Information Processing and Management 49 (2013) 757–772

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Growth dynamics of citations of cumulative papers of individual authors according to progressive nucleation mechanism: Concept of citation acceleration

Keshra Sangwal *

Department of Applied Physics, Lublin University of Technology, ul. Nadbystrzycka 38, 20-618 Lublin, Poland

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 7 January 2013
Accepted 10 January 2013
Available online 28 February 2013

Keywords:
Citation analysis
Citation acceleration *a*
h Index
Progressive nucleation mechanism

ABSTRACT

Using data generated by progressive nucleation mechanism on the cumulative fraction of citations of individual papers published successively by a hypothetical author, an expression for the time dependence of the cumulative number $L_{\text{sum}}(t)$ of citations of progressively published papers is proposed. It was found that, for all nonzero values of constant publication rate ΔN, the cumulative citations $L_{\text{sum}}(t)$ of the cumulative N papers published by an author in his/her entire publication career spanning over T years may be represented in distinct regions: (1) in the region $0 < t < \Theta_0$ (where $\Theta_0 \approx T/3$), $L_{\text{sum}}(t)$ slowly increases proportionally to the square of the citation time t, and (2) in the region $t > \Theta_0$, $L_{\text{sum}}(t)$ approaches a constant $L_{\text{sum}}(\max)$ at T. In the former region, the time dependence of $L_{\text{sum}}(t)$ of an author is associated with three parameters, viz. the citability parameter λ_0, the publication rate ΔN and his/her publication career t. Based on the predicted dependence of $L_{\text{sum}}(t)$ on t, a useful scientometric age-independent measure, defined as citation acceleration $a = L_{\text{sum}}(t)/t^2$, is suggested to analyze and compare the scientific activities of different authors. Confrontation of the time dependence of cumulative number $L_{\text{sum}}(t)$ of citations of papers with the theoretical equation reveals one or more citation periods during the publication careers of different authors.

1. Introduction

Several models have been developed and applied over years to analyze the number and growth characteristics of journals, articles and authors in different scientific fields. Among the different equations of various models, power-law, exponential and logistic functions are commonly used (Egghe & Ravichandra Rao, 1992; Glänzel, 1997, 2004; Gupta, Kumar, Sangam, & Karisiddappa, 2002; Gupta, Sharma, & Karisiddappa, 1995; Naranan, 1970; Price, 1963; Ravichandra Rao & Srivastava, 2010; Wong & Goh, 2010). Recently, Sangwal (2011a) proposed a new equation, based on progressive nucleation mechanism (PNM) of a solid phase during its crystallization in a closed liquid system of fixed volume. These models have also been used to analyze the time dependence of growth behavior of citations (Egghe, Ravichandra Rao, & Rousseau, 1995; Gupta, 1999; Sangwal, 2012a, 2012b).

For over two decades there has been an increasing interest in the evaluation of the scientific research output of scientists in terms of numerical indexes quantifying it unequivocally. In recent years, the *h* index proposed by Hirsch (2005) to quantify the research output of individual scientists has drawn constant attention in the academic literature. Apart from contributions dealing, among others, with improvement and modification of the *h* index (for example, see: Alonso, Cabrerizo,

* Tel.: +48 81 5384 504; fax: +48 81 5384 731.
E-mail address: k.sangwal@pollub.pl

0306-4573/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ipm.2013.01.003

Herrera-Viedma, & Herrera, 2009; Anderson, Hankin, & Killworth, 2008; Burrell, 2009; Csajbók, Berhidi, Vasas, & Schubert, 2007; Egghe, 2010a, 2010b, 2010c; Franceschini & Maisano, 2010a; Franceschini & Maisano, 2010b; Glänzel & Schubert 2010; Jin, Liang, Rousseau, & Egghe, 2007; Kosmulski, 2006; Navon, 2009; Prathap, 2006; Schubert, 2007) and discussion on the relationship between different bibliometric evaluation measures (Burrell, 2009; Van Raan, 2006), attempts have been made to give mathematical models to the h index and its modifications and to investigate its dependence on time (Burrell, 2007a, 2007b, 2009; Egghe, 2007, 2008, 2009; Glänzel, 2008; Hirsch, 2005; Nair & Turlach, 2012; Ye & Rousseau, 2008).

Among the various models proposed so far for the investigation of the dynamics of cumulative citations L_{sum} and h index of individual authors, the deterministic model of Hirsch (2005) and the stochastic model of Burrell (2007a) are of interest in the present work. Assuming that an author publishes a constant number of papers per year (ΔN) and that each published paper receives a constant number of citations per year (ΔL) every subsequent year, Hirsch (2005) gave relationships between total number L_{sum} of citations and h index and between h index and publication time t. For not-too-small publication time t, he concluded that

$$L_{\mathrm{sum}}(t) \approx Ah^2(t), \tag{1}$$

and

$$h(t) \approx bt, \tag{2}$$

where A is an empirical constant (called Hirsch constant hereafter) which Hirsch found empirically to lie between 3 and 5, and the slope b of $h(t)/t$ plot is related both to ΔN and ΔL of an author. In fact, using the values of b equal to 1, 2 and 3, Hirsch defined successful, outstanding and unique scientists, respectively.

In order to obtain insight into the h index as a measure of an author's research output and its impact, Burrell (2007a) proposed a stochastic model for an author's production/citation process and investigated the dependence of his/her h index on the publication rate ΔN of papers, the citation rate ΔL of published papers and his/her career length t. For the investigations, Burrell (2007a) assumed that the author publishes papers at regular intervals at a certain rate (ΔN = constant), after their publication these papers receive citations, and that both processes of accumulation of publications and citations are random. Moreover, he assumed that the citation rates of different papers are different (i.e. ΔL different for different papers). These modeling results suggested that the h index of an author is approximately proportional to his/her career length and approximately linear functions of the logarithm of the rates of publication of papers and their citations. Burrell (2007b) later applied his stochastic model to investigate, among others, the time dependence of the square of h index (called h-core) and found that the cumulative citations

$$L_{\mathrm{sum}}(t) \approx \frac{\Delta N \cdot \Delta L \cdot t^2}{2}, \tag{3}$$

and

$$h(t) \approx (2\Delta L)^{1/2} \cdot t. \tag{4}$$

According to these relations, the ratio $L_{\mathrm{sum}}(t)/h(t)$, defined as Jin's A index (see Burrell, 2007b), increases linearly with time t.

Egghe (2009) examined the dependence of h on t, assuming that citations of papers of a researcher are described by Lotka's law, i.e.

$$f(j) = \frac{K}{j^{\delta}}, \tag{5}$$

where $f(j)$ denotes the density of articles with a density j of citations, the constant $K > 0$, and the exponent $\delta > 1$, which remains constant in the time period t. He considered situations when the rate of publication ΔN of his/her papers during his/her career remains constant, grows following power law, and grows exponentially. He found that, with increasing t, h increases concavely in the former two cases whereas it increases either concavely in the entire t or shows an initial increase followed by a decrease (i.e. S-shaped $h(t)$). However, Egghe (2009) found that his own $h(t)$ data follows a linear dependence. It may be seen that a linear dependence between h and t is expected in the case of constant publication rate ΔN when $\delta = 1$ in Lotka's law.

The main prediction of the deterministic model of Hirsch (2005) and the stochastic model of Burrell (2007a, 2007b) is that the ratio $h(t)/t$ is expected to be a time-independent constant which, according to Hirsch (2005) "should provide a useful yardstick to compare scientists of different seniority". In a later work, Burrell (2007c) reexamined the previously published data of the evolution of $h(t)$ for eleven scientists by Liang (2006) to test the time independence of the ratio $h(t)/t$, called h rate, predicted by the stochastic model by distinguishing between: (1) raw h rate at given time t and (2) least-squares h rate at time t, given by the slope of the regression line passing through the origin of the data of evolving h index with time t. The results revealed that in many cases the h rate does not remain constant over long periods of time. Similar trends of h rate may be observed from the plots of the growth of Hirsch index h as a function of citation time t, reported by Anderson et al. (2008), of six scientists elected in 2006 to the membership of the Royal Society. Their plots of h against t also indicate that h rate of an individual author is different in different time intervals for practically all of these authors. However, Burrell

(2007c) concluded that tracking the evolution of the h index of an individual author and using the least-squares h rate is a better scientometric approach to investigate his/her scientific career.

Burrell (2007a) recognized the simplicity of his model and admitted that all of the assumptions on which the model is based are open to criticism. For example, neither the publication rate of an author nor the citation rate of his/her papers remains constant over time. It is also well known (Sangwal, 2012a) that the citation period of a paper usually lasts between 10 and 15 years, and the citation rate of a paper initially increases and then, after going through a maximum value, it finally approaches zero.

Recently, the present author (Sangwal, 2012b) proposed a general approach to explain the cumulative number $L(t)$ of items at time t produced by an individual source (system) using progressive nucleation mechanism (PNM). The final expression of the PNM is based on the postulate that, once a source for items is formed, items are nucleated in it progressively at a stationary rate until they attain a maximum value C. In a later paper (Sangwal, 2012a), the present author applied the PNM for the growth behavior of items to describe the cumulative citations $L(t)$ of an individual ith paper of an author. It was found that the PNM is indeed followed in the case of short citation durations less than about 15 years where the condition of stationary nucleation is *more or less* satisfied.

The aim of the present study is twofold: (1) to model data on the cumulative fraction of citations of individual papers published successively by a hypothetical author using the PNM, where his/her papers are characterized by different citation parameters, and, on the basis of the modeling data, to propose an expression for the time dependence of the cumulative number $L_{\text{sum}}(t)$ of citations received by all of his papers published successively during his/her career and to give physical interpretation of different constants of the final expression, and (2) to confront the time dependence of cumulative number $L_{\text{sum}}(t)$ of citations of papers published by selected authors with the theoretical predictions.

2. Progressive nucleation mechanism for growth behavior of citations

The PNM for the growth behavior of citations with time t in an individual system is based on the following postulates (Sangwal, 2012a, 2012b):

(1) Citations received by a paper published by an author and the paper earning these citations compose a closed system in which the process of occurrence of citations is stationary.
(2) The occurrence of citations of a paper occurs progressively with time and finally approaches a constant value C which is the maximum number of citations received by the paper at time T.
(3) The nature of the dependence of cumulative number of citations $L_i(t)$ of an author's ith paper (where i is a positive integer) at time t is determined by a maximum number of citations C_i, a time constant Θ_i and an exponent q_i.
(4) The citation behavior (pattern or trend) of different papers of an author is characterized by different values of C_i, Θ_i and q_i of each ith paper.
(5) The cumulative number $L_{\text{sum}}(t)$ of citations at time t is the sum of contributions at time t from each ith paper.

The above concepts are used below to discuss the citation behavior of an individual author's papers receiving different citations as characterized by different values of their time constants Θ and exponents q of each ith paper. However, for the analysis first we consider the fraction $\alpha(t)$ of cumulative citations $L(t)$ at time t for an individual paper. In the case of growth of citations of individual papers of an author with time t since the year Y_0, the time dependence of $\alpha(t)$ is given by (Sangwal, 2011a, 2012b)

$$\alpha(t) = \frac{L(t)}{C} = \left[1 - \exp\left\{-\left(\frac{t}{\Theta}\right)^q\right\}\right], \tag{6}$$

where C is the maximum number of citations that the paper can receive (i.e. citability of the paper), the time constant

$$\Theta = \frac{q^{1/q}}{\kappa J_s}, \tag{7}$$

and the exponent

$$q = 1 + \nu d. \tag{8}$$

In the above equations, J_s is the rate of stationary nucleation, κ is the shape factor (e.g. $\kappa = 4\pi/3$ for a sphere), d is the dimensionality of the growing nuclei, and the time

$$T = Y - Y_0, \tag{9}$$

where Y is the year of the citations $L(t)$ and Y_0 is the actual or extrapolated year when $\alpha(t) = 0$. In the case of growth of nuclei by diffusion and mass transfer processes, the parameter $\nu = 1/2$ and 1, respectively.

If $L_i(t)$ denotes the cumulative citations of the ith paper and α_i is the corresponding fraction of citations, the cumulative fraction α_{sum} of citations from a collective of n papers may be written as

$$\alpha_{\text{sum}}(t) = \sum_{i=0}^{n} \alpha_i(t) = \sum_{i=0}^{n} \frac{L_i(t)}{C_i}. \tag{10}$$

When new papers are published successively at equal time intervals Δ (a positive integer), the cumulative fraction $\alpha_{sum}(t)$ of citations may be given by

$$\alpha_{sum}(t) = \sum_{i=1}^{n} \alpha_i[t-(i-1)\Delta] = \sum_{i=1}^{n} \frac{L_i[t-(i-1)\Delta]}{C_i} = \sum_{i=1}^{n}\left[1-\exp\left\{-\left(\frac{t-(i-1)\Delta}{\Theta_i}\right)^{q_i}\right\}\right], \quad (11)$$

where C_i, Θ_i and q_i refer to the ith paper, and Δ is the time interval when a new paper is published. Eq. (11) does not have a simple solution because the citations of every new $(i+1)$th paper are characterized by different sets of C_{i+1}, q_{i+1} and Θ_{i+1}. However, Eq. (11) can be solved numerically for real collectives of n papers.

It should be mentioned that Eq. (11) satisfies the following conditions for n successively published papers: (1) when $i = 1$ or $\Delta = 0$, $\alpha_{sum}(t) = n\alpha_1(t)$, (2) for all values of $\Delta > 0$ and $i > 1$, $\alpha_{sum}(t) = \sum\alpha_i(t)$, and (3) the total citation duration $T = t > (i-1)\Delta$.

3. Modeling the citation behavior of individual and collectives of papers

As described previously (Sangwal, 2012a), the citation behavior of individual papers of an author can be described satisfactorily using Eq. (6) based on the PNM. Consequently, one can use the citation data generated by using these equations for individual papers of the hypothetical author and subsequently find solutions of Eq. (11). For this purpose we choose the following situations:

(1) The author publishes in the first year four papers simultaneously and these papers are characterized by a fixed value of time constant Θ and different values of the exponent q_i equal to 1.5, 2, 3 and 5, denoted hereafter as q_{1-4} for the set.
(2) During his/her publication career the author publishes one paper per year since the publication of his/her first paper in sets of four succeeding papers characterized by a preselected value of time constant Θ lying between 5 and 15 years and different values of q_i equal to 1.5, 2, 3 and 5 for the succeeding papers in each set.
(3) This situation is similar to that of situation (2) but now the sets of four succeeding papers are characterized by a preselected value of time constant Θ lying between 5 and 15 years and different values of q_i equal to 5, 3, 2 and 1.5 for every four-paper set. Obviously, the order of q_i is opposite to that in the previous situation.

It is well known that: (1) the publication career of an author usually lasts about 40 years, (2) many authors publish their papers during their entire career but there are also authors who publish a few papers only, and (3) some papers published by an author fetch high citations, others are cited poorly while the remaining papers remain uncited. Therefore, for the purpose of finding solution of Eq. (11), one requires data of cumulative citations $L(t)$ of individual and collectives of papers published by various hypothetical authors. However, the real situations of the citation behaviors of the publication output of different authors are more complicated than the simplified situations given above.

Using the data generated for the above situations, a general expression of the time dependence of cumulative fraction $\alpha_{sum}(t)$ of citations of collectives of papers in the form of approximate solution of Eq. (11) is given below. The time scale of 40 years for the citation data generated by using Eq. (6) for the analysis was selected in view of the fact that the research career of a majority of authors usually spans over 40 years. The value of the time constant Θ for the citations of different individual papers lies between about 5 and 15 years, whereas that of the exponent q usually lies between 1 and 3 (Sangwal, 2012b). Keeping these facts in mind, the above values of Θ and q were selected.

Fig. 1 shows the dependence of cumulative fraction $\alpha_i(t)$ of citations of four individual ith papers (i = 1, 2, 3 and 4) by a single author or four different authors (lower four curves), where the citability of the papers is characterized by the same

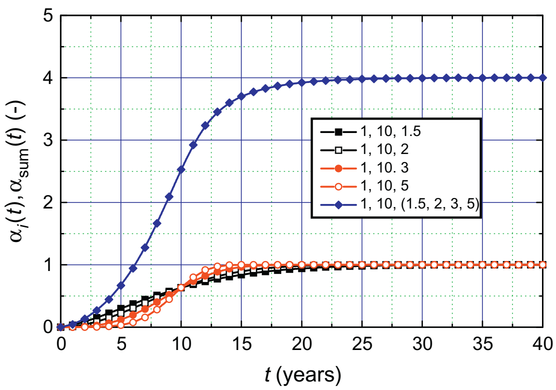

Fig. 1. Dependence of cumulative fraction $\alpha_i(t)$ of citations of four individual papers i characterized by Θ = 10 years and q_{1-4} = 1.5, 2, 3 and 5 (lower four curves), and resultant cumulative fraction $\alpha_{sum}(t)$ of citations produced by a set of four papers with Θ = 10 years and q_{1-4} = 1.5, 2, 3 and 5 (upper lone curve). See text for details.

Table 1
Constants of Eq. (11) for four papers published simultaneously.

Parameter	Fitting parameters			
	δ (years)	α_0 (year^{-1})	Θ_0 (years)	q_0 (–)
5, 3, 2, 1.5	0	3.962 ± 0.012	9.96 ± 0.05	2.528 ± 0.042
	0.5	3.959 ± 0.011	10.47 ± 0.05	2.682 ± 0.042
	1.0	3.957 ± 0.011	10.98 ± 0.05	2.837 ± 0.042

value of time constant Θ = 10 years but with different exponent q_i = 1.5, 2, 3 and 5 for papers 1, 2, 3 and 4, respectively. The plots were drawn using the values of Θ and q_i in Eq. (6). The cumulative fraction $\alpha_{sum}(t)$ of citations produced by these four papers with Θ = 10 years and q = 1.5, 2, 3 and 5 is presented in the upper lonely curve. This cumulative fraction $\alpha_{sum}(t)$ of citations represents the situation of an author who published only four papers in the beginning of his/her publication career.

From Fig. 1 the following two features may be noted:

(1) Irrespective of the citability of an individual paper as defined by the values of time constant Θ and exponent q in Eq. (6), they approach the maximum fraction $\alpha_i(\max)$ of citations equal to unity.
(2) The nature of the curve of the time dependence of the cumulative fraction $\alpha_{sum}(t)$ of citations is similar to that of the curves for the individual papers but now the maximum fraction $\alpha_{sum}(\max)$ of citations is equal to the sum of the fractions $\alpha_i(\max)$ of citations of individual papers, i.e. $\alpha_{sum}(\max) = \sum\alpha_i(\max)$. In the above example of Fig. 1 where $\alpha_i(\max) = 1$ for the ith paper and the number of individual papers is four, $\alpha_{sum}(\max) = 4$.

It should be mentioned that the data of the cumulative fraction $\alpha_{sum}(t)$ of citations, shown in the upper curve of Fig. 1, can also be represented by Eq. (6). The best-fit values of constants, denoted here as α_0, Θ_0 and q_0, for the above data are given in Table 1. In the table are listed two sets of the best-fit values of the constants corresponding to initial time $(t - \delta)$ in Eq. (6), where δ is an empirical correction time which gives a better fit for the data in terms of α_0. As seen from Table 1, the correction time δ leads to higher values of time constant Θ_0 and exponent q_0, but α_0 essentially remains unaltered. However, in terms of Θ_0 and q_0, the best fit is obtained when $\delta = 0$.

The generated data of cumulative fraction $\alpha_{sum}(t)$ as a function of time t of citations of 40 papers of different citability, published successively by a hypothetical author who published one paper each year since the beginning of their publication careers, were considered for the analysis. For the generation of the data the papers considered were characterized by three different values of Θ (i.e. 5, 10 and 15 years) and two sets of q_i of successively published four papers in the sequences: (i) 1.5, 2, 3 and 5, and (ii) 5, 3, 2 and 1.5.

Fig. 2 shows as an example the dependence of cumulative fraction $\alpha_{sum}(t)$ of citations on publication time t of successively published 40 papers of different citability by an author who published one paper each year since the beginning of his/her publication career. In this particular case, the citability of the papers is characterized by Θ = 10 years and a set of q_i of successively published four papers in the sequence: 1.5, 2, 3 and 5. The data points present situations when our hypothetical author publishes 4p papers (where the integer p increases from 1 to 10) in his/her publication career and the citability of his/her successive sets of four papers is characterized in the sequence: q_{1-4} = 1.5, 2, 3 and 5. However, a closely resembling dependence is observed for the citability of his/her successive sets of four papers characterized in the sequence: q_{1-4} = 5, 3, 2 and 1.5.

From Fig. 2 one notes that, as in the case of simultaneous publication of a set of papers, the maximum fraction $\alpha_{sum}(\max)$ of citations is equal to the sum of the fractions $\alpha_i(\max)$ of citations of individual papers i, i.e. $\alpha_{sum}(\max) = \sum\alpha_i(\max)$.

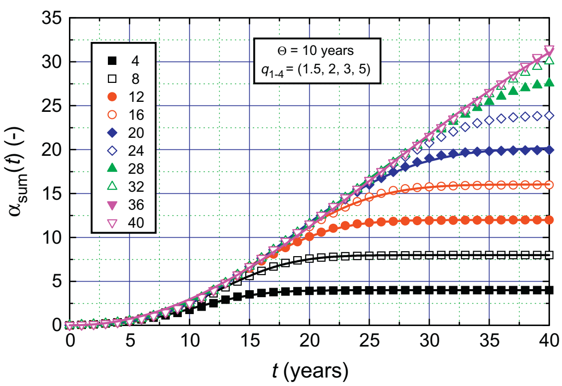

Fig. 2. Plots of cumulative fraction $\alpha_{sum}(t)$ of citations of successively published papers of different citability with Θ = 10 years and different sets of q_{1-4} of successive papers: (a) 1.5, 2, 3 and 5, and (b) 5, 3, 2 and 1.5.

Moreover, as found previously (Sangwal, 2012a), despite different sets of C_i, q_i and Θ_i corresponding to different sets of the four papers, the data can be represented satisfactorily by the relation

$$\alpha_{sum}(t) = \sum_{i=0}^{n} \alpha_i(t - i\Delta) = \alpha_0 \left[1 - \exp\left\{-\left(\frac{t}{\Theta_0}\right)^{q_0}\right\}\right], \quad (12)$$

where Θ_0 is a new time constant and q_0 is a new exponent describing the resultant growth behavior of the entire collective of papers, and α_0 is the sum of all of the maximum fractions. The best-fit values of the constants α_0, Θ_0 and q_0 for the citability data of the two sets of successively published four papers in the sequences (i.e. q_{1-4} = 1.5, 2, 3 and 5, and q_{1-4} = 5, 3, 2 and 1.5) considered above according to Eq. (12) are presented in Fig. 3 by open and filled points, respectively, as a function of publication duration t. In Fig. 3a the solid curve represents $\alpha_0 = t$ (i.e. the slope is unity) whereas the dashed curve is the best fit of $\alpha_{sum}(t)$ data for initial Θ = 10 years described by relation: $\alpha_0 = -1.74 + 1.16t$, $r^2 = 0.994$. In Fig. 3b the solid curves are drawn with slope 0.5 for initial Θ = 5, 10 and 15 years, dashed curve covering the entire $\Theta_0(t)$ data for initial Θ = 10 years represents the best fit according to the relation: $\Theta_0 = 7.71 + 0.72t$, $r^2 = 0.993$, whereas dotted curve representing $\Theta_0(t)$ data for initial Θ = 5 years is drawn with slope of 2.

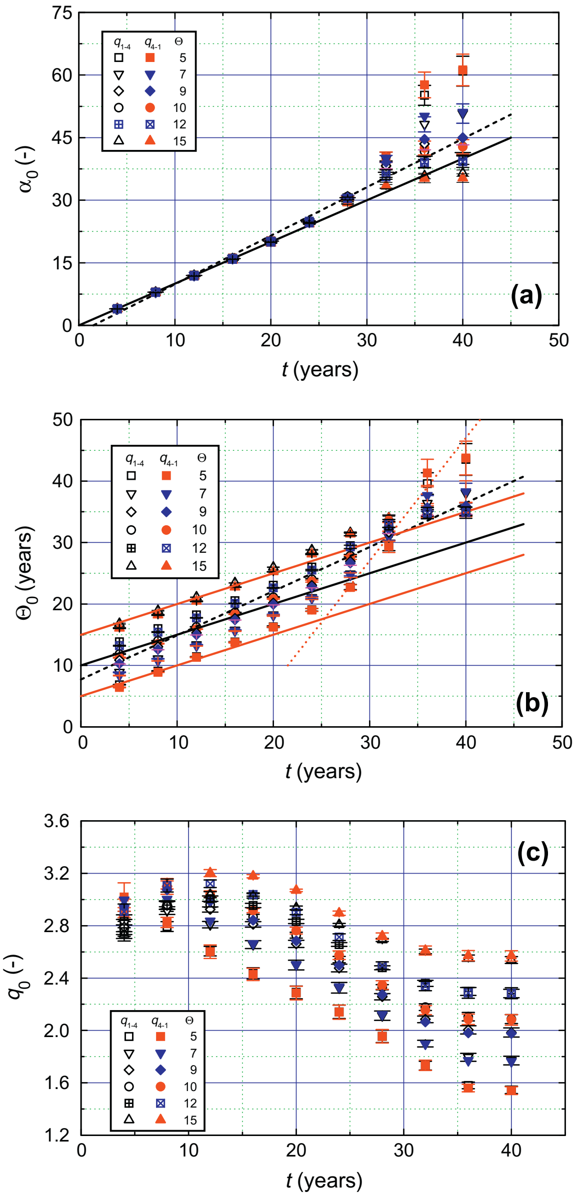

Fig. 3. Dependence of parameters: (a) α_0, (b) Θ_0 and (c) q_0 for the citations of progressively published papers in sets of two sequences of q_{1-4} for the citability of papers characterized by six different values of Θ as a function of publication duration t of hypothetical authors. See text for details.

Fig. 3a shows that the value of α_0 does not depend significantly on the choice of q_{1-4} for a particular value of initially used time constant Θ to generate the data and the value of α_0 also does not depend on the value of initial Θ up to about $t < 30$ years. However, among the different initial Θ used to generate the $\alpha_{\mathrm{sum}}(t)$ data, a linear dependence of α_0 on t in the entire t range up to 40 years is obtained when the value of Θ is about 10 years. Then $\alpha_0 \approx \alpha_{\mathrm{sum}}(\max) = \sum\alpha_i(\max)$ up to 40 years.

Fig. 3b also shows that the value of Θ_0 does not depend significantly on the choice of q_{1-4} for a particular value of initial Θ used to generate the data of $\alpha_{\mathrm{sum}}(t)$ as a function of time t. For $t < 32$ years, the higher the value of initial Θ used, the higher is the value of Θ_0. In this publication duration, the dependence of Θ_0 on t follows the relation

$$y(t) = y_0 + y_1 t + y_2 t^2, \tag{13}$$

where y_0, y_1 and y_2 are fitting parameters and $y_0 = \Theta$. At low t (<15 years), Θ_0 increases practically linearly with t with a slope equal to about 0.5, but the upper limit of t for this linear dependence somewhat increases with increasing values of the initial Θ used to generate the $\alpha_{\mathrm{sum}}(t)$ data. After this initial period of linear dependence, the slope increases first, then decreases such that finally Θ_0 tends to attain a constant value for $t > 32$ years. This behavior may be seen from the data for

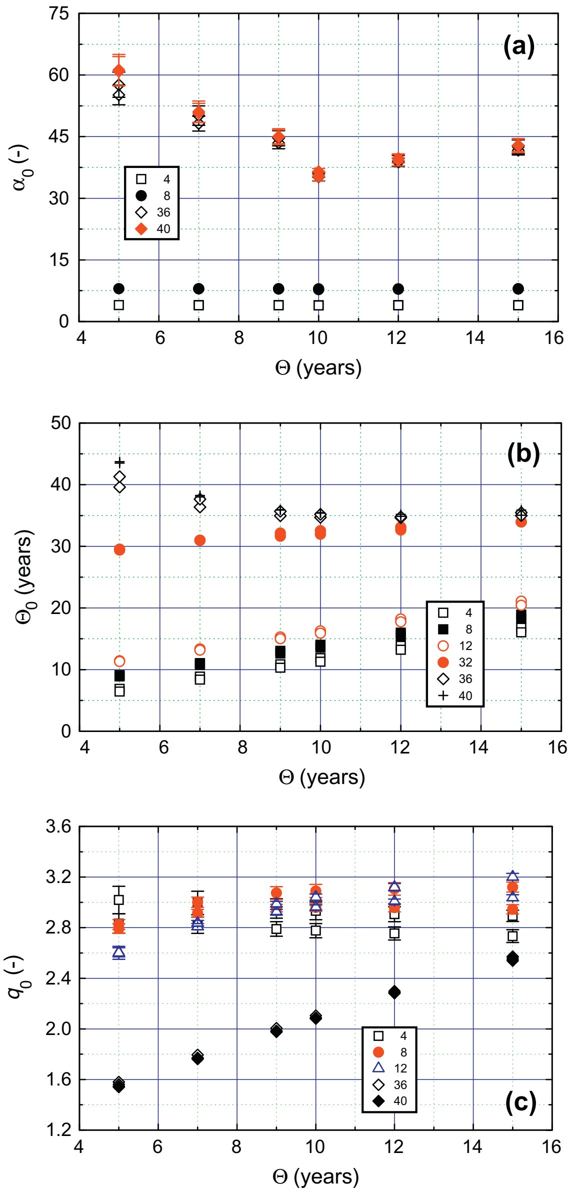

Fig. 4. Dependence of parameters: (a) α_0, (b) Θ_0 and (c) q_0 for the citations of progressively published papers in sets of two sequences of q_{1-4} for the citability of papers characterized by different values of publication duration t of hypothetical authors as a function of time constant Θ.

the initial Θ equal to 10 and 15 years. The value of the slope of the plots of Θ_0 on t in the later period increases with decreasing value of Θ and are about 2, 1 and 0.7 for Θ equal to 5, 10 and 15 years, respectively. However, only in the case of initial Θ = 10 years, the entire data follow a linear dependence: $\Theta_0 = 7.71 + 0.72t$, $r^2 = 0.993$.

In contrast to the behavior of α_0 and Θ_0 noted above, the exponent q_0 lies between a wide range of values and its dependence on t, shown in Fig. 3c, is relatively complex. Irrespective of the value of initial Θ used to generate the data, the value of q_0 for q_{1-4} in the sequence 1.5, 2, 3 and 5 is lower than that for q_{1-4} in the sequence 5, 3, 2 and 1.5 up to t = 28 years. The differences in the values of q_0 are relatively enormous for t < 24 years. However, for t > 28 years the two values of q_0 are practically the same. With an increase in t, q_0 increases initially and then, after passing through a maximum at a particular t, it decreases steeply first up to t = 36 years and then attains a constant value. The value of t (denoted by t_{max}) corresponding to the maximum value of q_0 depends on the initial Θ and is approximately equal to Θ. Moreover, the nature of the plots of q_0 on t for Θ equal to 5 and 15 years is different below and above t_{max}. For example, with respect to the data corresponding to Θ = 10 years, the values of q_0 corresponding to Θ = 5 years is higher than those corresponding to Θ = 15 years for $t < t_{max}$, but the order is reversed when $t > t_{max}$.

From Fig. 3a and b it may be noted that practically linear dependences of α_0 and Θ_0 on t during the entire publication period are observed when initial Θ = 10 years. In this case, $\alpha_0 \approx \alpha_{sum}(\max) = \sum \alpha_i(\max)$, $11.2 < \Theta_0 < 35.4$ and $2.08 < q_0 < 3.1$. This means that, with increasing number of successively published papers, $\Theta_0 > \Theta$ and q_0 takes values intermediate between the initially chosen values of 1.5 and 5 and the lowest value of 2.08 ± 0.04 is attained at t = 40 years The data of Fig. 1 are a special case of the above situations. In that case, $\Delta = 0$ and $n = 4$.

The relative complicated dependences of α_0, Θ_0 and q_0 for the citations of progressively published papers on the publication duration t of hypothetical authors for various values of time constant Θ presented in Fig. 3 become more informative when the data of α_0, Θ_0 and q_0 for low and high publication duration t are plotted as a function of time constant Θ as shown in Fig. 4. As seen from Fig. 4a, the dependence of α_0 on Θ can be splitted into two regions of t. For t below about 24 years, the value of α_0 is practically independent of Θ. However, for t > 28 years, with an increase in Θ the value of α_0 initially decreases, attaining a minimum value at $\Theta \approx 10$ years and then it increases. Fig. 4b shows that for t < 32 years the constant Θ_0 steadily increases with Θ, but for t = 36 and 40 years Θ_0 first decreases from $\Theta_0 > t$ and approaches a steady value of about 34 years above $\Theta \approx 10$ years. From Figs. 4a and 4b it may be seen that the dependences of α_0 and Θ_0 on Θ are somewhat similar. In a particular range of Θ, when α_0 increases with Θ, Θ_0 also increases. In contrast to the dependences of α_0 and Θ_0 on Θ, the dependence of q_0 on Θ is different, as seen from Fig. 4c. For t > 12 years, the value of q_0 increases with Θ; but the value of q_0 for t = 36 and 40 years is practically the same for a given Θ. The value of q_0 is practically independent of Θ for t = 8 years, but its value somewhat decreases with increasing Θ for t = 4 years.

From the above plots of α_0 and Θ_0 against predefined time constant Θ for different publication time t, shown in Fig. 4a and 4b, it may be seen that, corresponding to $\alpha_0 = t$, there is a particular value of $\Theta \approx 10$ years when Θ_0 also attains a constant value equal to α_0 (i.e. $\Theta_0 = t \approx 36$ years). Corresponding to this transition value of $\Theta \approx 10$ years, the value of the exponent q_0 is about 2 (see Fig. 4c). Thus, it may be concluded that, for the citations of the papers of an author publishing papers successively at a constant rate ΔN in his/her publication career t lasting over 40 years, there is an average time constant $\Theta \approx 10$ years when the exponent q_0 approaches a constant value of about 2. This means that in this particular case the publication career length $t \approx 4\Theta$.

It should be noted that there are two time constants Θ_0 and T in the plots of $\alpha_{sum}(t)$ against t. The former time constant Θ_0 corresponds to the time up to which the ascending part of the resultant $\alpha_{sum}(t)$ curve follows the linear dependence between Θ_0 and t (see Fig. 3b). The latter time constant T corresponds to the citation time t on a particular $\alpha_{sum}(t)$ curve when $\alpha_{sum}(t)$ approaches its maximum value $\alpha_{sum}(\max)$ as represented by the linear dependence between α_0 and citation time t (Fig. 3a). The time constant T is approximately equal to $3\Theta_0$ (see Fig. 2).

A linear dependence between α_0 and t with a slope of unity is expected for cumulative citations of papers where a new paper is published per year. Higher values of Θ_0, and accompanied with them lower values of q_0, for q_{1-4} in the sequence 1.5, 2, 3 and 5 than those for q_{1-4} in the sequence 5, 3, 2 and 1.5 are associated with higher citations produced by papers with the lower value of q_1 in the set of q_{1-4} of the four papers (see Fig. 1).

In the range of relative time t/Θ_0 when the approximation $e^x = 1 + x$ holds, Eq. (13) transforms to the power-law expression:

$$\alpha_{sum}(t) = \alpha_0 \left(\frac{t}{\Theta_0}\right)^{q_0}. \tag{14}$$

The practically linear increase in the time constant Θ_0 with increasing publication duration t for the cumulative citation fraction $\alpha_{sum}(t)$ from an initial extrapolated value of about 8 years at t = 0 to t = 40 years (see Fig. 3b) implies that the citation time t for cumulative citations increases when Eq. (12) transforms to the power-law relationship (14). In the range of long citation periods t, q_0 is the lowest, and for t > 36 years, q_0 is approximately 2.0 (cf. Fig. 3c). Moreover, since $q_0 \approx 2$ it may be seen that the approximation $e^x = 1 + x$ holds reasonably well with an error less than 20, 40 and 70% up to t/Θ_0 = 0.6, 0.8 and 1, respectively. This means that the power-law relation (14) may be used in a wide range of t/Θ_0. Thus, it may be concluded that there are three regions of the dependence of $\alpha_{sum}(t)$ on t: (1) $0 < t/\Theta_0 < 1$ when power-law relation (13) holds, (2) $1 < t/\Theta_0 < 3$ when $\alpha_{sum}(t)$ slowly deviates from the power-law relation and approaches a maximum value at $T \approx 3\Theta_0$, and (3) $t > 3\Theta_0$ when $\alpha_{sum}(t)$ remains independent of t.

It should be mentioned that the value of $q_0 = 2$ means that the term $\nu d = 1$ (cf. Eq. (8)). According to the PNM, this implies that dimensionality $d = 2$ when citation process is controlled by diffusion (i.e. $\nu = 1/2$) whereas $d = 1$ when citation process is controlled by surface reactions (i.e. $\nu = 1$). In the context of citations of papers published by an author, the former process may be attributed to the dissemination of the contents of the papers to the potential readers whereas the latter process to the absorption of the disseminated contents by the reader.

It may be noted that the physical model proposed in the present paper shows similarities with the model of "bursty and hierarchical structure in streams" developed by Kleinberg (2002). According to the latter model, the appearance of a topic in a document stream is indicated by a burst of activity with certain features rising sharply in frequency as the topic grows continuously over time and the document stream is modeled using an infinite-state automation, in which bursts appear naturally as state transitions. Document stream of this approach is the equivalent of receiving citations (i.e. items) over time by an author's individual paper (i.e. source) in the PNM used in this paper. Like the PNM (Sangwal, 2012a, 2012b), the approach proposed by Kleinberg also predicts that every document stream characterized by a topic, such as citations received by a paper, appears, grows in intensity for a certain period of time and finally fades away.

4. Predictions of progressive nucleation mechanism

Eq. (12) is the general expression which can easily be extended to analyze the cumulative citations $L_{sum}(t)$ of the papers of individual authors as functions of citation duration t. For this purpose, in Eq. (12) the ratio α_{sum}/α_0 may be considered as an average cumulative fraction α_{av} for the citations of papers published successively. Eq. (12) is exactly of the form of Eq. (6). Therefore, if $L_{sum}(t)$ is the sum of citations produced by successively published papers at time t, the ratio $\alpha_{sum}(t)/\alpha_0$ may be defined in terms of cumulative citations $L_{sum}(t)$ by the relation

$$\frac{\alpha_{sum}(t)}{\alpha_0} = \frac{L_{sum}(t)}{C_0}, \tag{15}$$

where C_0 is sum of the maximum numbers of citations from the collective of papers. Note that Eq. (15) follows from the definition of the fraction $\alpha(t)$ of cumulative citations $L(t)$ at time (t); see Eq. (6) where $\alpha_0 = 1$. Then Eq. (12) may be expressed as follows:

$$L_{sum}(t) = C_0\left[1 - \exp\left\{-\left(\frac{t}{\Theta_0}\right)^{q_0}\right\}\right], \tag{16}$$

Eq. (16) is the same as Eq. (6) but now the term $\alpha(t)$ of Eq. (6) represents $\alpha_{av}(t) = \alpha_{sum}(t)/\alpha_0$ and takes into consideration all successively published papers. According to Eq. (16), in the region $0 < t < \Theta_0$ when the approximation $e^x = 1 + x$ is roughly valid, the cumulative fraction $L_{sum}(t)$ of citations of cumulative N papers published by an author in his/her entire publication career limited to Θ_0, such that $\Theta_0 < T/3$, may be represented by (cf. Eqs. (15) and (16))

$$L_{sum}(t) = \lambda_0 \Delta N \cdot \frac{t^{q_0}}{\Theta_0^{q_0-1}}, \tag{17}$$

where the cumulative papers $N = \Delta N \cdot \Theta_0$ and λ_0 is a citability parameter relating $L_{sum}(t)$ to $\alpha_{sum}(t)$, i.e. $\lambda_0 = L_{sum}(t)/\alpha_{av}(t) = L_{sum}(\Theta_0)/\alpha_{av}(\Theta_0)$. Obviously, since $q_0 = 2$, in this region $L_{sum}(t)$ slowly increases following t^2 dependence. Moreover, the number of cumulative citations L_{sum} is directly proportional to the publication rate ΔN and citability constant λ_0 and inversely proportional to the time constant Θ_0. Note that here publication period is synonym with citation period for $t < \Theta_0$.

From Eq. (17) the following two parameters relating cumulative citations $L_{sum}(t)$ to the publication duration t may be introduced:

$$a = \frac{\lambda_0 \cdot \Delta N}{\Theta_0} = \frac{L_{sum}(t)}{t^2} = \frac{\upsilon_{sum}}{t}, \tag{18}$$

where

$$\upsilon_{sum} = \frac{L_{sum}(t)}{t} = \lambda_0 \cdot \Delta N\left(\frac{t}{\Theta_0}\right). \tag{19}$$

Here the citability proportionality constant $\lambda_0 = L_{sum}(\Theta_0)/\Delta N \cdot \Theta_0$. Note that the parameters a and υ_{sum} denote acceleration and velocity of growth of citations of an author and are the analogs of linear acceleration and linear velocity in kinematics. Eq. (19) represents the time dependence of υ_{sum} in the time interval between t_0 and Θ_0, where t_0 is a small time constant beyond which the time constant Θ_0 increases linearly with publication duration t (see Fig. 3b).

It may be noted that the time dependence of the citations of our hypothetical author predicted by Eq. (17) of the PNM is in agreement with the predictions of the deterministic model of Hirsch (2005) and the stochastic model of Burrell (2007a, 2007b). In the case of the deterministic model, from Eqs. (1) and (2) one finds the citation acceleration

$$a = Ab^2. \tag{20}$$

Comparison of this equation with Eq. (18) suggests that, for the given time constant Θ_0 for the citations of the papers of an author, the proportionality constant Ab^2 is directly proportional to the citability-related parameter λ_0 of papers and the publication rate ΔN. According to the stochastic model the citation acceleration (see Eq. (3))

$$a = \frac{\Delta N \cdot \Delta L}{2}. \tag{21}$$

This relation is very similar to that described by Eq. (18), implying that the average citation rate $\Delta L = 2\lambda_0/\Theta_0$. Thus, it can be concluded that the basic concepts of our approach are sound.

Eq. (17) contains three parameters, viz. the citability parameter λ_0, the publication rate ΔN and the time constant Θ_0, which determine the citation behavior of the publication activity of an author in his/her publication career. Since the time constant Θ_0 for the predefined time $\Theta \approx 10$ years used for modeling is essentially proportional to the publication career t of an author (see Figs. 3b and 4b), it may be concluded that the value of the λ_0/Θ_0 ratio is determined, among others, by the citation culture of papers in a discipline, whereas the publication rate ΔN of the papers is an author-related factor. In fact, such effects on the citations of scientific papers are well known (for example, see: Albarran & Ruiz-Castllo, 2011; Vieira & Gomes, 2010; Wu & Wolfram, 2011). Therefore, if an author remains active in a particular scientific field and publishes papers at a constant rate ΔN, Eq. (18) of the citation acceleration a is a useful scientometric age-independent measure to analyze and compare the scientific activities of different authors.

The above predictions of the constancy of citation acceleration a according to Eq. (17) can be tested by analyzing the dependence of cumulative citations $L_{sum}(t)$ on citation duration t of papers published successively by different authors. Such data are easily accessible from websources such as Google Scholar, Elsevier's Scopus and Thomson Reuters' ISI Web of Knowledge (Web of Science).

5. Application of Eq. (17) to the citation data of selected authors

We analyzed the expected constancy of citation acceleration a of Eq. (17) from the dependence of cumulative citations $L_{sum}(t)$ on citation duration t of papers published successively by six arbitrarily selected Polish authors. The six authors considered are: M. Kosmulski (MK), K. Sangwal (KS), S. Krukowski (SK), G. Gładyszewski (GG), W. Stępniewski (WS) and Z.R. Żytkiewicz (ZZ). The data on the $\Delta L_{sum}(t)$ against t for these authors have been reported previously (Sangwal, 2011b) and were collected from Thomson Reuters' ISI Web of Knowledge (Web of Science) and covered the period up to 2010. The cumulative citations $L_{sum}(t)$ were calculated from the reported data (see also Table 4)

Fig. 5 presents the data of cumulative citations $L_{sum}(t)$ of the papers of the above authors against t. The data were analyzed in the entire citation duration according to Eq. (16). It was observed that the data are better described by Eq. (16) when the first year of citation was chosen during the best fitting of the data. However, there were situations when the data could equally be fitted with two or more sets of the constants C_0, Θ_0 and q_0. Some examples are given in Table 2. The curves shown in Fig. 5 were drawn with the values of C_0, Θ_0 and q_0 indicated by arrows in Table 2. The curves indeed suggest that Eq. (16) represents the citation data of different authors reasonably well. An easily recognized exception in this regard are the data of KS, where the fitting curve lies below and above the citation data between 1980 and 1995 and between 1995 and 2004, respectively.

From Table 2 one finds that the value of the time constant Θ_0 lies between 15 and 24 years for SK, GG and ZZ and between 30 and 90 years for the other authors. These values of Θ_0 are two to ten times higher than the average value of the time constant Θ for the citations of individual papers of an author (cf. Sangwal, 2012b). This observation is consistent with the finding of resultant citation behavior of collectives of papers characterized by different sets of time constant Θ and exponent q (see Section 3). In contrast to the behavior of Θ_0 for collectives of papers, the value of exponent q_0 is 1.35 for GG, about 2.0 for WS, and between 2.8 and 3.9 for the remaining authors. According to the analysis of the modeling of cumulative citations of papers one expects that q_0 is about 2 for the citations of a collective of papers (see Fig. 3c). This value of q_0 equal to about 2 is observed only for WS. The values of q_0 deviating from 2 are a consequence of citations originating from papers characterized by enormously different citation parameters Θ_0 and q_0 in two or more citation periods during the publication career of an author. This point is discussed below.

A careful examination of the citation data for different authors reveals that the best fit is obtained in the entire citation duration only in the case of GG and WS (see Fig. 5). However, in the initial citation duration of practically all other authors one observes the best-fit curves either below (e.g. MK and ZZ) or above the citation data (e.g. SK). These observations are associated with different citabilities of the papers published by different authors in the initial and later periods of their publication careers. To illustrate this idea we reexamine in more detail the cumulative citation data of MK and KS.

In view of the fact that the time constant Θ_0 for the cumulative citations $L_{sum}(t)$ of a collective of papers increases with the publication career t, it is expected that the citation data of an author follows power-law relation when Θ_0 is relatively high. Therefore, the $L_{sum}(t)$ data of MK and KS were fitted in two different citation regions with appropriately selected values of Y_0 according to the power-law relation

$$L_{sum}(t) = at^{q_0} = a(Y - Y_0)^{q_0}, \tag{22}$$

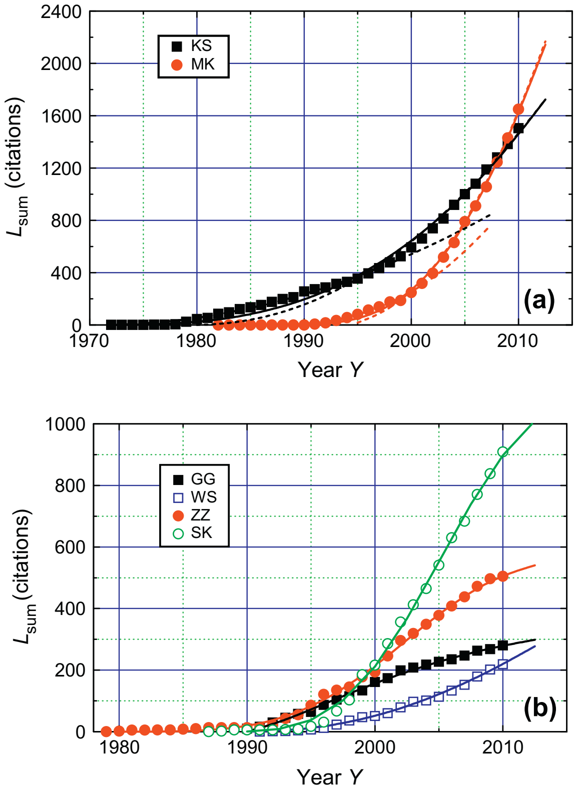

Fig. 5. Plots of cumulative citations $L_{sum}(t)$ against citation duration t of different authors: (a) KS and MK, and (b) SK, GG, WS and ZZ. Solid curves are drawn according to Eq. (11) with the values of C_0, Θ_0 and q_0 indicated by arrows in Table 2. In (a) dashed curves are drawn according to Eq. (21) with the values of a and q_0 indicated by arrows in Table 3.

Table 2
Values of constants of Eq. (16) for different authors.

Author	Y_0 (year)		C_0 (cites)	Θ_0 (years)	q_0
K. Sangwal (KS)	1971		5194	58.07	2.929
		→	16586	89.77	2.863
M. Kosmulski (MK)	1988		5341	30.02	3.284
		→	9209	36.88	3.154
S. Krukowski (SK)	1988		1118	19.06	3.374
G. Gładyszewski (GG)	1990		367	15.34	1.357
W. Stępniewski (WS)	1990		1734	51.29	2.022
		→	1453	45.53	204
Z. Żytkiewicz (ZZ)	1980		564	24.21	3.925

which follows from Eq. (17), where $a = C_0/\Theta_0^{q_0}$ and $t = (Y - Y_0)$. The best-fit values of the constants a and q_0 in different citation regions for the authors are listed in Table 3. The dashed curves in Fig. 5a show the best-fit plots in two different regions according to Eq. (22) with the values of a and q_0 indicated by arrows in Table 3.

Comparison of the dashed curves, drawn according to Eq. (22) with the values of a and q_0 listed in Table 3, with the solid curves, drawn according to Eq. (16) with the values of C_0, Θ_0 and q_0 given in Table 2, in Fig. 5a reveals that higher values of q_0 obtained from the best-fit of the citation data according to Eq. (16) in the entire citation duration are associated with the inclusion of the data of the initial stage in the analysis. However, it may be seen from Table 3 that q_0 is approximately 2 in the initial and later stages of citations of both of the authors analyzed here. The main change lies in the value of a which is higher at the later stage. A higher value of a in the later stage is associated with higher citability of the papers published by these authors during this period.

In order to establish trends of the citation behavior during their publication careers, the citation data of all of the authors mentioned above were examined again using Eq. (22) with $q_0 = 2$. When $q_0 = 2$, Eq. (22) predicts a linear dependence between $L_{sum}^{1/2}$ and Y with slope $a^{1/2}$. Fig. 6 shows the plots of $L_{sum}^{1/2}$ against citation year Y for different authors. It may be seen from these plots that an approximately linear dependence between $L_{sum}^{1/2}$ and Y is observed only in the case of WS. In all other

Table 3
Constants a and q_0 of Eq. (22).

Author	Data		Y_0 (year)	a	q_0
K. Sangwal	<1995	→	1971	0.72 ± 0.15	1.964 ± 0.071
	>1992	→	1980	1.07 ± 0.17	2.126 ± 0.050
	>1992		1982	2.37 ± 0.39	1.930 ± 0.052
M. Kosmulski	<2000		1985	0.105 ± 0.036	2.870 ± 0.131
	<2000	→	1987	0.43 ± 0.14	2.479 ± 0.136
	>1998		1992	1.96 ± 0.19	2.328 ± 0.187
	>1995	→	1993	4.12 ± 0.67	2.109 ± 0.061

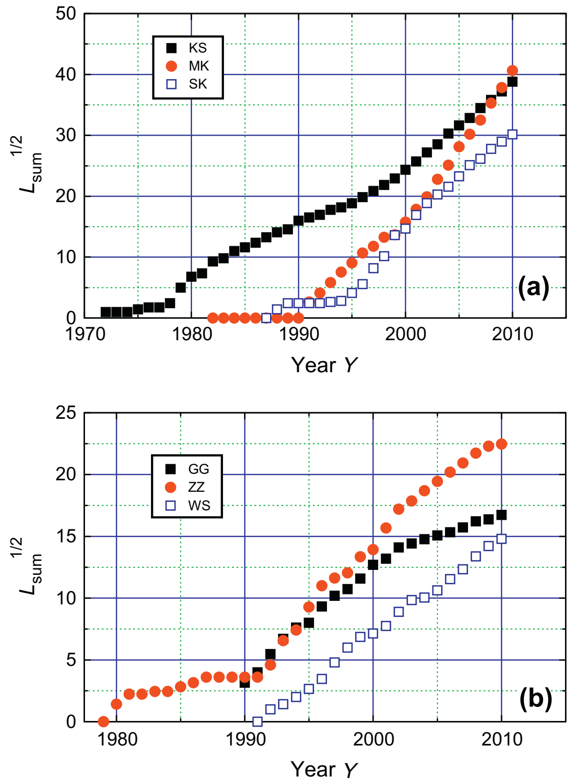

Fig. 6. Plots of $L_{sum}^{1/2}$ against citation year Y of different authors: (a) KS, MK and SK, and (b) GG, WS and ZZ.

cases, there are two or more regions of linear dependences. In the case of MK for example, after an initial inert citation period between 1982 and 1990, there are two regions of linear increase: between 1990 and 2001 and between 2001 and 2010.

From the above, it may be concluded that analysis of growth of cumulative citations of various authors in their publication careers using Eq. (17) reveals different periods of citations. Since Eq. (17) is based on the concept of stationary nucleation involving constant values of citability parameter λ_0 and the publication rate ΔN for an author, different citation periods during her/her publication career may be attributed to changes in these parameters.

It is a common observation that the publication rate ΔN of an author fluctuates enormously in successive years during his/her publication career and even the average values of ΔN over various decades are not constant. Fig. 7 shows the dependence of average values of ΔN per five years (ΔN_{av}) on the publication year for the authors of Fig. 6. It may be seen that ΔN_{av} decreases linearly with career length for GG, increases linearly with career lengths for MK and SK except for the initial period, and is practically career independent for WS. In contrast to the above cases, it shows wavy nature for the $\Delta N_{av}(Y)$ data of KS and ZZ. Therefore, it may be suspected that $L_{sum}(t)$ and $\Delta N_{av}(t)$ are interrelated.

The interdependence between ΔN and a may be established from Eq. (18) rewritten in the form

$$\left(\frac{L_{sum}(t)}{\Delta N}\right)^{1/2} = \left(\frac{\lambda_0}{\Theta_0}\right)^{1/2} \cdot t = \left(\frac{a}{\Delta N}\right)^{1/2} \cdot t. \tag{23}$$

According to this relation, $(L_{sum}/\Delta N)^{1/2}$ increases linearly with t for an author, with slope $(a/\Delta N)^{1/2} = (\lambda_0/\Theta_0)^{1/2}$. Two or more linear parts of the plots of $(L_{sum}/\Delta N)^{1/2}$ against t for an author are indicators of periods of different citability during his/her publication career. Plots of $(L_{sum}/\Delta N)^{1/2}$ against t for the authors discussed above are shown in Fig. 8. As seen from Fig. 8, there is indeed a linear dependence of $(L_{sum}/\Delta N)^{1/2}$ on t for MK and GG, with slope $(a/\Delta N)^{1/2} = 0.63$. Probably there is also a linear dependence for the data of WS. In other cases, the dependence is relatively complex, but linear dependence with a constant slope may also be noted for these authors if the initial and/or later career lengths are excluded from the analysis. For example, in the case of KS if the last point is omitted during the analysis, one finds that $(L_{sum}/\Delta N)^{1/2}$ increases linearly with his publication career t (with slope 0.41).

Finally, it should be mentioned that the citation acceleration a of Eq. (21) is related to the traditionally-defined impact factor IF used for journals:

$$IF = \frac{L_{sum}}{N} = \frac{a}{\Delta N}t, \tag{24}$$

where L_{sum} and N now denote the total number of citations and papers published by an author during t years, respectively. Since IF of a journal for a given year Y is defined with reference to the total number L_{sum} of citations received in the year Y by the number N of papers published in the years $(Y - 1)$ and $(Y - 2)$, relation (24) holds for $t \geqslant 3$ years. The impact factor IF defined by Eq. (24) now refers to a given author.

6. Analysis of ranking of selected authors according to their citation data

It has been found that the so-called least-squares h rate obtained from h sequences of authors at time t (Burrell, 2007c), their average of decade-based h index (Kosmulski, 2009; Abt, 2012) and the raw h rate at time t (Burrell, 2007, 2012; Sangwal, 2012c) are practically independent of their career lengths. This finding suggests that, despite fluctuations in the values of ΔN and λ_0 of an author during his/her publication career spanning over t years, one can use Eq. (17) with $q_0 = 2$ to analyze and compare the overall citation behavior of various authors. In this case, citation acceleration a described by Eq. (18) as a measure to establish the ranking of the scientific impact of different authors.

In this study, for the analysis we used the citation data of eight Polish professors selected arbitrarily by the present author (Table 4) and six scientists elected to membership of the Royal Society in 2006 (Table 5), randomly chosen by Anderson et al. (2008). The data of the Polish professors have previously been published (Sangwal, 2011b, 2012b) whereas those of the Royal Society scientists have been given by Anderson et al. (2008). The authors are listed in the order of decreasing h index. The

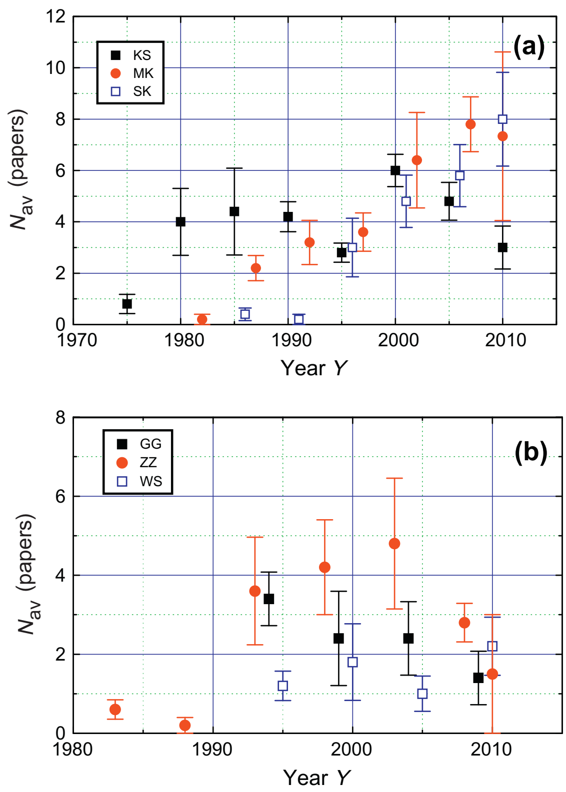

Fig. 7. Dependence of average values of ΔN per five years on the publication years Y for different authors: (a) KS, MK and SK, and (b) GG, WS and ZZ.

Table 4
Activity parameters of selected Polish authors.

Author	N (t)	ΔN	L	h	$b = h/t$	$A = L/h^2$	$a = L/t^2$
T. Dietl (TD)	289 (36)	8.03	10278	41	1.14 (1)	6.114 (1)	7.931 (1)
J. Barnaś (JB)	286 (28)	10.21	2922	30	1.07 (2)	3.247 (6)	3.727 (2)
M. Kosmulski (MK)	138 (33)	4.18	1759	23	0.70 (3)	3.325 (5)	1.615 (3)
K. Sangwal (KS)	152 (40)	3.80	1487	20	0.50 (5)	3.718 (3)	0.928 (5)
S. Krukowski (SK)	103 (29)	3.55	909	17	0.59 (4)	3.145 (7)	1.081 (4)
Z. Żytkiewicz (ZZ)	85 (32)	2.66	505	12	0.38 (7)	3.507 (4)	0.493 (8)
G. Gładyszewski (GG)	53 (22)	2.41	281	10	0.45 (6)	2.810 (8)	0.581 (7)
W. Stępniewski (WS)	31 (20)	1.55	219	7	0.35 (8)	4.469 (2)	0.548 (6)

values of raw h rate (equal to constant b), Hirsch constant A, and citation acceleration a for the above scientists are included in the tables.

The Hirsch index h is defined as the highest number of papers of an author that received h or more citations. The set of h most cited papers is also said to form the so-called h-core (for example, see: Burrell, 2007a; Franceschini & Maisano, 2010b) which contains h^2 citations, i.e. $L_{sum} = h^2$. This means that the Hirsch constant $A = 1$ when h is equal to the number N of papers. When $A = 2$, h^2 citations are outside the h-core. Thus, higher the value of A, the increasing number of citations lies outside the h-core.

With reference to Tables 4 and 5 one notes that the h index is not a suitable measure to compare the impact of the scientific output of different authors of different career lengths t and different publication rates ΔN. For example, both Jackson and Badford published the same number of papers and have $h = 44$, but the latter achieved his goal during a short publication career. Similarly, Lockwood and Becke have roughly similar h indexes and similar career lengths t, but the latter published three-times less papers than the former. The h rate eliminates the effect of publication career t and publication rate ΔN, which results in changes in the ranks of some of the authors. In the case of Polish scientists for example, the ranks of SK and GG have improved whereas those of KS and ZZ have deteriorated. Among the "Royal Society" scientists, Jackson ranked 1 according to the h index has tumbled down to 4 according to the h rate.

The shortcoming of the h rate during the comparison of the scientific acitivities of different authors is that, like the original h index, it does not take into account ($L_{sum} - h^2$) citations outside the h-core. For example, Becke with $h = 35$ has received citations five times higher than Jackson with $h = 44$, but Becke is at rank 4 in comparison with Jackson at rank 1.

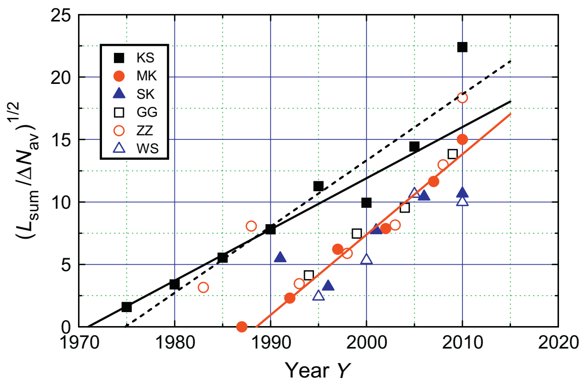

Fig. 8. Plots of $[L_{sum}(t)/\Delta N]^{1/2}$ against citation year Y of different authors: (a) KS, MK and SK, and (b) GG, WS and ZZ. As a guide, linear plots are drawn for KS and MK. KS (dashed plot): $y = -1023 + 0.518x$ ($r^2 = 0.9496$); KS (solid plot, without last point): $y = -808.1 + 0.41x$ ($r^2 = 0.9749$); MK (solid plot): $y = -1251 + 0.629x$ ($r^2 = 0.9906$).

Table 5
Activity parameters of authors selected by Anderson et al. (2008).

Author	N (t)	ΔN	L	h	$b = h/t$	$A = L/h^2$	$a = L/t^2$
R.J. Jackson	79 (36)	2.194	10778	44	1.22 (4)	5.567 (2)	8.316 (3)
D. Badford	78 (20)	3.90	6281	44	2.20 (1)	3.244 (6)	15.703 (2)
M. Lockwood	176 (25)	7.04	5101	39	1.56 (2)	3.354 (5)	8.162 (4)
A.D. Becke	55 (28)	1.964	40,094	35	1.25 (3)	32.730 (1)	51.140 (1)
H.R. Saibil	80 (30)	2.667	4234	33	1.10 (5)	3.888 (3)	4.704 (5)
M.R.E. Proctor	89 (31)	2.871	2356	26	0.84 (6)	3.485 (4)	2.452 (6)

Consequently, in the case of Becke the high ($L_{sum} - h^2$) citations outside the h-core have resulted in his exceptionally high Hirsch constant A equal to 32.7, although its value lies between 2.8 and 6 for most of the authors.

The citation acceleration a based on cumulative citations L_{sum} of authors of different career lengths is devoid of the above shortcomings of the h index and h rate and is a useful scientometric measure to compare the scientific impact of different authors during their scientific careers (Sangwal, 2012c, 2012d). Moreover, the value of a for different authors covers a wide range and it is easy to calculate it. For example, for the authors listed in Tables 4 and 5, it lies between 0.5 and 51. This wide range makes it very sensitive to even small changes in ΔN and t of the authors. This changes the rankings of some of the authors significantly. For example, among the Polish scientists, ranks of four of them have changed. Similarly, among the six Royal Society scientists, Becke, Lockwood and Jackson have changed their ranks to 1, 4 and 3 from previous h-based ranks 4, 3 and 1, respectively. However, it should be noted that, except for ranking based on h index, all measures indicate that Becke is the leader.

Finally, it should be mentioned that the present author (Sangwal, 2012d) discussed previously the time dependence of citation acceleration $a = L(t)/t^2$ using Eq. (3) due to Burrell (2007c) and introduced the concept of the ratio $R(t)/t$, defined as R rate, where R is the radius of a circle of mesh area equal to L citations. From Eq. (18) one obtains the relation for the R rate in the form

$$\frac{R(t)}{t} = \left(\frac{a}{\pi}\right)^{1/2} = \left(\frac{\lambda_0 \cdot \Delta N}{\pi \Theta_0}\right)^{1/2}, \tag{25}$$

implying that R-rate for an author is also a constant. Since $R \approx h$ for a majority of authors (Sangwal, 2012d), similar trends are expected for R and h rates. However, in contrast to the trends of h rate, it is relatively easily to follow and compare the trends of R rates of different authors.

7. Conclusions

Modeling the growth behavior of cumulative citations $L_{sum}(t)$ of cumulative N papers published by a hypothetical author in his/her entire publication career spanning over t years according to the progressive nucleation mechanism (PNM) reveals that, for all nonzero values of constant publication rate ΔN, the time dependence of $L_{sum}(t)$ may be represented in two distinct regions: (1) in the region $0 < t < \Theta_0$, $L_{sum}(t)$ slowly increases approximately proportional to the square of citation time t (see Eq. (17)), and (2) in the region $t > T$, $L_{sum}(t)$ approaches a constant $L_{sum}(\max)$ at T. The former prediction of the PNM is in agreement with the predictions of the deterministic model of Hirsch (2005) and the stochastic model of Burrell, 2007.

Eq. (17) of the time dependence of $L_{sum}(t)$ contains three parameters, viz. the citability parameter λ_0, the publication rate ΔN and the time constant Θ_0, which determine the citation behavior of the publication activity of an author in his/her publication career t which is essentially equal to the time constant Θ_0. It is concluded that the value of the λ_0/Θ_0 ratio is determined, among others, by the citation culture of papers in a discipline, whereas the publication rate ΔN of the papers is an author-related factor. Therefore, if an author remains active in a particular scientific field and publishes papers at a constant rate ΔN, from Eq. (17) a new useful scientometric age-independent measure to analyze and compare the scientific activities of different authors in the form of citation acceleration a described by Eq. (18) is suggested. The definition of citation acceleration a also follows from the two relationships (20) and (21) based on the deterministic model of Hirsch (2005) and the stochastic model of Burrell (2007a, 2007b), respectively.

Confrontation of the time dependence of cumulative number $L_{sum}(t)$ of citations of papers published by different authors with the theoretical equation expressed as $L_{sum}^{1/2}(t) = a^{1/2}t$ (cf. Eq. (22)) reveals one or more citation periods during their publication careers. It is suggested that plots of $(L_{sum}(t)/\Delta N)^{1/2}$ against t (cf. Eq. (23)) should be used to analyze the effects of λ_0 and ΔN for an author.

Acknowledgement

The author is indebted to one of the referees for pointing out similarities between the physical model proposed in this paper and the model of ‘‘bursty and hierarchical structure in streams’’ proposed by Jon Kleinberg.

References

Abt, H. A. (2012). A publication index that is independent of age. *Scientometrics, 91*(3), 863–868.

Albarran, P., & Ruiz-Castllo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology, 62*(1), 40–49.

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: An review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics, 3*(4), 273–280.

Anderson, T. R., Hankin, R. K. S., & Killworth, P. D. (2008). Beyond the Durfee square: enhancing the h-index to score total publication output. *Scientometrics, 76*(3), 577–588.

Burrell, Q. L. (2007a). Hirsch's h-index: a stochastic model. *Journal of Informetrics, 1*(1), 16–25.

Burrell, Q. L. (2007b). On the *h*-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics, 1*(2), 170–177.

Burrell, Q. L. (2007c). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. *Scientometrics, 73*(1), 19–28.

Burrell, Q. L. (2009). On Hirsch's h, Egghe's g and Kosmulski's h(2). *Scientometrics, 79*(1), 79–91.

Burrell, Q. L. (2012). Comments on ''A publication index that is independent of age'' by Abt. *Scientometrics, 91*(3), 1059–1060. http://dx.doi.org/10.1007/s11192-011-0605-5.

Csajbók, E., Berhidi, A., Vasas, L., & Schubert, A. (2007). Hirsch index for countries based on essential science indicators data. *Scientometrics, 73*(1), 91–117.

Egghe, L. (2007). Dynamic h-index: The Hirsch index in function of time. *Journal of the American Society for Information Science and Technology, 58*(3), 452–454.

Egghe, L. (2008). Modelling successive h-indices. *Scientometrics, 67*(3), 491–502.

Egghe, L. (2009). Mathematical study of h-index sequences. *Information Processing and Management, 45*(3), 288–297.

Egghe, L. (2010a). Characteristic scores and scales based on h-type indices. *Journal of Informetrics, 4*(1), 14–22.

Egghe, L. (2010b). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology, 44*, 65–114.

Egghe, L. (2010c). Conjugate partitions in informetrics: Lorenz curves, h-type indices, Ferrar graphs and Durfeee squares in a discrete and continuous setting. *Journal of Informetrics, 4*(3), 320–330.

Egghe, L., & Ravichandra Rao, I. K. (1992). Classification of growth models based on growth rates and its applications. *Scientometrics, 25*(1), 5–46.

Egghe, L., Ravichandra Rao, I. K., & Rousseau, R. (1995). On the influence of production on utilization functions: obsolescence or increased use? *Scientometrics, 34*(2), 285–315.

Franceschini, F., & Maisano, D. (2010a). The Hirsch spectrum: A novel tool for analyzing scientific journals. *Journal of Informetrics, 4*(1), 64–73.

Franceschini, F., & Maisano, D. (2010b). The citation triad: An overview of a scientist's publication output based on Ferrers diagrams. *Journal of Informetrics, 4*(4), 503–511.

Glänzel, W. (1997). On the possibility and reliability of predictions based on stochastic citation processes. *Scientometrics, 40*(3), 481–492.

Glänzel, W. (2004). Towards a model for diachronous and synchronous citation analysis. *Scientometrics, 60*(3), 511–522.

Glänzel, W. (2008). On some bibliometric applications of statistics related to the h-index. *Scientometrics, 77*(1), 187–196.

Glänzel, W., & Schubert, A. (2010). Hirsch-type characteristics of the tail of distributions. The generalized h-index. *Journal of Informetrics, 4*(1), 118–123.

Gupta, U. (1999). Obsolescence of physics literature: exponential decrease of the density of citations to Physical Review articles with age. *Journal of the American Society for Information Science and Technology, 41*(4), 282–287.

Gupta, B. M., Kumar, S., Sangam, S. L., & Karisiddappa, C. R. (2002). Modeling the growth of world social science literature. *Scientometrics, 53*(1), 161–164.

Gupta, B. M., Sharma, L., & Karisiddappa, C. R. (1995). Modeling the growth of papers in a scientific speciality. *Scientometrics, 33*(2), 187–201.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102*(46), 16569–16572.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin, 52*(6), 855–863.

Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In *KDD '02: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery (ACM), pp. 91–101.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original h-index. *ISSI Newsletter, 2*(3), 4–6.

Kosmulski, M. (2009). New seniority-independent Hirsch-type index. *Journal of Informetrics, 3*(4), 341–347.

Liang, L. (2006). H-index sequence and h-index matrix: Constructions and application. *Scientometrics, 69*(1), 153–159.

Nair, G. M., & Turlach, B. A. (2012). The stochastic *h*-index. *Journal of Informetrics, 6*(1), 80–87.

Naranan, S. (1970). Bradford's law of bibliography of science. an interpretation. *Nature, 227*, 631–632.

Navon, D. (2009). The h-index: A proposed new metric of individual scientific output. Cybermetrics, 13, paper 3.

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. *Current Science, 91*(1), 10.

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

Ravichandra Rao, I. K., & Srivastava, D. (2010). Growth of journals, articles and authors in malaria research. *Journal of Informetrics, 4*(1), 249–256.

Sangwal, K. (2011a). Progressive nucleation mechanism and its application to the growth of journals, articles and authors in scientific fields. *Journal of Informetrics, 5*(4), 529–536.

Sangwal, K. (2011b). On the growth of citations of publication output of individual authors. *Journal of Informetrics, 5*(4), 554–564.

Sangwal, K. (2012a). Progressive nucleation mechanism for the growth behavior of items and its application to cumulative papers and citations of individual authors. *Scientometrics, 92*(3), 575–591.

Sangwal, K. (2012b). Application of progressive nucleation mechanism for the citation behavior of individual papers of different authors. *Scientometrics, 92*(3), 643–655.

Sangwal, K. (2012c). On the age-independent publication index. *Scientometrics, 91*(3), 1053–1058.

Sangwal, K. (2012d). On the relationship between citations of publication output and Hirsch index *h* of authors: conceptualization of tapered Hirsch index h_T, circular citation area radius *R* and citation acceleration *a*. *Scientometrics, 93*(3), 987–1004.

Schubert, A. (2007). Successive h-indices. *Scientometrics, 70*(1), 201–205.

Van Raan, A. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics, 67*(3), 491–502.

Vieira, E. S., & Gomes, J. A. N. F. (2010). Citations of scientific articles: Its distribution and dependence on article feature. *Journal of Informetrics, 4*(1), 1–13.

Wong, C.-Y., & Goh, K.-L. (2010). Growth behavior of publications and patents: A comparative study on selected Asian economies. *Journal of Informetrics, 4*(2), 460–474.

Wu, Q., & Wolfram, D. (2011). The influence of effects and phenomena on citations: a comparative analysis of four citation perspectives. *Scientometrics, 89*(1), 245–258.

Ye, F. Y., & Rousseau (2008). The power law model and total career h-index sequences. *Journal of Informetrics, 2*(4), 288–297.