

# Revisiting citation aging: a model for citation distribution and life-cycle prediction

Hamid Bouabid

Received: 27 November 2010 / Published online: 30 March 2011  
© Akadémiai Kiadó, Budapest, Hungary 2011

**Abstract** The study of citation distribution provides retrospective and prospective picture of the evolving impact of a corpus of publications on knowledge community. All distribution models agree on the rise of the number of citations in the first years following the publication to reach a peak and then tend to be less cited when time passes. However, questions such as how long it will continue being cited and what is objectively the rate of the decline remain unanswered. Built up of simple polynomial function, the proposed model is proven to be suitable to represent the observed citation distribution over time and to interestingly identify with accuracy when the major loss of citations happens. I calculate from the model the ‘*residual citations*’ representing the citations kept after a long time period after publication year. I demonstrate that the residual citations may be greater than or equal to zero, meaning that the ‘life-cycle’ of the corpus is infinite, contrary to what some researches termed to be around 21 years. This model fits the observed data from *SCI* according to R-sq which is greater than 98.9%. Rather, it is very simple and easy to implement and can be used by not highly-skilled scientometric users. Finally, the model serves as a citation predictive tool for a corpus by determining the citations that would obtain at any time of its life-cycle.

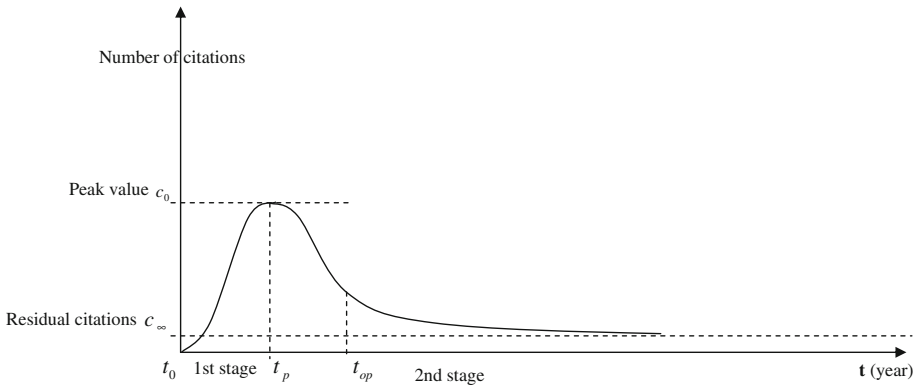
**Keywords** Citation distribution · Observed citations · Citation aging · Model · Life-cycle · Life-time · ISI-data · OECD-countries · Citation-prediction

## Introduction

The study of citation distribution allows both historical and prospective analysis of the impact of a corpus of publications over time. In general, it is agreed that an influential paper is likely to earn a number of citations in its first following years (second to sixth years) after publication, called “immediacy effect”. Then, rise to a citations-peak, after what it tends to be gradually less cited with time. A typical curve of the observed citation

---

H. Bouabid (✉)  
University Ibn Tofail, B. P 133, 14000 Kenitra, Morocco  
e-mail: h.bouabid@hotmail.com



**Fig. 1** Schematic representation of the two stages of the distribution of citations

distribution over time is shown in Fig. 1. There have been two approaches to study the citation distribution. The first considers papers cited by a publication during a particular year and then analyzes retrospectively the distribution of their ages. This approach is called ‘synchronous distribution’ (Nakamoto 1988), ‘citations from’ approach (Redner 2004) and ‘retrospective citation’ approach (Burrell 2002; Glänzel 2004). The second approach consists of analysing the distribution of citations gained over time by a paper (or papers) published in a given year. This is called ‘diachronous distribution’ (Nakamoto 1988), ‘citations to’ approach (Redner 2004) and ‘prospective citation’ approach (Burrell 2002; Glänzel 2004). Nakamoto suggested that the synchronous and diachronous distributions follow similar curves and are symmetric. The two approaches were compared by Stinson and Lancaster (1987) in measuring the obsolescence.

Avramescu (1979) has proved the reliability of his model for citation distribution based on a combination of the diffusion theory (in physics) to represent the decay and the exponential law to represent the growth of literature. However, he confessed that even if the model fits the curve, it cannot explain neither the decay rate nor the immediacy effect. He added that when the model fits better in small times it is not for larger times and vice versa. Egghe and Ravichandra Rao (1992) have advocated a lognormal distribution rather than Avramescu’s model which they believe is ‘not the right citation age model’. In comparing various models, Pollmann (2000) has concluded that an inverse function and an exponential function are the best candidates to represent the citation distribution over time. However, he claims that from age ‘4’ (i.e. starting from the 4th year after publication) the decay curve of the aging of literature is better predicted by an inverse model than by an exponential model. In 2003, Burrell presented and compared two functions: exponential and gamma distributions; even though he presented these functions as a purely theoretical exercise in stochastic modelling without using real data to support or examine the fit of these functions. Some other researches has been carried out by Redner (2004) when examining statistically the complete citations distributions (citations to) of articles published in Physical Review from more than a century, which has shown that the distribution is described by a power law age distribution with an exponent close to  $-1$  over a time period of 2–20 years. Later, Nadarajah and Kotz (2007) have examined a collection of 17 distribution functions: uniform, exponential, gamma, beta (two kinds), Rayleigh, Stacy, Pareto (two kinds), Inverse Gaussian (two kinds), half normal, half logistic, half Cauchy, half  $t$ , Fréchet, and Pearson type VI. Their aim was not far than providing a useful

reference on citation modelling. They actually plotted only the beta, exponential and normal distributions without giving information about the data used or comparing the models. Simkin and Roychowdhury (2007) have found average rate of citation diminishing with a paper's age as a power law with an exponential cut-off. They yet calculated the number of years for the transition from a hyperbolic to an exponential distribution. These findings are in accordance with the previous statements of Redner (2004). However, Simkin and Roychowdhury (2007) refute the empirical finding of Nakamoto (1988) concerning the exponential decay of citations with age. Mingers and Burrell (2006) have adopted quite different method by determining the future number of additional citations given the number of citations earned at a time  $t$  including obsolescence functions. It is also the case for Burrell (2002) when addressing the citation age using the  $n$ -th citation distribution that is the distribution of sets of articles receiving their  $n$ -th citation at a given time after publication. This method is to some extent similar to that of the cumulative citation distribution function (dimensionless) recently improved by Yu and Li (2010) but originally used to overcome the fitting error of classical citation distributions particularly at the beginning of referencing time (or citing time).

The model proposed in this article is based on a diachronous analysis which is the natural one and also being proved by Glänzel (2004) to be, from an empirical point of view, the appropriate method (than the synchronous) to characterise citation processes (even if he has used the relative frequency citation). This model is proposed with a special focus to bring responses to questions of how quickly a corpus of publications is losing citations and getting lesser impact over knowledge community?, how long a publication will continue being cited?, what is objectively the rate of decline of the curve? and when precisely the major loss of citations occurs? Two corpuses can earn the same number of citations but over a completely different period, and also experience a fall in the rate of citation over a completely different period of time even if they follow broadly the same rules for publishing and respond extensively in the same way to science and technology advances (Bouabid and Martin 2009; Osareh and Wilson 1997).

Rather, the model proposed is interestingly used to predict future or expected citations for a corpus of publications. Indeed, at any time after publication we could predict the number of citations a corpus would receive and consequently, predict the 'life-cycle' or 'life-time' (interpreted as the period when the corpus is still cited) which I found to be either a finite period or infinite one.

The model provides a reasonable fit to the observations that are made using the *ISI*-data with a R-sq score always greater than 98.9%. I have also kept in mind that any proposed model should not only be reasonably accurate, but also easy to use and implement, particularly for less highly-skilled users of scientometric.

## Analysis and distribution model

### Assumptions

- I. A macroscopic analysis is made for the corpus of publications under observation (macro-level) rather than an individual citation analysis inside the corpus (micro-level). The observed curve is reported in Fig. 1 and is assumed to reflect observed data (macro-level) which may include so far, but would not distinctively analyse, phenomena such as factual growth of literature (Avramescu 1979; Stinson and Lancaster 1987), event-data problems related to citation (MacRoberts and

- MacRoberts 1989), delay effect (Egghe and Rousseau 2000a; Yu et al. 2006), sleeping beauty or other endogenous effects (van Raan 2004; Simkin and Roychowdhury 2007) or uncited papers or highly cited papers (Wallance et al. 2009). All studies agree that in general macro-level analysis encounter less distortions and irregularities that may raise in micro-level;
- II. Time  $t$  is assumed to be continuous variable and the citation function is also continuous for all  $t \in [0, \alpha[$ . In reality, time is continuous but is simply discrete and limited to ‘whole year’ unit only for publication requirement;
  - III. Stages one and two in Fig. 1 are not systematically linked. Each one is governed by rather different social behaviours of citing communities. Scientists might react more impressively when responding to recent advances which results in a high increase in citing. In addition, the sense of ‘immediacy’ of a paper is greater when it cites very recent publications. Such a phenomenon would result in a lower rate of citations to older publications. Moreover, immediate interest in a newly released scientific work is not the same as that after a few years. The distinction of the two stages is supported by Egghe and Rousseau (2000b) who came up to the conclusion that the two approaches considering the citation distribution as one phenomenon or as composed of two stages (grow, decay) have their value; it only depends on the topics one wants to study.

### Mathematical model for the 2nd stage

In Fig. 1, we note that the number of citations  $c(t)$  falls from a peak value of citations (or maximum)  $c_0$  at a time  $t_p$  (peak time) to a lower value  $c_\infty$  which I called ‘*residual citations*’. As we will see later the residual value may vanish to zero meaning that the corpus of publications ceases to earn any citations in long term.

Thus:

$$\text{for } t = t_p \text{ we have } c(t_p) = c_0. \quad (1)$$

The second stage starts from  $t = t_p$  and on.

$t_p$  is simply obtained by an iterative process : for any  $t_i \geq t_0$  we calculate  $c(t_{i+1}) - c(t_i)$ . If  $c(t_{i+1}) - c(t_i) < 0$  and  $c(t_{i+2}) - c(t_{i+1}) < 0$  then  $t_i = t_p$ . In other words, we compare to a given value of citation the two successive ones starting from  $t_0$ ; and for two subsequent times  $t_{i+1}$  and  $t_{i+2}$  if  $c(t)$  is successively decreasing ( $c(t_i) > c(t_{i+1}) > c(t_{i+2})$ ) it is then considered as the peak value and the corresponding time as being the peak time.

To focus the study on the critical and the larger time period beyond the peak (2nd stage), I introduce a change in the time variable as follows:  $t = t - t_p$ . Therefore, one obtains:

$$t = 0 \Rightarrow c(0) = c_0. \quad (2)$$

The curve from Fig. 1 becomes as shown in Fig. 2 where  $t \in [0, \infty[$ .

Starting from the findings of Redner (2004), Pollmann (2000) or Simkin and Roychowdhury (2007), I first suggest a power law decline in the rate of citation that is given by  $c(t) = c_0 - \alpha t^p$ , where  $\alpha$  is a constant and  $p$  is the power to reflect the rate of decay and is of sure greater than 1. Nevertheless, to satisfy change of curve from concave to convex form (see Fig. 2); and also when  $t$  is greater (or  $t \rightarrow \infty$ ),  $c(t)$  is approaching an asymptote (residual citations), I suggest that the citation distribution (overall observed citations)  $c(t)$  at a time  $t$  has the same power law both in the numerator and the denominator as:

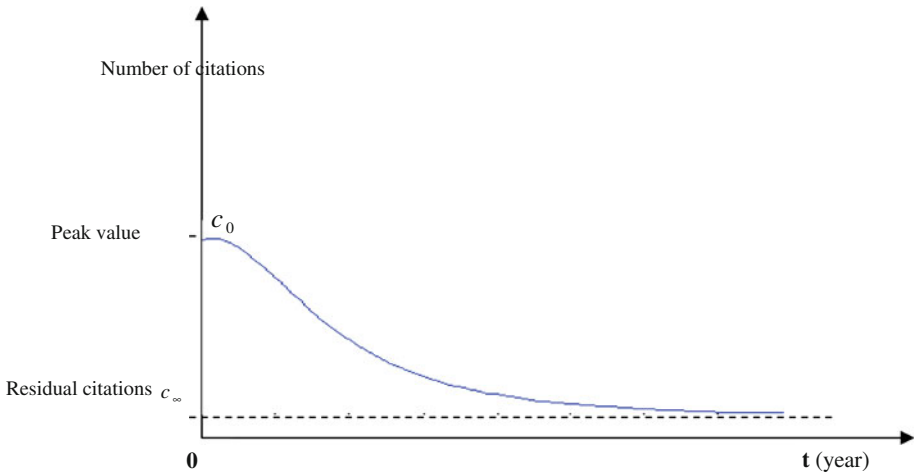


Fig. 2 Representation of the second stage of the distribution of citations with time

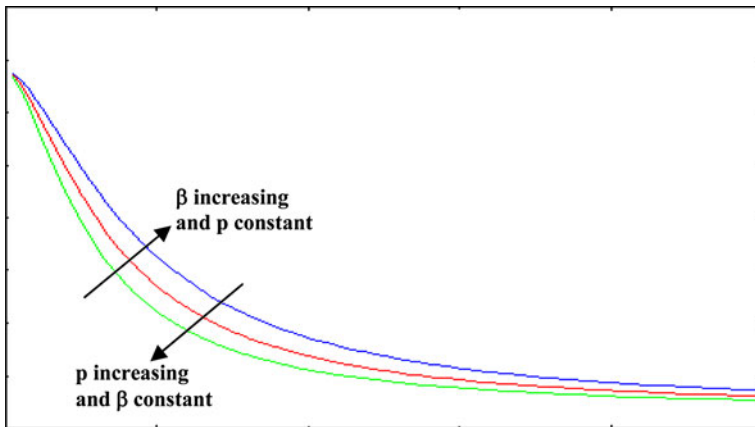


Fig. 3 Various distributions of citations with different values of  $p$  and  $\beta$

$$c(t) = c_0 - \alpha \frac{t^p}{t^p + \beta} \tag{3}$$

with  $\alpha$  positive constant and  $\beta$  introduced to avoid singularity (an indeterminacy) if  $t$  equal to zero.  $\beta$ -factor also plays a key role in accelerating or slowing down the distribution decay since denominator is made larger than the numerator (see Fig. 3). Finally, we ensure from Eq. 3 that if  $t = 0 \implies c(0) = c_0$ .

The model above in its polynomial form is very simple than other complicated mathematical formula of citation distributions. Consequently, extracting and deducing significant indicators will be meaningful in citation analysis.

Distributions from Eq. 3 are plotted in Fig. 3 with different  $p$  and  $\beta$  factors.

Figure 3 reveals when  $p$  increases, the distribution tends to rapidly loose many citations with time. Distribution with smaller value of  $p$  is likely to maintain citations much longer.

In Eq. 3 when  $t$  is very large ( $t \rightarrow \infty$ ), we obtain  $c(\infty) = c_\infty = c_0 - \alpha$  representing residual citations over a long period of time. Mathematically speaking, we face three cases:

1.  $c_0 > \alpha$ : this means that some publications succeed in maintaining their scientific value despite time passing ( $c_\infty > 0$ ). The ‘life-time’ or ‘life-cycle’ is infinite. In this case, we rewrite the model as  $c(t) = c_0 \left(1 - \gamma \frac{t^p}{t^p + \beta}\right)$ , where  $\gamma = \frac{\alpha}{c_0}$  represents the percentage of lost citations compared to citations at peak. Thus the quantity  $(1 - \gamma)$  denotes the percentage of remaining citations after a long period of time compared to those at peak time. This ratio provides, as shown later, an objective measure of the real rate of decline in the number of citations for the corpus under observation.
2.  $c_0 = \alpha$ : means that  $\gamma = 1$  and the  $t$ -axis represents the asymptote of the curve. The residual citations  $c_\infty$  is equal to zero but the distribution does not completely vanish. The corpus loses 100% of its maximum citations earned at its peak. We nevertheless assume in this case that the life-cycle is infinite as in case one.
3.  $c_0 < \alpha$ : ( $c_\infty$  is negative) means the total number of citations vanishes to zero within a given time  $t_v$  calculated by setting  $c(t_v) = 0$  since the number of citations  $c(t)$  is always greater or equal to zero (and is never negative). In other words, all the publications are being forgotten by scientific community within a time period  $t_v$ . Using the model in Eq. 3 and the previous equations, we deduce:

$$t_v = \sqrt[p]{\frac{\beta}{\gamma - 1}} \text{ with } \gamma > 1. \tag{4}$$

### Indicators for citation aging and life-cycle

Using the model, I suggest in this part the *speed of withdrawal* index as the maximum rate of decline (loss is higher) in citations earned before the corpus of publications is getting forgotten. The smaller the index the greater its impact on the science community. The second is the *Optimum-time* index, as time point when major loss of citations happens during ‘life-cycle’ of these citations.

The existence of a specific time point  $t^* > t_p$  when the function (obsolescence) rate has a minimum has been proved by Egghe and Ravichandra Rao (1992). Using the lognormal, they suggest however a finite  $t^* > t_{op} > t_p$  (see Fig. 1) in one case or  $t^* = \infty$  in the second case.

According to Assumption II, let us take the derivative of the citations  $c(t)$  as:

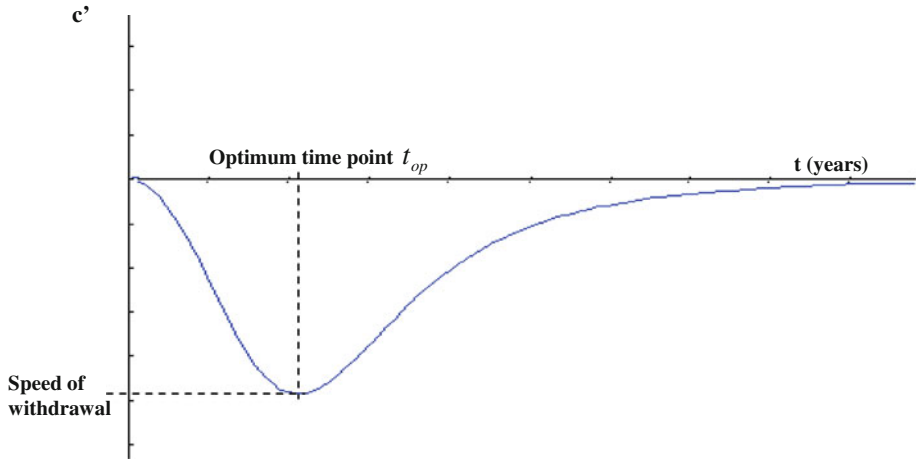
$$c' = dc/dt \tag{5}$$

representing the change in citations with time. In other terms, it represents the speed of citation change with time. Using Eq. 3 one calculates  $c'(t)$  as:

$$c'(t) = -\alpha p \beta \frac{t^{p-1}}{(t^p + \beta)^2}. \tag{6}$$

We note first that the derivative vanishes to zero when time  $t$  is equal to zero. The curve of  $c'(t)$  with respect to time  $t$  is given in the Fig. 4.

We see from Fig. 4 that a mathematical particularity of the derivative is its minima (or the maxima in absolute terms). It is the point that corresponds to the highest rate of



**Fig. 4** Evolution of the derivative of  $c$  with time

citations with time, or simply the maxima of the speed of citations distribution called previously *speed of withdrawal*.

The corresponding time point of the speed of withdrawal is the *Optimum-time* that is the inflection point  $t_{op}$  of the citation distribution  $c(t)$ . Consequently, the Optimum-time is the time point when the second derivative  $c''$  vanishes to zero:

$$c'' = \frac{d^2c}{dt^2} = \frac{dc'}{dt} = 0. \tag{7}$$

The inflection point means the time point when the acceleration of citations vanishes to zero. Using the derivative distribution in Eq. 6 one easily obtains from Eq. 7 the inflection time point  $t_{op}$  as:

$$t_{op} = \sqrt[p]{\frac{\beta(p-1)}{(p+1)}} \tag{8}$$

where  $p > 1$  and  $\beta$  is a positive constant.

We note from Eq. 8 that the inflection point, which is a characteristic of the citation distribution of the corpus under study, is interestingly written of only two specific parameters from the distribution model.

### Results and discussion

Once the model is set up, it is applied to *SCI* data for a sample of countries. I checked the validity of the model using the subfield ‘Biochemistry & Molecular biology’ (based on the *ISI* field classification). In fact, Narin and Hamilton (1996) have mentioned in their research on bibliometric performance measures that this subfield is one of the ‘very heavily citing scientific subfields’. Rather, Osareh and Wilson (1997) have shown that almost 58% of citing documents earned by 42 cited Third World Countries (TWC) during the period from 1985 to 1993 are in the field of biochemistry and biomedicine far ahead of others. To

compare their model of cumulative citations distribution to others, YU and Li (2010) used the publications and citations of *Journal of Biological chemistry and Biochemistry*. Thus, reasonable size of citations, stable patterns and regular trends of citation distribution are assured.

However, no emphasis should be attached to the particular subfield chosen for the modelling process. Pollmann (2000) has shown that the distribution of citations over time and the speed of decay is largely independent of the field and the medium of publication. He has noted this similarity might be found in some property of human cognition, and this is presumably broadly the same in all fields, languages and cultures for all times. Consequently, I believe the model can be accurately extended to other scientific fields or other corpuses than countries.

The time series for citations covers the period from 1985 to 2008. Citations being analysed are only those made to journal articles. The countries considered are major OECD-countries with more than 20000 articles published either in 2003 and 1996 (OECD 2006). This threshold is the only one used to build the sample. Germany is not included due to certain technical problems related to the situation of Germany (East and West Germany) in 1985, starting year for this analysis. The sample is then composed of Australia, England, France, Italy, Japan, and USA.

#### Citation aging model

The model is computed using Least Squares Method (LSM). A specific computational program is elaborated to obtain the model parameters. Further, *Eviews* software (version 4.1) based also on LSM, is used retrieving the same results. The fitting of the model to the observed data as a means of  $R - sq$  is reported in Table 1. Accordingly, the model fits the observed data with an error less than 1.1% in worse case. The model lines and the observed data from SCI for the countries sample are plotted in the Figs. A1, A2 and A3 in the Appendix.

Table 2 presents the model parameters for each of the countries considered. The values for  $t_v$  in Table 1 are rounded off to the nearest whole number; so is the life-cycle too.

From Table 2,  $\gamma$  is almost equal to 91% for Japan which means that the citation distribution of articles published in 1985 have lost more than 91% and kept just 9% of the maximum citations (gained at peak) after long time. Italy succeeded in maintaining almost 2% of the citations earned at peak, this is so even though Japan's  $p$ -factor is the greater (1.63) and  $\beta$ -factor is lesser (18.18) which results in steeper citations lost (see Fig. 3). Table 2 also shows that France's loss of citations is sharp (within 32 years) compared to England (35 years), USA (36 years) or Australia (72 years) after peak.

With respect to variable change leading to Eq. 2 and Fig. 1, one states that:

$$\text{Life-cycle} = t_v + (t_p - 1985) \text{ in Table 2.}$$

Life-cycle or life-time refers to the period of time starting from the publication year during which the corpus is still cited. As a result, all citations for France are lost within a time period of:  $32 + (1987 - 1985) = 34$  years, and even later for Australia (75 years).

**Table 1**  $R^2$  score for the model to the observed SCI data for countries sample

Country	Australia	USA	Italy	England	Japan	France
$R^2$ score	99.72	99.87	98.90	99.60	99.79	99.64



**Table 2** Model parameters results for countries sample

Country	$t_p$ (Eq. 1)	$c_0$	$\beta$	$p$	$\alpha$	$\gamma$ (%)	$t_v$ (Eq. 4)	Life-cycle*
USA	1988	30206	13.07	1.23	35011.58	–	36	39
England	1988	7856	10.71	1.18	9140.68	–	35	38
Japan	1987	6811	18.81	1.63	6171.77	90.61	–	$\infty$
France	1987	5255	25.95	1.50	5996.47	–	32	34
Australia	1988	1990	15.92	1.60	2023.72	–	72	75
Italy	1987	1469	22.28	1.59	1442.59	98.20	–	$\infty$

\* Life-cycle =  $t_v + (t_p - 1985)$

Even if the USA had, and still continues to have, the highest number of citations out of the sample of countries studied here, it has a finite life-cycle of 39 years ( $=t_v + 1988 - 1985$ ). France, with a life-cycle of 34 years, does not seem to maintain longer impact on knowledge world community in the subfield.

From what preceded, one can note that the ‘life-cycle’ is extensively different from one country to another, which in turn might also be quite different if the subfield is to be changed. This life-cycle is demonstrated to be either an infinite or finite period. Even in the latter case the period ranges from 34 to 75 years, unlike what Glänzel (2007) has showed that the particular choice of the citation window is, except for a short initial period, not important for citations class-size (25% highly cited, 5% remarkably cited, 18% fairly cited, 75% poorly cited). In fact, according to Glänzel’s analysis, the citations distribution is truncated. Setting up a time frame of 21 years means that all citations received beyond this time frame, which are not inconsequential, are simply unaccounted for in his analysis. In contrast, we have seen that a country (or a corpus) may maintain around 9% of its citations gained at peak over an indefinite period of time.

### Speed of withdrawal and optimum-time

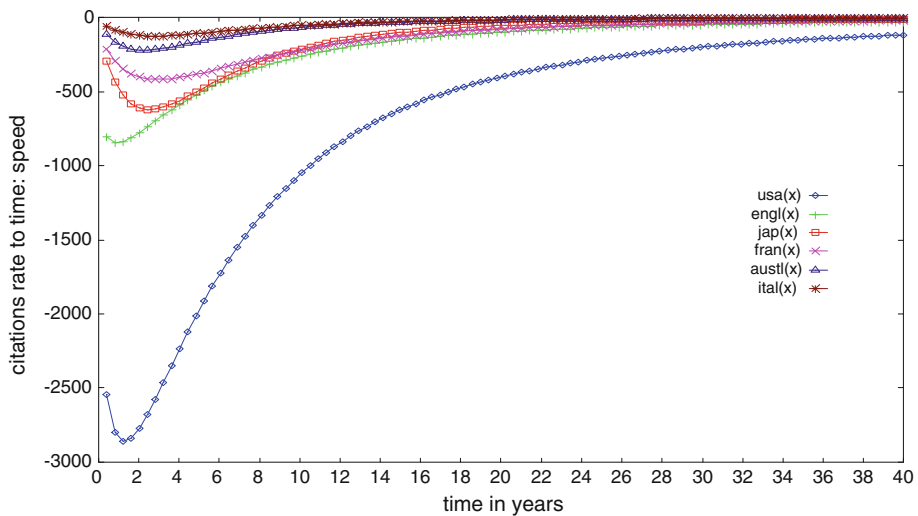
Using the model parameters  $p$  and  $\beta$ , for the sample of countries considered in this analysis, I find the values listed in the Table 3 for the Optimum-time and the corresponding speed of withdrawal. I should remind that, considering the Eq. 1 and 8 and that the starting year is 1985, the real Optimum-time is then:  $t_{op} + (t_p - 1985)$ .

Table 3 shows that the Optimum-time, when major decrease of citations happens, is almost the same for all the countries (4–5 years) after publication year. It then mathematically proves, relying on observed data, the finding of some researches that proposed a

**Table 3** Optimum-time and speed of withdrawal index for countries sample

Country	$t_p$ (Eq. 1)	$p$	$\beta$	$t_{op}$ (Eq. 8)	Optimum-time*	Speed of withdrawal $c'(t_{op})$
USA	1988	1.23	13.07	1.3	4	–2863
England	1988	1.18	10.71	0.9	4	–843
Japan	1987	1.63	18.81	2.5	5	–623
France	1987	1.50	25.95	3.0	5	–417
Australia	1988	1.60	15.92	2.3	5	–219
Italy	1987	1.59	22.28	2.8	5	–125

\* Optimum-time =  $t_{op} + (t_p - 1985)$ . The value is rounded off to the nearest whole number



**Fig. 5** Speed distribution with time for countries sample

fixed period. But it happens to be different and ranges from less than 1 year after the peak time such as the case of England, to 3 years such as the case of France or Italy.

Now that the Optimum-time is known, we can fairly calculate for each of the countries the speed of the citations decrease at this particular point. But prior to that, one should note from Table 3 that the findings made in the first part are once again confirmed. Indeed, England seems to be the country with lowest value of  $\beta$ , making it losing all its citations within a time period of 38 years, and also with the closer decay year from the peak year time, with less than 1 year (see Table 3).

Remarkably, the other finding is the fact that all the countries (corpuses) loose their major citations gained at peak within almost the same optimum-time period but maintain these citations over different life-cycle. Indeed, Italy and Japan would save a part of their citations indefinitely rather than other countries such as France or Australia even if they both have the same Optimum-time of 5 years.

Even though the USA had the highest number of citations out of the sample, and its life-cycle is 39 years, the *speed of withdrawal* of citations is the highest. Almost 3000 citations are lost within a time unit (year) at the Optimum-time (Table 3). The speed of citation decrease for USA is reported in Fig. 5 with the other countries of the sample.

## Conclusion

In this study I propose a model to represent the naturally observed citations distribution and citation aging in a diachronous-based-method. Using this model one can easily obtain the '*residual citations*' which are citations kept by a corpus after a long time. The model proves that the residual citations may be greater than zero meaning that the life-cycle of this corpus is infinite. In the other case, the life-cycle is a finite period and found to vary widely from one corpus to another.

The proposed model is proved to be suitable to represent the observed citation distribution over time and citation aging. Based on Least Squares Method, the model fits well

the observed data extracted from *SCI* with an error not exceeding 1.1%. This error may rather be attributed to the random increase of the number of citations due to phenomenon such as ‘sleeping beauties’, publication delay or literature growth, in a given year than to a regular trend in the distribution; since the error is lesser with great numbers of citations offering reasonable size of citations for analysis and consequently, tend to hide any ‘side-effects’ of this phenomenon.

Once the model is set up, it becomes a straightforward task to determine precisely the Life-cycle (period of time after-which the corpus is not cited anymore), the *Optimum-time* when the major loss of citations happens and to predict the number of citations that would a corpus obtain at any time of its life-cycle.

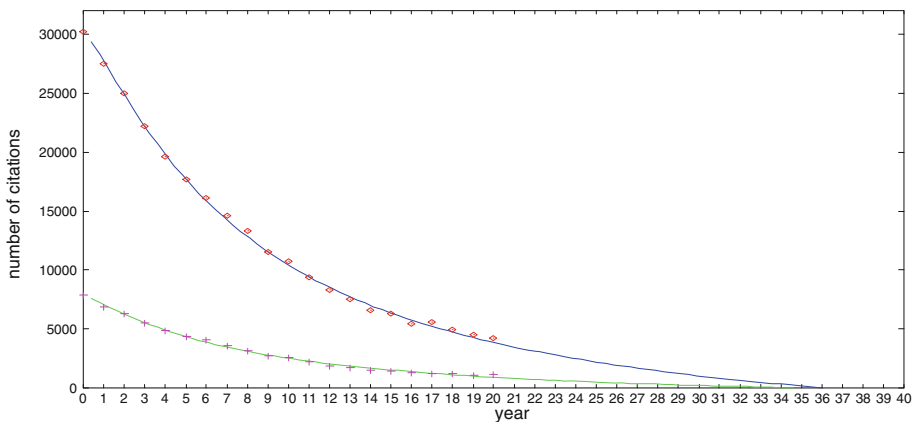
After answering the question of how long a publication will continue being cited by calculating the life-cycle, I determine with more accuracy the *Speed of withdrawal* that is simply the rate of decline of the distribution at the *Optimum-time* (i.e. the maximum of the rate). The *Optimum-time* is almost the same for the countries of the sample ranging from 4 to 5 years time period. However, we find out that even if this period, from the publication year to the optimum-time is the same, the peak time to the *Optimum-time* is likely different from one country to another moving then like a trigger inside this fixed period.

Furthermore, even if the *Optimum-time* is found not to be characteristic of the corpus and is independent of its citations size, the speed of withdrawal is interestingly directly related to the size of the citations at peak time.

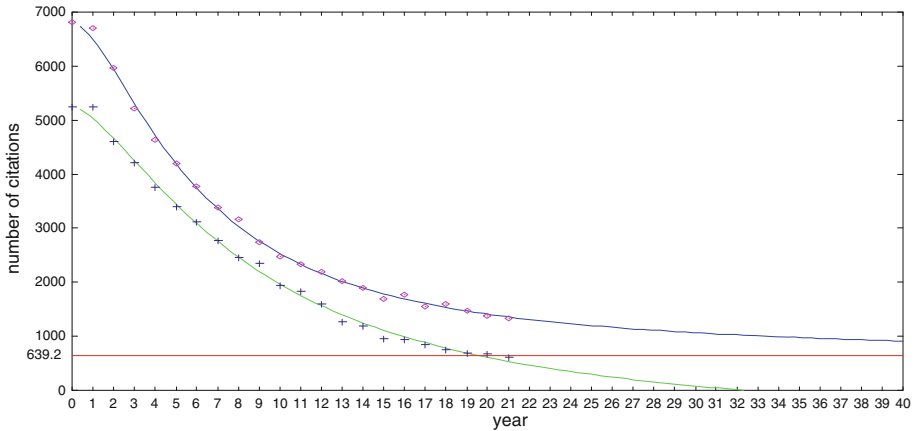
**Acknowledgment** The author is grateful to the anonymous referees for the very critical and valuable remarks leading to substantial improvements of the initial version of the paper. The author is also grateful to Dr. Elhamine S. for his help in data computing.

## Appendix

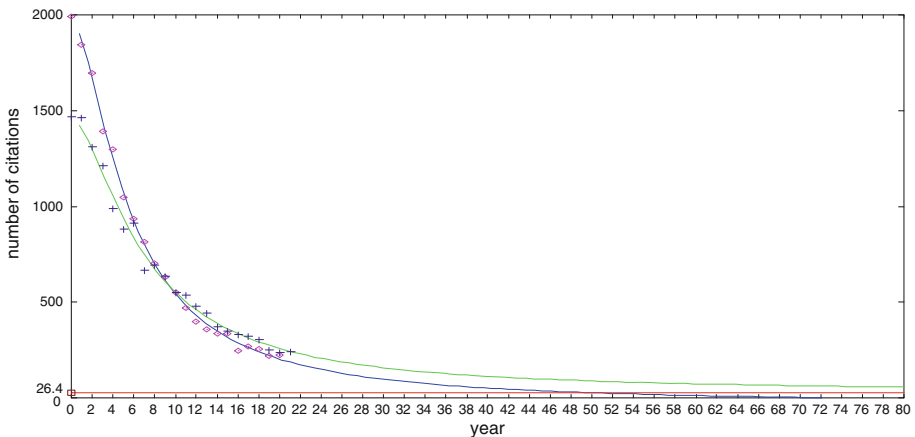
See Figs. [A1](#), [A2](#), [A3](#).



**Fig. A1** Model (line) and SCI data (scattered points) for USA (diamond) and England (plus)



**Fig. A2** Model (*line*) and SCI data (scattered points) for Japan (*diamond*) and France (*plus*)



**Fig. A3** Model (*line*) and SCI data (scattered points) for Australia (*diamond*) and Italy (*plus*)

## References

- Avramescu, A. (1979). Actuality and obsolescence of scientific literature. *Journal of the American Society for Information Science (pre-1986)*, 30(5), 296–303.
- Bouabid, H., & Martin, B. R. (2009). Evaluation of Moroccan research using a bibliometric-based approach: Investigation of the validity of the h-index. *Scientometrics*, 78(2), 203–217.
- Burrell, Q. L. (2002). The nth-citation distribution and obsolescence. *Scientometrics*, 53(3), 309–323.
- Burrell, Q. L. (2003). Predicting future citation behavior. *Journal of the American Society for Information Science and Technology*, 54(5), 372–378.
- Egghe, L., & Ravichandra Rao, I. K. (1992). Citation age data and the obsolescence function: Fits and explanations. *Information and Processing Management*, 28(2), 201–217.
- Egghe, L., & Rousseau, R. (2000a). The influence of publication delays on the observed aging distribution of scientific literature. *Journal of the American Society for Information Science*, 51(2), 158–165.
- Egghe, L., & Rousseau, R. (2000b). Aging, obsolescence, impact, growth, and utilization: Definitions and relations. *Journal of the American Society for Information Science*, 51(11), 1004–1017.
- Glänzel, W. (2004). Towards a model for diachronous and synchronous citation analyses. *Scientometrics*, 60(3), 511–522.

- Glänzel, W. (2007). Characteristic scores and scales. A bibliometric analysis of subject characteristics based on long-term citation observation. *Informetrics*, 1(1), 29–102.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*, 40(5), 342–349.
- Mingers, J., & Burrell, Q. L. (2006). Modeling citation behavior in management science journals. *Information Processing and Management*, 42, 1451–1464.
- Nadarajah, S., & Kotz, S. (2007). Models for citation behaviour. *Scientometrics*, 72(2), 291–305.
- Nakamoto, H. (1988). Synchronous and dyachronous citation distributions. In L. Egghe & R. Rousseau (Eds.), *Informetrics 87/88* (pp. 157–163). Amsterdam: Elsevier Science Publishers.
- Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, 36(3), 293–310.
- OECD. (2006). *Science, technology and industry outlook 2006*. Paris: OECD.
- Osareh, F., & Wilson, C. S. (1997). Third World countries (TWC) research publications by disciplines: A country by country citation analysis. *Scientometrics*, 39(3), 253–266.
- Pollmann, T. (2000). Forgetting and the aging of scientific publication. *Scientometrics*, 47(1), 43–54.
- Redner, S. (2004). Citation statistics for more than a century of physical review. *Physics Today*, 58(6), 49–54.
- Simkin, M. V., & Roychowdhury, V. P. (2007). A mathematical theory of citing. *Journal of the American Society for Information Science and Technology*, 58(11), 1661–1673.
- Stinson, E. R., & Lancaster, F. W. (1987). Synchronous versus diachronous methods in the measurement of obsolescence by citation studies. *Journal of Information Science*, 13, 65–74.
- van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, 59, 467–472.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2009). Modeling a century of citations distributions. *Journal of Informetrics*, 3(4), 296–303.
- Yu, G., Guo, R., & Li, Y. J. (2006). The influence of publication delays on three ISI indicators. *Scientometrics*, 69(3), 511–527.
- Yu, G., & Li, Y.-J. (2010). Identification of referencing and citation processes of scientific journals based on the citation distribution model. *Scientometrics*, 82, 249–261.