TOWARDS A MODEL FOR DIACHRONOUS AND SYNCHRONOUS CITATION ANALYSES

WOLFGANG GLÄNZEL

Katholieke Universiteit Leuven, Steunpunt O&O Statistieken, Dekenstraat 2,
B-3000 LEUVEN, Belgium

and

Hungarian Academy of Sciences, Institute for Research Organisation, Nádor u. 18,
H-1051 BUDAPEST, Hungary

Abstract

This paper gives an overview of the *diachronous* (*prospective*) and *synchronous* (*retrospective*) approach to ageing studies of scientific literature from the perspective of technical reliability, visualising the different aspects that can be analysed by the two approaches. The main objective is to deepen the understanding of the mechanism and the theory underlying the two aproaches, and is to show that the difference between the *diachronous* and *synchronous* model is not “just counting into opposite directions”. In this context, a stochastic model is presented showing that one and the same model can be used to describe both *diachronous* and *synchronous* perspectives of citation processes.

On the basis of this model, it is explained how some *diachronous* and *synchronous* citation-based indicators can be re-calculated for changing publication periods and citation windows underlying their construction. The paper is concluded by several applications such as the definition and calculation of *diachronous* (*prospective*) and *synchronous* (*retrospective*) journal impact measures and other citation indicators used in research evaluation.

Introduction

Recently, a number of bibliometric studies related with modelling citation age data have been published. These studies were concerned with the analyses of the age distribution of reference literature (for instance, *Wallace*, 1986), of citations to a given set of papers (e.g., *Glänzel* and *Schoepflin*, 1994) or of first-citation distributions (e.g., *Glänzel*, 1992, *Rousseau*, 1994, *Egghe*, 2000, *Burrell*, 2001). In these studies, two basic approaches have been distinguished, namely, the *diachronous* and the *synchronous* model. The *diachronous* approach is concerned with the use of a given set of publications in successive years, whereas *synchronous* studies proceed from the present to the past.

In some recent publications, *Burrell* (2001, 2002) has used the terms *retrospective* and *prospective* citation studies, where the terms ‘retrospective’ and ‘prospective’ are practically synonyms for ‘synchronous’ and ‘diachronous’, respectively.

Wolfgang Glänzel, *Towards a model for diachronous and synchronous citation analyses*, In: Guohua Jiang, Ronald Rousseau, Yishan Wu (Eds), Proceedings of the 9th International Conference on Scientometrics and Informetrics – ISSI 2003, Dalian University of Technology, Dalian, China, 2003, 68-78.

First *Wallace* (1986) has studied ageing of scientific literature. In particular, he analysed, the relationship between journal productivity and obsolescence, and assumed an exponential distribution. In his model, ageing is analogously related to the radioactive decay characterised by the "half-life" being the median of the distribution. Wallace's study is based on the *synchronous* (*retrospective*) approach, that is, he analysed the age of reference literature. From the pragmatic point of view, one can say that synchronous analyses are easier to conduct since it does not require the observation of citations in a quite large citation window of ten, fifteen or even more years. Nevertheless, the synchronous approach cannot serve as a substitute for *diachronous* (*prospective*) studies since the two approaches shed light on quite different aspects of citation processes, in general, and of ageing, in particular.

Although most *synchronous* (*retrospective*) ageing studies are based on the analysis of references in selected papers, *synchronous* studies can also be concerned with the analysis of citation received by publication sets. In the latter case, the citation window is fixed and the publication year or period is variable. Thus, both the Citing and the Cited Journal Package in the annual up-dates of the Journal Citation Reports (JCR) have, for instance, to be considered *synchronous* citation approaches.

The terms *synchronous* and *diachronous* are nowadays also used in a more general context of citation analyses that are not directly related to ageing. *Ingwersen* et al. (2000, 2001) have introduced a distinction between *synchronous* (*retrospective*) and *diachronous* (*prospective*) impact measures for scientific journals. According to their approach, the 'Garfield Impact Factor' produced by the Institute for Scientific Information is a synchronous Impact Factor since the citation year is fixed and the two-year publication period lie in the "past". In the above-mentioned studies, *Ingwersen* et al. have given a methodological discussion why the 'diachronous' approach should be preferred to the 'synchronous' one. The fact that only diachronous (*prospective*) impact measures can be calculated for non-serials is one of the advantages. The journal impact measures built since 1995 at LHAS in Hungary and at RASCI in Germany may serve as examples for such diachronous impact measures; here the publication year is fixed and citations are counted in a three-year observation period, namely, in the year of publication and the two subsequent years (for instance, *Glänzel*, 1996).

In the following, I will give a brief overview of ageing from the perspective of technical reliability visualising the different aspects that can be analysed by the two approaches. In the subsequent section a stochastic model will be given showing that one and the same model can be used to describe both *diachronous* (*prospective*) and *synchronous* (*retrospective*) perspectives of citation processes.

Diachronous and synchronous ageing studies in the context of technical reliability

Statistical functions provided by *synchronous* (*retrospective*) studies depend on too many factors to be uniquely interpreted in terms of ageing alone. In order to illustrate this effect I will refer to a simple model adopted from technical reliability processes (see, for example, *Watson* and *Wells*, 1961, *Gupta*, 1981). This may help to interpret the two approaches. It should be stressed at once that the technical reliability model can not be applied to explain information processes, moreover, basic principles such as *lack of memory property* (technical reliability) and *cumulative advantage principle* (bibliometric processes) can be regarded as almost diametrical. The basic questions concerning *synchronous* (*retrospective*) and *diachronous* (*prospective*) approach to obsolescence are, however, very similar in informetrics and technical reliabilty.

Within the framework of technical reliability analyses, function and lifetime of a system, a machine, a device or equipment is studied. The equipment may, for example, consist of machines in factories or of automobiles, but it might also be a simple device such as a rubber tyre or an ordinary electric switch. The usual definition of reliability of a system or a device is the probability that it will give satisfactory performance of its intended function for a specified period under specific operating conditions (see, *Juran* and *Gryna*, 1988). According to *Juran* and *Gryna*, this definition has four key elements.

1. The quantification of reliability in terms of probability;
2. A statement defining the required product performance;
3. A statement of the required operating time between failures;
4. A statement defining the environmental conditions in which the equipment must operate.

Using a set of systems or devices assumed to have identical parameters, reliability measures then give in the context of the above key elements information about the following.

1. The probability that a given system or device will operate for a specific time;
2. The number of failures that will occur over a specific period of time;
3. The average operating time between failures;
4. The expected lifetime of a system or device provided it works for a certain period in a given environment.

Burrell (2002) has already pointed to the fact that it might perhaps be unfortunate to think of citation as a “failure”, however, this approach allows adopting model and terminology of technical reliability. On the other hand, the term “failure” can be translated as a hit or incident and therefore be viewed from a positive perspective in citation analysis. Moreover, if ageing is measured by the number of received citations and the time elapsing between successive citations, the assumed “analogy” between failures in technical reliability and citations in the informetric ageing model does not at all seem to be absurd.

Lifetime is usually interpreted whether as time until failure or as time until death or destruction. In more general terms one could also consider ‘lifetime’ *the time during which a system is or can be used*; beyond this time it will not perform its intended function anymore. This definition can straightaway be applied to information science and bibliometrics as well. In particular, if citations are interpreted as *one important form of use of scientific information within the framework of documented science communication* (*Glänzel* and *Schoepflin*, 1999) then “technical reliability” of a scientific paper expresses the performance of its intended function, namely, that it is read and that it is has an impact on scientific research. The latter one can at least in part be measured by citations. Lifetime can consequently be interpreted as the period until it is not cited anymore. Analogously to a brand-new device that is not operating satisfactorily, a paper that is never cited can be considered not to give satisfactory performance of its intended function already at the time when it was published.

The concept of technical reliability implies a *diachronous* (*prospective*) approach. A device is produced, it is operating, and failures and finally death or destruction is observed. Nevertheless, this concept allows a *synchronous* (*retrospective*) approach, too, namely, in the context of the age of systems, machines or devices in use. This approach leads, however, rather to economical and social questions than to technical ones. To take just an example, the ageing of cars produced in a given year, represented by a certain brand and measured through their technical reliability corresponds to the diachronous approach. The comparison of the results of this diachronous analysis with those reflecting the situation, say, 15 years ago might reveal details concerning changing quality and technical standards. On the other hand, the analysis of the age of cars in a given environment, such as a region or country and the comparison of the present situation with that 15 years before sheds light on different aspects of obsolescence. In particular, the results are influenced by a variety of factors since the *synchronous* (*retrospective*) approach is based on the superposition of *diachronous* (*prospective*) ageing processes with different parameters that should by rights be decomposed before the results can correctly be interpreted. In the fictitious example of analysing and comparing the average age of cars in use, one finds a complex of technological, economical, ecological, social, climatic and infrastructure factors.

Despite the different background and the different interpretation of the diachronous and synchronous approach, it might, however, occur that the mathematical laws derived from the two approaches are very similar, so that the same formula could be used in both cases. In order to disprove this possible assumption, I will give a simple empirical evidence for the necessary distinction between the diachronous and the synchronous approach to ageing of scientific literature. The following example may just serve to visualise why diachronous bibliometric processes are not merely the "reflection" of the corresponding synchronous process. In order to show this, I have selected three journals representing three different subject fields, particularly, the journals *Cell* (Biosciences), *JACS* (Chemistry) and *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* (since 1986: *Probability Theory and Related Fields*) (Mathematics). All citable papers published in 1980 have been selected. The age of references not older than 21 years and all citations received in the 21-year period 1980-2000 are presented in Figure 1. It should be mentioned, that the charts in Figure 1 have a certain methodological shortcoming. The citations to papers published in 1980 are citations in SCI journals whereas the references include all kind of literature, that is, both journal literature and non-serial literature, too. Since references go back till 1960, this cannot be corrected by computerised techniques alone. Nevertheless, the results can be considered valid. I refer in this context to an earlier study by *Glänzel* and *Schoepflin* (1999). The authors of that study have found that the mean age of *references in serials* (a_S) and that of *references in non-serials* (a_N) are strongly correlated and the relationship can be expressed by the linear functions $a_N = 1.01 \cdot a_S - 0.25$ in the sciences and by $a_N = 1.04 \cdot a_S + 0.94$ in the social sciences.

Therefore, one can practically assume $a_N \equiv a_S$, that is, the age of serial references corresponds to that of the non-serial references.

Figure 1 Relative frequency of references and citations for three selected journals in 1980 as a function of time (top: CELL, centre: JACS, bottom: Z Wahrscheinlichkeit)

Proceedings from the above result, the charts in Figure 1 allow the following interpretation. The life-time curve of citations reflecting the diachronous process deviates significantly from that of references representing the synchronous process in all three cases. In particular, the citation curve is flatter and less skewed than the corresponding reference curve. This effect is most pronounced in the life sciences. It should be mentioned, that the results of the analysis of other journals (not presented here) show similar patterns.

One of the most important indicators derived from the life-time function is the *half-life*. It is defined as the number of years from the publication year that account for 50% of current citations received (*diachronous*, i.e., *prospective* approach), and for 50% of the current citations published in the references (*synchronous*, i.e., *retrospective* approach), respectively. The corresponding half-life values t_{m_d} (diachronous case) and t_{m_s} (synchronous case) calculated as interpolated medians take the following values for the three examples. 1. *Cell*: $t_{m_s} = 2.3 < 5.3 = t_{m_d}$, 2. *JACS*: $t_{m_s} = 4.7 < 6.7 = t_{m_d}$, and 3. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*: $t_{m_s} = 5.6 < 7.5 = t_{m_d}$. The diachronous (prospective) half-life of all journals is significantly greater that the corresponding synchronous (retrospective) one; it exceeds the synchronous half-life by two or three years. This again illustrates that results derived from synchronous analyses cannot simply applied to the diachronous case.

A stochastic model for diachronous (prospective) citation processes and its application to synchronous (retrospective) studies

In this section, a concise introduction into diachronous (prospective) and synchronous (retrospective) citation processes, respectively, will be given. It is based on a more general stochastic action-reaction model introduced by *Glänzel* (1983). In order to do without an excessive use of mathematical formalism, a verbal description of the rudiments is given. Nevertheless, the use of equations and formulae later on in the text will not completely be avoidable. The citation process will be defined as a diachronous (prospective) process since this approach seems to be the natural one: a paper is published and the citations it receives cumulated over the years form the process. Then the synchronous (retrospective) process, namely, the number of references as a function of time, i.e., the year in which the cited work has appeared can be derived from the individual diachronous (prospective) citation processes of the cited papers. It should, however, be mentioned that the reverse way is also possible by defining a synchronous process and deriving the diachronous one through the reference "processes" of citing papers.

In the following, I give the all definitions and tools necessary to model diachronous (prospective) and synchronous (retrospective) citation processes.

1. The basic idea of an action-reaction model is the introduction of stochastic processes through *timing-functions* defined on two discrete finite subsets in the original probability space. The two subsets will be denoted by **A** and **A'**, where these subsets need not be disjoint; they may even be identical. The elements of these sets are in the present paper assumed to be publications. The probability measure is defined as an elementary measure on the basis of a counting measure ϕ.

2. There are two subsets **M** and **N** of **R** (the reel axis) or **Z** (the set of integers), depending on whether a continuous or discrete time model is used. Without loss of generality we assume that $\mathbf{M} \subseteq \mathbf{N}$. These sets represent the time-parameter. In particular, the mappings $\nu: \mathbf{A} \to \mathbf{M}$ and $\mu: \mathbf{A'} \to \mathbf{N}$ are called *timing-functions* indicating when an event has happened, that is, in the present case, when a paper is published. In the following, time is assumed to be discrete, that is, $\mathbf{M}, \mathbf{N} \subseteq \mathbf{Z}$. The elements of these sets can, for instance, be the publication years of scientific papers.

3. The mapping $\tau: \mathbf{A'} \to \mathbf{A}$ describes the link between citing (**A'**) and cited papers (**A**). The "inverse" mapping can be defined using complete origins in the following way. Let $\underline{\mathbf{A}}\mathbf{'}_{\mathbf{N}}$ denote the set $\{\mathbf{A'}_{t_n,a} : t_n \in \mathbf{N}, a \in \mathbf{A}\}$, where $\mathbf{A'}_{t_n,a} = \{a' \in \mathbf{A'}: \mu(a') = t_n \wedge \tau(a') = a\}$. In verbal terms, $\underline{\mathbf{A}}\mathbf{'}_{\mathbf{N}}$ is the set of papers published in the period **N** and citing papers of the set **A**. Then the following mapping is uniquely defined: $\Lambda: \mathbf{N} \times \mathbf{A} \to \underline{\mathbf{A}}\mathbf{'}_{\mathbf{N}}$, where $\Lambda(t_n, a) = \mathbf{A'}_{t_n,a}$. The mapping $X^* = \phi \circ \Lambda$ then defines the *increments* of a stochastic process, namely the number of citations received by a paper a at time t_n . The process X then is the number of citations received in the period $t \leq t_n$, that is, $X(t_n) = \Sigma_{t \leq t_n} X^*(t)$. $X: \mathbf{A} \times \mathbf{N} \to \mathbb{N}_0$, with $\mathbb{N}_0$ being the set of non-negative integer numbers is an appropriate stochastic process that can be used as a model for *diachronous* (*prospective*) *citation process*.

4. *Random selection.* It is clear that in the above paragraphs a stochastic process X_s is defined for each $s \in \mathbf{M}$. The complete set of processes defined this way is $\underline{X} = \{X_s\}_{s \in \mathbf{M}}$. Then the measurable mapping $\nu: \mathbf{A} \to \mathbf{M}$ defines a *random selection* of processes from the set $\underline{X}$. Without loss of generality we can extend the definitions of the processes by putting $\mathbf{N} = \mathbf{M}$ and defining $X_s(t) \equiv 0$ for all $t < s$. In the discrete case, a process $X_s(t_n)$ can then be selected with probability $p_s = \mathrm{P}(\nu = s)$ for all $s \in \mathbf{M} = \mathbf{N}$. An alternative presentation of probability measures or expectations based on these processes is possible through the use of conditional distributions, where the condition is given by the selection of publication period, that is, for instance, $\mathrm{P}(X_s(t) = k) = \mathrm{P}(X(t) = k \mid \nu = s)$, or $\mathrm{E}\, X_s(t) = \mathrm{E}(X(t) \mid \nu = s)$ for any given time s. The random selection can be restricted to given subsets of **M**, and be made under certain conditions like publications in given journals, by authors from given countries, etc. This results in corresponding conditional probabilities p_s.

5. The appropriate *synchronous* (*retrospective*) process can now readily be derived from the *diachronous* (*prospective*) one by fixing t and through random selection from the corresponding set of increments $\underline{X}^* = \{X_s^*\}_{s \in \mathbf{M}}$.

> Random selection from the set $\underline{X}$, or from $\underline{X}^*$ with variable t, respectively, will result in *hybrid* diachronous-synchronous processes.

I give two simple examples for functions that can be used to measure ageing properties of the process. First, we can assume that if inf sup $\mathbf{N} = \infty$ and the process is convergent, that is, $X_s(t_n)$ converges to a non-degenerate random variable $X_s(\infty)$ with probability 1 for each $s \in \mathbf{M}$. We can further more assume that the sequence is uniformly integrable. Under these conditions, which are quite natural for citation processes, we can define the following measures.

The *mean-value function* is the simplest of these measures. It is defined as the mean value of $\mathrm{E}\,X_s(t_n)$ for any fixed $s \in \mathbf{M}$ at any time t_n. Here t_n denotes a discrete time, say, representing years. Its basic properties are obvious from the definition and the above assumptions:

$$\mathrm{E}\,X_s(t_n) \text{ is a non-negative, increasing function and } \mathrm{E}\,X_s(t_n) \rightarrow \mathrm{E}\,X_s(\infty) < \infty.$$

Hence we can define the second function, the *life-time function* of the process (see *Glänzel*, 1983). In particular, it takes the following form:

$$\mathrm{P}(\mu \le t_n \mid \nu = s) = \frac{\mathrm{E}\,X_s(t_n)}{\mathrm{E}\,X_s(\infty)} \ .$$

This life-time function can also be called life-time distribution, since $0 \le \mathrm{P}(\mu \le t_n) \le 1$ and $\lim_{t_n\to\infty} \mathrm{P}(\mu \le t_n) = 1$. This distribution has already been used, for instance, in the studies by *Glänzel* and *Schoepflin* (1994) and, more recently, by *Burrell* (2002).
The citation half-life is one of the most important indicators characterising ageing of scientific literature. This indicator can be obtained as the median of the life-time distribution of the diachronous (prospective) process in the following way.

$$t_m = \min\{t : \mathrm{P}(\mu \le t_n \mid \nu = s) \ge 0.5\} = \min\{t : \mathrm{E}\,X_s(t_n) \ge 0.5 \cdot \mathrm{E}\,X_s(\infty)\}.$$

In verbal terms, the citation half-life is the period $(t_m - s)$ beginning with the publication year in which a paper receives 50% of its citations.

Application to synchronous analyses

In the following the terms *publication period* ($S \subseteq \mathbf{M}$) and *citation window* ($T \subseteq \mathbf{N}$) will be used. The two periods will be assumed to form discrete sets of subsequent years. The fact that journal impact measures, such as the ISI Impact Factor, can be considered empirical mean-value functions of diachronous (prospective) citation processes in the sense of the previous section is quite obvious. In order to be in keeping with the standards in mathematical statistics, a distinction between *empirical* and *theoretical* impact measures will be made. Theoretical "journal impact measures" (cf., *Schubert* and *Glänzel*, 1983) can be defined as conditional expectations of citation distributions, where the condition is based on a partition defined through scientific journals. Thus the *theoretical* impact measure of a journal J_i is according to *Schubert* and *Glänzel*, the expectation $\mathrm{E}(Y_S(T) \mid J_i)$, where $S = [s, s+1]$ and $T = [t] = [s+2]$. I just mention in passing, that the empirical *IF* is the statistical estimator which is based on the corresponding means for papers of given document types and published in the journal under study. Using the above notations, one can write for the theoretical IF_i of journal i since Garfield's IF is a synchronous measure:

$$\mathrm{IF}_i = \mathrm{E}(X^*_{[s,s+1]}(s+2) \mid J_i) = p_s\mathrm{E}(X^*_s(s+2] \mid J_i) + p_{s+1}\mathrm{E}(X^*_{s+1}(s+2) \mid J_i),$$

where $p_s = \mathrm{P}(\nu = s \mid J_i)$ is the probability that a paper has been published in the year s and p_{s+1} is the probability that a paper has been published in the year s+1 in the journal i. From the empirical point of view, this equation has to be interpreted in the following way. We assume that the journal i has N_s (N_{s+1}) publications in the year s (s+1). These papers have received $C_s(s+2)$ and $C_{s+1}(s+2)$ citations, respectively. The estimates of the expected annual citation increments according to the diachronous (prospective) model are $C_s(s+2)/N_s$ and $C_{s+1}(s+2)/(N_{s+1})$, respectively. Random selection for the synchronous (retrospective) impact factor is here restricted to $S \subseteq \mathbf{M}$, that is, the estimate that a paper has been published in the year $s+j$ (provided it has appeared in journal i) is then $N_{s+j}/(N_s + N_{s+1})$ with $j \in \{0, 1\}$. Consequently, we can derive Garfield's impact factor from the annual diachronous (prospective) measures as follows:

$$\overline{\mathrm{IF}}_I = \frac{N_s}{N_s + N_{s+1}} \cdot \frac{C_s(s+2)}{N_s} + \frac{N_{s+1}}{N_s + N_{s+1}} \cdot \frac{C_{s+1}(s+2)}{N_{s+1}} = \frac{C_s(s+2) + C_{s+1}(s+2)}{N_s + N_{s+1}},$$

which is exactly the definition of the ISI Impact Factor.

This can be generalised for any synchronous (retrospective) journal impact measure (JIM_i):

$$\mathrm{JIM}_i(T|S) = \mathrm{E}(X^*_S(t) \mid J_i) = \Sigma_{s\in S}\, p_s\cdot\mathrm{E}(X^*_s\,(T) \mid J_i\,) = \Sigma_{s\in S}\, p_s\cdot\mathrm{E}(X^*_s\,(t) \mid J_i),$$

where inf sup $S < \infty$, $T = \{t\}$, $t \geq \max S$ and the random selection under the condition $\{J_i\}$ is restricted to S.

Hybrid diachronous-synchronous journal impact measures $\mathrm{JIM}_i(T|S)$ of the journal J_i for any publication period S and citation window T is then defined through increments of diachronous (prospective) citation processes as

$$\mathrm{JIM}_i(T|S) = \mathrm{E}(X^*_S(T) \mid J_i) = \Sigma_{s\in S}\, p_s\cdot\mathrm{E}(\Sigma_{t\in T}\, X_s^*(t) \mid J_i\,) = \Sigma_{s\in S,\, t\in T}\, p_s\cdot\mathrm{E}(X_s^*(t) \mid J_i),$$

where $X^*(T) = \Sigma_{t\in T}\, X^*(t)$, in general, and $X(t) = X^*(T) = \Sigma_{t\in T}\, X^*(t)$ if $T = [s, t]$, in particular.

Purely diachronous (prospective) or synchronous (retrospective) impact measures are obtained in the discrete case if the period S or T, respectively, reduces to a one-year period. The above properties immediately lead to the interesting conclusion that results obtained from synchronous (retrospective) studies can only be considered valid in the diachronous (prospective) case if ν has an approximately uniform distribution, otherwise results may be distorted by built-in biases caused, for instance, by the growth of literature or changes in the coverage of the bibliographic database used for the study (see also *Egghe*, 1993).

These example may illustrate how synchronous or hybrid measures can be derived from the diachronous (prospective) model. By contrast, the derivation of impact measures or even the change of the citation window is quite easy in the diachronous case. The theoretical journal impact measure of the journal indicator based on a 3-year window used in several recent studies (for instance, *Glänzel*, 1996) can be expressed by the following equation.

$$\mathrm{E}(X_s(s+2) \mid J_i) = \mathrm{E}(X_s^*(s) \mid J_i) + \mathrm{E}(X_s^*(s+1) \mid J_i) + \mathrm{E}(X_s^*(s+2) \mid J_i) \text{ with } T = [s, s+2].$$

The extension of the citation window results in adding the expectations of the increments of the new citation years to the equation. In particular, if $T_1 = [s+3, s+n]$ with $n>2$ then $\mathrm{E}(X_s(s+n) \mid J_i) = \mathrm{E}(X_s(s+2) \mid J_i) + \mathrm{E}(X_s^*(T_1) \mid J_i)$ is the theoretical journal impact measure for a $(n+1)$-year citation window beginning with the year of publication.

The two types of journal impact measures can be considered special cases of the mean-value functions introduced above. The second function, the life-time function, is designed to measure ageing properties of the process. For any fixed time s, the distribution $\mathrm{P}(\mu \le t) = \mathrm{E}\, X_s(t)/\mathrm{E}\, X_s(\infty)$ gives the probability of being cited not later than t. In the discrete case, we can write $\mathrm{P}(\mu = t_n) = \mathrm{E}\, X_s(t_n)/\mathrm{E}\, X_s(\infty)$, which is the probability that a paper is cited at time t_n. This measure does, however, not give any information about first or last citation since the event that a paper is cited at time t_n does not exclude possible citations at other time.

Although the proof is not straightforward, it is not difficult to show that under the assumption of the above diachronous (prospective) model the life-time function of the corresponding synchronous (retrospective) process for given fixed time t takes the following form.

$$\mathrm{P}(\nu \le s_n) = \{\sum_{s_n}^{t} p_{s_i}\cdot \mathrm{E}\, X_{s_i}^*(t)\}/\{\sum_{-\infty}^{t} p_{s_i}\cdot \mathrm{E}\, X_{s_i}^*(t)\}$$

These results may illustrate how synchronous (retrospective) ageing model can be developed on the basis of a diachronous (prospective) one, but it also shows the differences between the two approaches. The same applies to the corresponding half-life indicators that are defined with the help of the two life-time distributions. In this context it should be mentioned that the development of other synchronous (retrospective) measures than those based on mean values from the diachronous (prospective) model is more difficult.

Conclusions

From the empirical viewpoint, the diachronous (prospective) approach is the appropriate method to characterise citation processes. Since ageing of scientific literature has to be considered a real “process” (not only in the mathematical sense) with maturing and decay (cf., *Glänzel* and *Schoepflin*, 1995, *Moed* et al., 1998), this process can best be reflected by a measure of the use of (scientific) information. The idea of measuring ageing through the change of citedness in time, and thus of regarding ageing as a diachronous (prospective) process, is therefore quite obvious. It is also straightforward that indicators measuring citation impact should be based on this approach. Advantages of purely diachronous impact measures have been shown by *Ingwersen* et al. (2001) and *Rousseau* (1997). Besides the possibility of calculating impact measures also for non-serial literature (cf., *Rousseau*, 1997), diachronous (prospective) impact measures can easily take account to changing citation windows. Moreover, it could be shown that also synchronous (retrospective) impact-related measures can be derived from the diachronous (prospective) model.

References

BURRELL, Q.L., Stochastic Modelling of the First Citation Distribution. *Scientometrics,* **52**, 2001, 3-12.

BURRELL, Q.L., Modelling citation age data: Simple graphical methods from reliability theory. *Scientometrics*, **55**, 2002, 273-285.

EGGHE, L., RAVICHANDRA RAO, I. K., Citation age data and the obsolescence function: fits and explanations. *Information Processing and Management*, **28**, 1992, 201–217.

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

EGGHE, L., ROUSSEAU, R. Aging, obsolescence, impact, growth and utilization: Definitions and relations*. Journal of the American Society for Information Science*, **51**, 2000, 1004-1017.

GLÄNZEL, W., Stochastic action-reaction processes and their application to scientometric problems (in Hungarian). Doctoral thesis. University Budapest, 1983

GLÄNZEL, W., On Some Stopping Times on Citation Processes. From Theory to Indicators. *Information Processing and Management*, **28**, 1992, 53-60.

GLÄNZEL, W., SCHOEPFLIN, U., A Stochastic Model for the Ageing Analyses of Scientific Literature. *Scientometrics*, **30**, 1994, 49-64.

GLÄNZEL, W., SCHOEPFLIN, U., A Bibliometric Study on Ageing and Reception Processes of Scientific Literature. *Journal of Information Science*, **21** (1), 1995, 37-53.

GLÄNZEL, W., A Bibliometric Approach to Social Sciences. National Research Performances in Six Selected Social Science Areas. 1990-1992, *Scientometrics*, **35**, 1996, 291-307.

GLÄNZEL, W., SCHOEPFLIN, U., A Bibliometric Study of Reference Literature in the Sciences and Social Sciences. *Information Processing and Management*, **35**, 1999, 31-44.

GUPTA, R.C., On the Mean Residual Life Function in Survival Studies. In: C. Teil et al. (eds.), *Statistical Distributions in Scientific Work*, Vol. 5, D. Reidel, Dordrecht-Holland, 1981, 327-334

INGWERSEN, P., LARSEN, B., WORMELL, I.. Applying Diachronic Citation Analysis to Ongoing Research Program Evaluations. In: *The Web of Knowledge* : *A Festschrift in Honor of Eugene Garfield*. B. Cronin & H.B. Atkins [eds]. Medford, N.J. Information Today, Inc. & American Society for Information Science, 2000, 373-387.

INGWERSEN, P., LARSEN, B. ROUSSEAU, R., RUSSELL, J. The publication-citation matrix and its derived quantities. *Chinese Science Bulletin*, **46** (6), March 2001, 524-528; 700-704.

JURAN, J. M, GRYNA, F. M., Juran’s Quality Control Handbook. McGraw-Hill, 1988.

MOED, H.F., VAN LEEUWEN, T.N., REEDIJK, J., A new classification system to describe the ageing of scientific journals and their impact factors. *Journal of Documentation*, **54** (4), 1998, 387-419.

ROUSSEAU, R., Double exponential models for first-citation processes. *Scientometrics*, **30**, 1994, 213–227.

ROUSSEAU, R., The proceedings of the first and second international conferences on bibliometrics, scientometrics and informetrics: a data analysis. In: PERITZ, B., and EGGHE, L., eds., *Proceedings of the sixth conference of the international society for scientometrics and informetrics* 1997, 371-380.

ROUSSEAU, R., Jin, B., Yang, N. and Liu, X. Observations concerning the two- and three-year synchronous impact factor, based on the *Chinese Science Citation Database*. *Journal of Documentation*, **57**, 2001,349-357.

SCHUBERT, A., GLÄNZEL, W. Statistical Reliability of Comparisons Based on the Citation Impact of Scientific Publications. *Scientometrics*, **5,** 1983, 59-74.

WALLACE, D.P., The relationship between journal productivity and obsolescence. *Journal of the American Society for Information Science*, **37,** 1986, 136-145

WATSON, G.S., WELLS, W.T., On the Possibility of Improving the Mean Useful Life of Items by Eliminating those with Short Lives. *Technometrics*, **3**, 1961, 281-298

A revised version of this contribution has been published as:

Wolfgang Glänzel (2004). Towards a model for diachronous and synchronous citation analyses. Scientometrics, 60(3), 511-522.

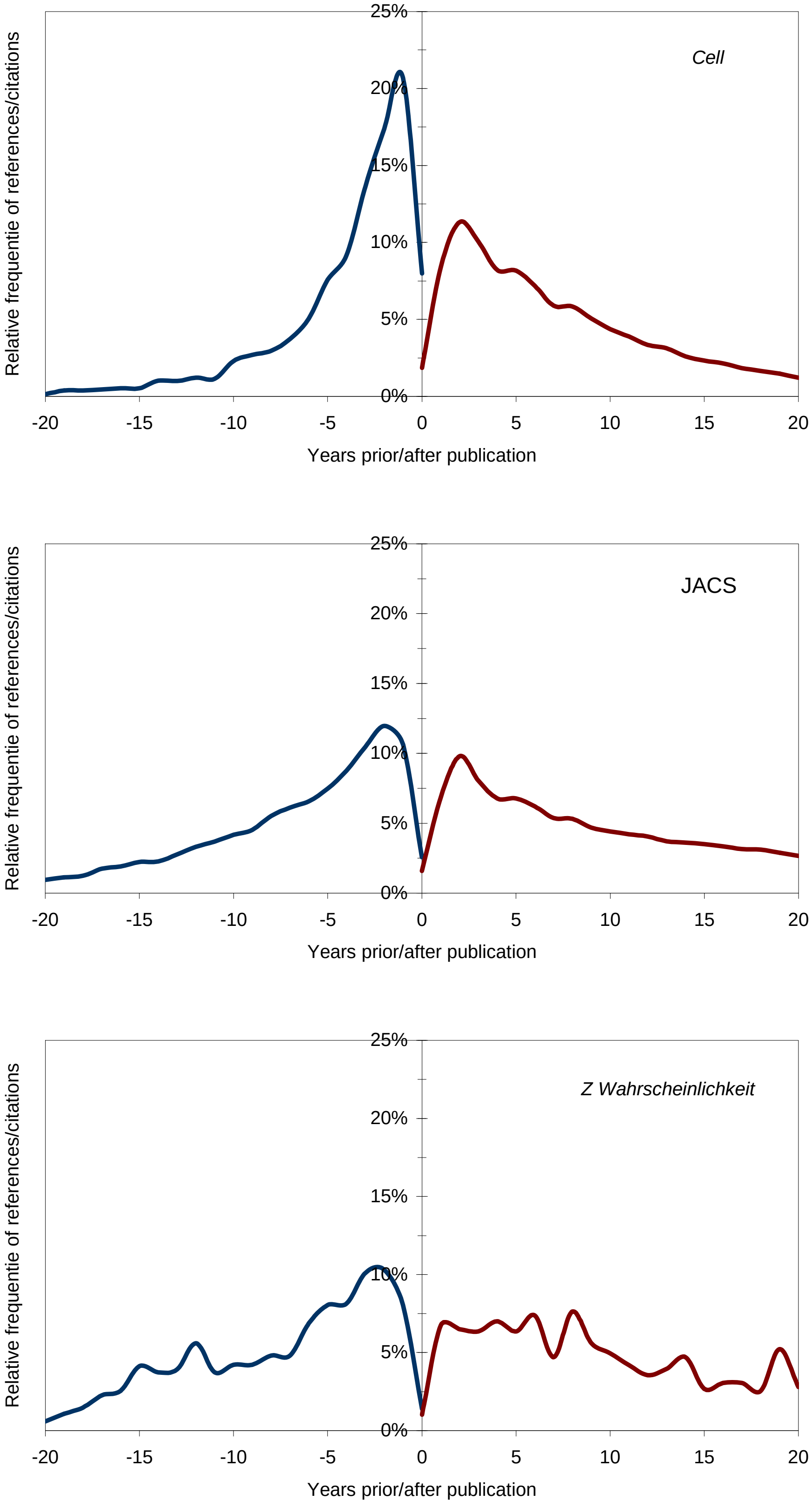

Figure 1 Relative frequency of references and citations for three selected journals in 1980 as a function of time (top: CELL, centre: JACS, bottom: Z Wahrscheinlichkeit)