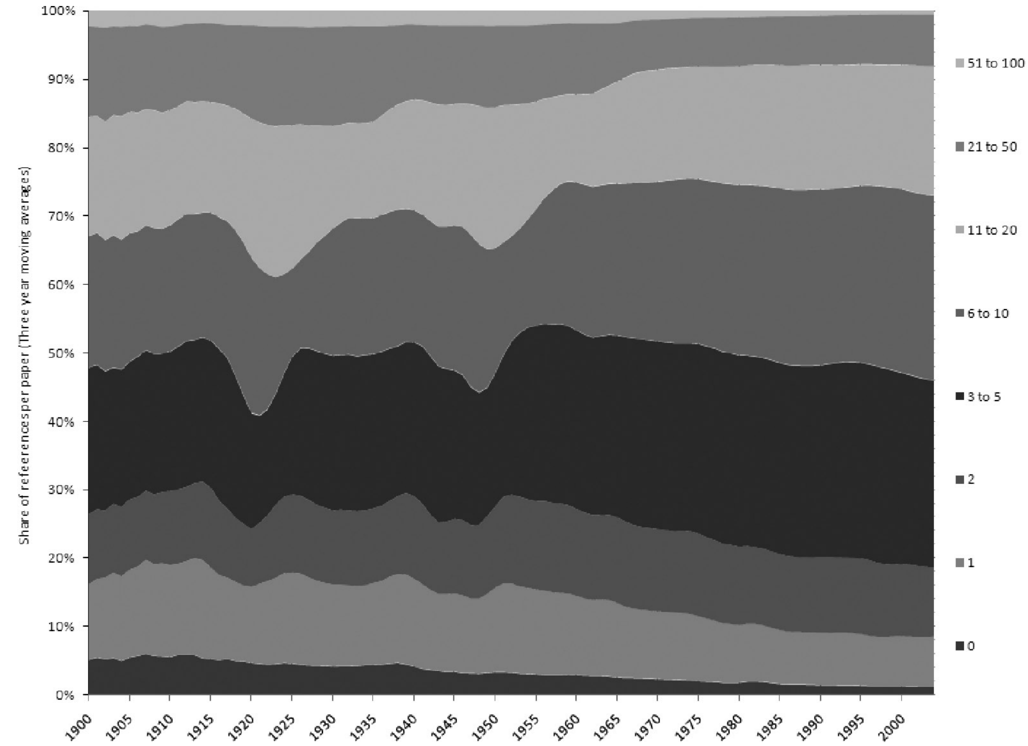

FIG. 8. Share of the average number of references per article, by class of age of cited document, for medical fields, 1900–2004.

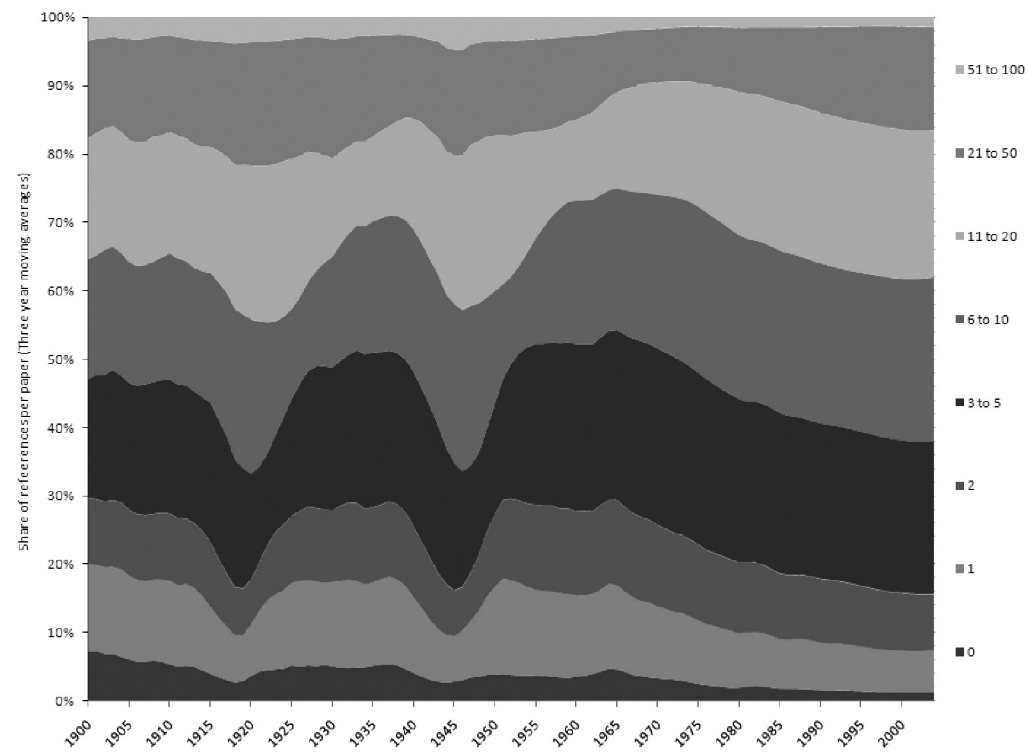

FIG. 9. Share of the average number of references per article, by class of age of cited document, for natural sciences and engineering, 1900–2004.

while those aged 2 years stagnated after the end of World War II and then decreased in NSE since the beginning of the 1980s. Also note that documents aged 0 (i.e., 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. On the other hand, the increase in the age of cited documents seems to be caused by the increasingly high importance, among the cited references, of *middle-aged* or *mature* literature (i.e., articles that are 6–10 and 11–20 years of age). In NSE, the importance of articles aged between 21 and 50 years has increased tremendously since the beginning of the 1970s, with their relative importance having more than doubled. These older articles are not excluded from the science system but rather still play an active role in it. The increase of these relatively old articles has had, in NSE, a high impact on the changes in the age of cited literature. Finally, very old articles (i.e., those aged 51–100 years) have seen their importance decrease in both fields since World War II.

Discussion

The figures presented in this article reveal a striking feature of the cited literature. Indeed, in contrast to a widely held belief, scientific literature is not becoming obsolete more rapidly than it used to; in fact, quite the opposite is observed. The useful life of scientific publications has been increasing steadily since the 1970s. A first explanation for this rise in the age of cited literature can be inferred from the effects both world wars had on the age of cited documents. As shown in Figures 1, 2, and 3, because of the smaller number of articles published during those wars, a significant increase in the average and median age of the documents cited by these articles can be observed. Given that a small decrease in the number of articles published has had a significant effect in the age of the cited material, a stabilization of the annual number of published articles will have a similar, albeit less pronounced, effect and increase the age of references.

Compiling the average age, for a given year, of the articles indexed in Thomson's database (Figure 10) shows that the average age grew linearly until 1945 and then diminished due a larger growth rate of recent literature until the mid-1970s. A third period (1980–2004) is characterized by a return to a linear growth of average age.

A comparison of these characteristics with the average age of cited literature (excluding cited literature published before 1900) suggests that the distribution of cited years roughly reflects the distribution of the average age of the cited literature, as if cited articles were randomly taken from the basket of existing articles with no strong bias toward recent articles. The rise and fall of the age of cited literature is thus a mechanical reflection of the rise and fall of the growth rate of the existing literature. The fact that the increase in age is much higher in NSE than it is in MED suggests that the latter disciplines tend to cite more recent science than is the case in NSE.

These findings are consistent with Egghe's model (1993), which suggests that in synchronous analyses, "the higher the growth rate of the literature is, the faster it becomes obsolete" (p. 199). Conversely, if the rate of growth of the literature slows down, one could expect a lengthening of the life of documents. Even though some countries such as China (Zhou & Leydesdorff, 2006) still have an exponential growth rate, most countries have turned either to a fairly low exponential growth rate or to stagnation, suggesting that their research systems are in a steady state (Ziman, 1994). On the whole, this would tend to suggest that scientific research, especially in the NSE, would be in a period of *normal science* (Kuhn, 1962); a period defined as "research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice" (p. 10).

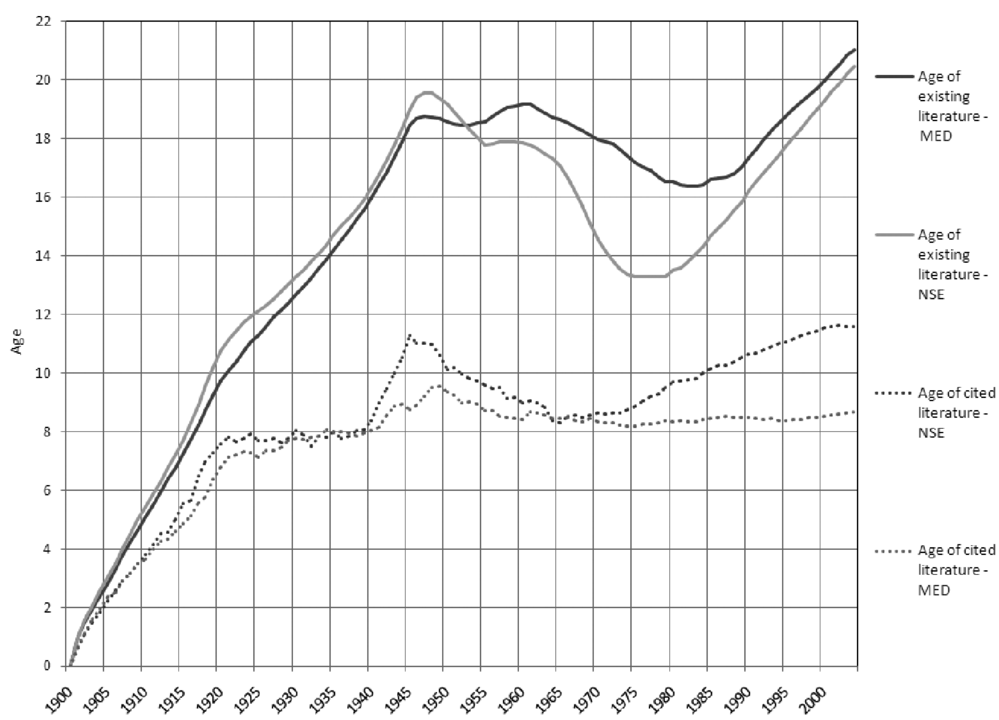

FIG. 10. Average age of the existing literature and average age of cited literature (1900+ citation window), 1900–2004.

DOI: 10.1002/asi

Other phenomena that 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 can cite more frequently (see Boyce, King, Montgomery, & Tenopir, 2004). Although this might have contributed to the variation in the aging process in more recent years, Figures 8 and 9 clearly show that citing older material more frequently started as early as 1960 (e.g., citing material that is 11–20 years old) and grew steadily afterward. The increased availability of computerized search tools since the mid-1960s (Neufeld & Cornog, 1986) and of older literature through systems such as JSTOR certainly contributed to this change.

Conclusion

As the evolution of the Price Index strikingly shows (Figure 4), contrary to a widely held belief, science as a whole—and especially NSE—relies on an increasingly older body of literature. After the golden age of science (i.e., 1945–1975), scientists had solved many of the important bottlenecks they faced, and no major "scientific revolutions" have appeared since. Science now seems to be in a period of steady state (Ziman, 1994). However, at the micro-level, data for the fields of astronomy and astrophysics and nuclear and particle physics showed that after an increase in the age of cited material analog to what is observed at the macro level, the age of cited literature started to decrease at the beginning of the 1990s, most likely as a consequence of the creation of e-print servers such as arXiv. Hence, added to the argument that open-access articles receive a higher number of citations (Harnad et al., 2004), our data show that foundation of an e-print archive is good for the field as a whole since it provides faster access to new research, therefore increasing the speed of use of research results. One could thus expect a new trend toward a lower age of cited literature with the general adoption of e-print servers; however, given that very few fields have adopted e-prints as much as have researchers in astronomy and astrophysics and nuclear and particle physics, the global effect of bibliographical databases and the global steady state of science are currently more important than those of online preprint servers.

Acknowledgments

We thank François Vallières and Gilles Renaud for the construction of the bibliometric databases. We also acknowledge Jean-Pierre Robitaille, Professors Jamshid Beheshti and Stevan Harnad, and the three anonymous referees for their valuable comments and suggestions.

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.

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature. Journal of Information Science, 21, 37–54.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. Information Processing and Management, 35, 31–44.

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

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H., & Hilf, E.R. (2004). The access/impact problem and the green and gold roads to open access. Serial Reviews, 30, 310–314.

Kuhn, T.S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

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.

Merton, R.K. (1968). Social theory and social structure. New York: The Free Press.

Moed, H.F., van Leeuwen, Th.N., & Reedijk, J. (1998). A new classification system to describe the ageing of scientific journals and their impact factors. Journal of Documentation, 54, 387–419.

Neufeld, M.L., & Cornog, M. (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 C.E. Nelson & D.K. Pollack (Eds.), Communication among scientists and engineers (pp. 155–179). New York: Columbia University Press.

van Raan, A.F.J. (2000). On growth, ageing, and fractal differentiation of science. Scientometrics, 47, 347–362.

Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. Research Policy, 35, 83–104.

Ziman, J.M. (1994). Prometeus bound: Science in a dynamic steady state. Cambridge, England: Cambridge University Press.