e-Pathshala पाठशाला
A Gateway to All Post Graduate Courses

MHRD
Govt. of India

An MHRD Project under its National Mission on Education throught ICT (NME-ICT)

Subject: Library and Information Science

Production of Courseware
e-Content for Post Graduate Courses

Paper No : 10 Informetrics and Scientometrics

Module : 08 Different Models to Explain the Phenomena of Growth and Obsolescence of Literature

ज्ञान-विज्ञान विमुक्तये

e-Pathshala पाठशाला

MHRD Govt. of India

Development Team

Principal Investigator & Subject Coordinator	Dr. Jagdish Arora, Director INFLIBNET Centre, Gandhinagar
Paper Coordinator	Dr I K Ravichandra Rao Retd Professor, Documentation Research and Training Centre
Content Writer	Dr I K Ravichandra Rao Professor, Documentation Research and Training Centre
Content Reviewer	Prof A Y Asundi Retd Professor, Bangalore University

Unit 8

Different Models to Explain the Phenomena of Growth and Obsolescence of Literature

I. Objectives

The objectives of the Unit are to discuss:

- Different models which are useful to explain the data on growth of literature--exponential, logistic, power, Gompertz, etc.
- Obsolescence of literature
- Relation between growth of sources and items, growth rates and obsolescence rates, etc.

II. Learning Outcome

It is an important concept and you have now learned it. At the end of this module, you have also learnt various growth models and their characteristics; also, the relations among the various models; how to identify the trend and how to compute growth rates, doubling time, etc.

III. Module Structure

1. Introduction
2. Different Models of Growth
    2.1 The Exponential Model
    2.2 The Logistic Model
    2.3 The Power Function
    2.4 The Gompertz Model
    2.5 Ware's Model
3. Selecting a Trade Type
    3.1 Relationship between Growth of Sources (Journals) and Items (Articles)
4. Obsolescence of Literature
    4.1 Relation between Growth rate and Obsolescence rate
5. Summary
6. References

1. Introduction

The numbers of scientific journals including the abstracting periodicals are simple indicators of scientific growth. Price in 1963 argued observed that literature doubles approximately once in fifteen years. Neelameghan (1963) analyzed the documents on the history of medicine in India for the period 1954-61, during which period Indian contribution was 65% and foreign was 30%. He studied the growth of Indian medical societies and medical periodicals between 1780 and 1920. He also studied the coverage of Indian medical literature in Index Medicus and Experta Medica and it was found that they covered respectively only 38% and 13.5% of the Indian literature. Since then a number of articles were published on this topic, particularly the growth of literature in different subjects and on various growth models. The number and the growth characteristics (of articles, journals, scientists, discoveries, etc.) have been matters of some debate for considerable time. For instance, Price (1963) argued that:

- Once in fifty years the number of universities, labor force, population, etc. doubles;
- Once in twenty years GNP, discoveries, scientists, college entrants/1000 population doubles;
- Once in fifteen years the number of scientific journals doubles;
- Once in ten years the number of articles / literature in a field (particularly in science) doubles; and
- Once in two years the number of web sites doubles

As observed by many, the growth of publications passes through the following four stages (Price (1963), Michael Mabe (2003) and many others):

- The preliminary period of growth in which the absolute increments are small although the rate of increase is large – this is the first stage;
- During the period of exponential growth, the number of publications in a field double at regular intervals as a result of a high rate of growth – this is the stage 2;
- The rate of growth declines but the annual increments still remain approximately constant, in stage 3; and
- In the fourth stage, both the rate of increment and the absolute increase decline and eventually approach to zero.

The growth curves which explain the above four steps are also referred to as S-shape curve. A typical S-shape curve is given in Figure 1. This Figure clearly shows the four different stages – in X – axis, 0 to 3 is the first stage; 3 to 5 is the second stage; 5 to 7 is the third stage and 7 to 8 is the fourth stage.

The growth generally adopts an S-shaped pattern and is symmetrical about its point of infection. Price in 1963 estimated that the number of scholarly periodical titles being published at the end of the twentieth century would exceed one million. This however has not come true.

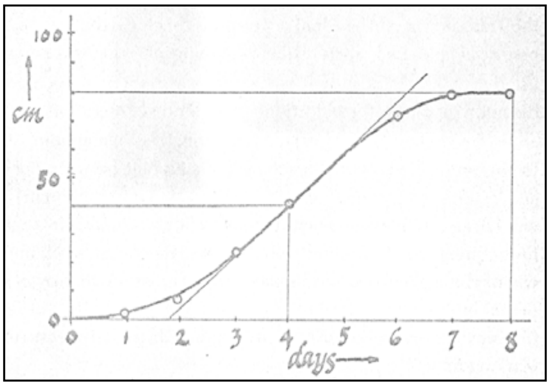


Fig. 1: A Typical S-shape Curve.

He argued that the logistic growth of knowledge over a period of time is a result of a number of applications of intellectual innovations. A logistic curve was fitted to the cumulative number of new publications appearing every year in science. He has also studied the growth of literature covered by Physics Abstracts during 1900-1950. He observed that except for the interruptions during the two world wars the literature has been increasing at exponential rates with a doubling time of about twelve years. Meadows (1993, 1998) on the other hand observed that estimates of the number of journals varied from 10,000 in 1951 to 70,000 in 1987.

Mabe (2003) in his study also observed that journal growth rates have been remarkably consistent over time with average rates of 3.46%, since 1800. He has in fact observed:

- From 1900 to 1940, the number of active journal titles grew at an actual rate of 3.23%, a doubling time of twenty-two years.
- From 1945 to 1976, the number of journals grew at an annual rate of 4.35%, representing a doubling time of sixteen years.
- Since 1977, the number of journals grew at 3.26%. Growth rates were very high; this trend continued until mid 1970s. Mabe pointed out that the slow growth rates after the mid 1970s were due to
    - The oil crisis of the 1970s
    - The increasing public awareness of potential ecological disaster and
    - The turning away from nuclear technology in the 1950s.

These factors, certainly lead to a slow down of government support for research. Yamazaki (1998), with the following three assumptions

i. The number of journals in a given subject is growing exponentially in time,
ii. Concurrently each journal is also augmenting the number of papers on the subject exponentially in time, and
iii. The rate of growth of articles in individual journal is the same for all journals,

Studied the number and rate of growth of scientific journals. His review critically assesses studies based on the 'ecological approach' to journal publishing growth. He concluded that the annual rate of growth of scientific journals is 1.85% from the end of the Eighteenth century. Naranan (1970) has shown that 'a frequency distribution (J(p)) of the number of journals with p articles is of the form

$$J(p) \propto p^{-\alpha}$$

With $\alpha \approx 2$, this model is similar to that of Bradford's law. The mathematical concepts of growth have become popular through the largely circulated report of the Club of Rome (Donella and Others (1972)). There are several models which are very useful to apply growth of literature. Some of these models are discussed below.

2. Different Models of Growth

2.1 The Exponential Model

An exponential model is associated with the name of Thomas Robert Malthus (1766-1834) who first realized that any species could potentially increase in numbers according to a geometric series. Exponential growth represents an increase with a fixed proportion of total population for each unit of time. For example, if a species has non-overlapping populations (e.g., annual plants), and each organism produces "b" offspring, then, the initial size of the population, say "a", in time t=0,1,2, is equal to (12):

$$Y_t = a.b^t$$

In this case, the growth rate is (b-1) * 100. This model is also popularly known as exponential model or log-linear model, and often expressed it as:

$$\ln y = \ln \alpha + \beta x,$$

Here β is the slope of curve, and measures proportional changes in y for a given absolute change in x. The model not only provides the rate of growth (the exponential parameter), but also the rate at which the size of the population doubles. The exponential growth has also been linked and compared with the size of compound interest. The exponential function assumes a convex shape in its graphical presentation. In exponential growth, the increase is proportional to population size, i.e. if the population is y at time t then

$$\frac{1}{y}.\frac{dy}{dx} = \beta,$$

where β is the Malthusian parameter of the population. In terms of differential equations,

if Y is the population, and dy/dx its growth rate, then its relative growth rate is $\frac{1}{y}.\frac{dy}{dx} = \beta$ If the relative growth rate is constant, it is not difficult to verify that the solution to this equation is P(x) = exp(βx). When calculating or discussing relative growth rate, it is important to pay attention to the units of time being considered.

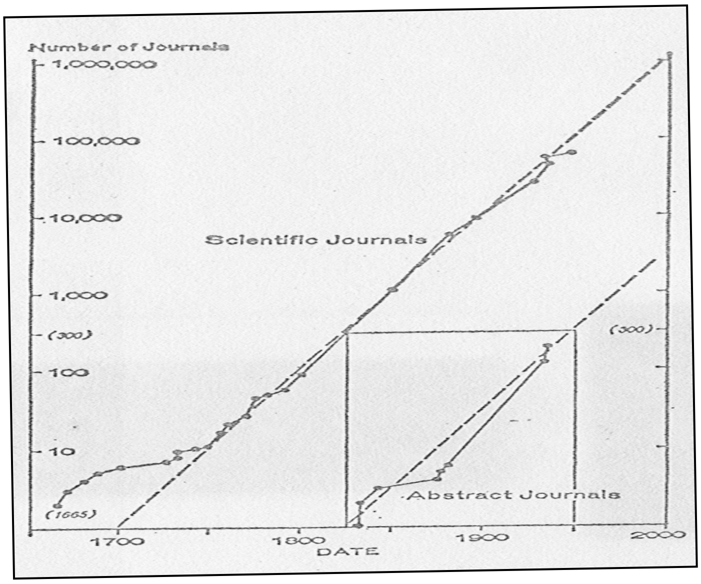


Fig. 2: Total # of scientific journals and abstract journals, as a function of date

(Source: Little Science Big Science)

If the growth rate is relative to the size of the population, then it is generally referred to as relative growth rate. It is also called the **exponential growth rate**, or the continuous growth rate. The exponential function can be applied to both growth process as well as decay process (obsolescence studies). The examples of growth process are given below (Croxton and Cowden (1966)):

- Growth of literature – growth of articles, journals, author population, universities, etc.;
- Microbiology (growth of bacteria);
- Conservation biology (restoration of disturbed populations);
- Insect rearing (prediction of yield);
- Plant or insect quarantine (population growth of introduced species); and
- Fisheries (prediction of fish dynamics). A typical exponential curve is shown below:

The **doubling time** is the period of time required for a quantity to double in size or value. It is applied to population growth, library collection, number of universities or colleges or students and many other things which tend to grow over time. When the relative growth rate (not the absolute growth rate) is constant, the quantity undergoes exponential growth and has a constant doubling time or period which can be calculated directly from the growth rate. This time can be calculated by dividing the natural logarithm of 2 by the exponent of growth, or approximated by dividing 70 by the percentage growth rate; that is:

$$D_t = \frac{\ln 2}{\ln(1 + \frac{r}{100})} \simeq \frac{70}{r}$$

The doubling time helps us to understand the long-term impact of growth than simply viewing the percentage growth rate. The doubling time is a characteristic unit (a natural unit of scale) for the exponential growth equation and its converse for exponential decay is the half life. For example with an annual growth rate of 4.8%, the doubling time is 14.78 years and a doubling time of 10 years corresponds to a growth rate between 7% and 7.5% (actually about 7.18%). Some doubling times calculated with this formula are shown in the following Table:

r%	**D_t**	**r%**	**D_t**	**r%**	**D_t**	**r%**	**D_t**
0.1	693.49	3.0	23.45	6.0	11.90	9.0	8.04
0.5	138.98	3.5	20.15	6.5	11.01	9.5	7.64
1.0	69.66	4.0	17.67	7.0	10.24	10.0	7.27
1.5	46.56	4.5	15.75	7.5	9.58	15.0	4.96
2.0	35.00	5.0	14.21	8.0	9.01	20.0	3.80
2.5	28.07	5.5	12.95	8.5	8.50	20.0	3.80

Table1: Doubling times D_t given constant r% growth

Given the two measurements of a growing quantity, q_1 at time t_1 and q_2 at time t_2, and assuming a constant growth rate, you can calculate the doubling time as

$$D_t = (t_2 - t_1) x \frac{\ln 2}{\ln\left(\frac{q_2}{q_1}\right)}$$

The equivalent concept to doubling time for a material undergoing a constant negative relative growth rate or exponential decay is the half-life

The Calculation of Simple Percentage Growth rate

The percent change from one period to another is calculated from the formula:

$$GR\ (\%) = \frac{(y_{t+1} - y_t)}{y_t}\ x\ 1$$

Where:

GR(%)	=	Percent Growth Rate
y_{t+1}	=	Value at time (t+1)
y_t	=	Value at time t

The annual percentage growth rate is simply the percent growth divided by N, the number of years.

Example

In 1980, the number of documents in Library A was 250,000. This grew to 280,000 in 1990. What is the annual percentage growth rate of library collection in Library A?

y_t = 250,000, y_{t+10} = 280,000 and N = 10

N = 10

$$GR\ (\%) = \left(\frac{(280000 - 250000)}{250000}\ x\ 100\right)$$

$$= 1.2\%$$

The library collection grew 12 percent between 1980 and 1990 or at a rate of 1.2 percent annually

2.2 The Logistic Model

The Belgian mathematician Pierre Verhulst (1838) developed the Logistic model (17). A typical logistic curve is shown in Figure 3. He suggested that the rate of population increase might depend on population density:

$$y = \frac{a}{bc^t}$$

The algebra of the logistic family is something of a hybrid. It mixes together the behaviors of both exponentials and powers (proportions, like rational functions). The parameters b and c are simply the y-intercept and the base of the component exponential function bc^t. The rate at which a logistic function falls from or rises to its limiting value is completely determined by the exponential function in the denominator, by the parameters b and c.

Here, as observed in exponential function, one notice that logistic function depends on the population itself, i.e., $\frac{dy}{dx} = \lambda(1 - \frac{y}{k})y.$ Pearl and Reed have independently developed and they used the curve to describe the growth of albino rat and tadpole's tail, the number of yeast cells, and most interesting of all, the number of human beings in a geographical area. The curve is also known as Pearl and Reed curve.

The curve depicts three things, as explained in section 1:

i. Show growth in the early stage.
ii. Intermediate period of rapid growth
iii. An approach to maturity.

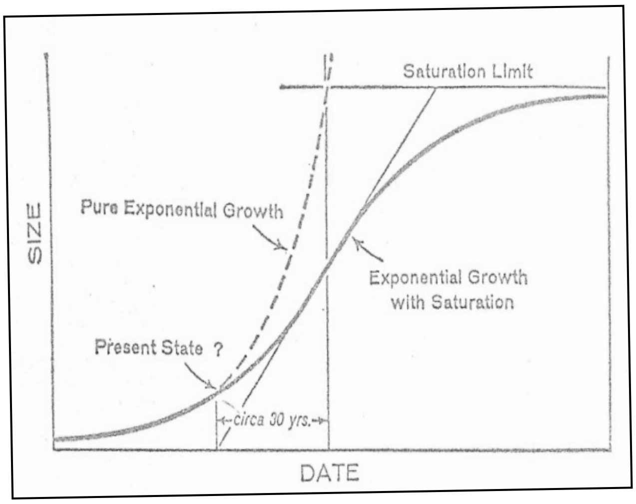


Fig. 3: A typical logistic curve

2.3 The Power Function

It is a log-log or double log model. It is mathematically represented as:

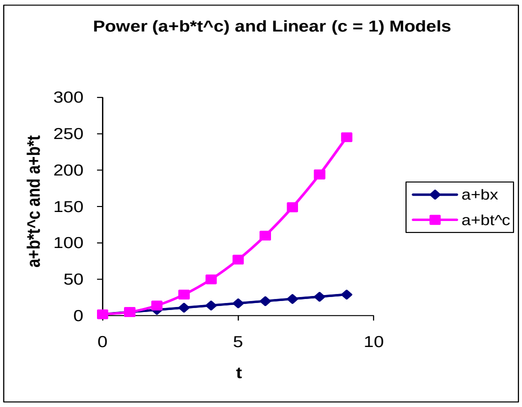
$$y = \alpha\, t^{\beta}$$

i.e., $\log y = \alpha + \beta \log t$

Sometimes, the power model is also represented as

$$y_t = \alpha + \beta t^{\gamma}$$

where α, $\beta > 0$. For $0 < \gamma < 1$, the function y takes a concave shape, but without an upper limit. For $\gamma = 1$, the function y assumes a linear shape. For $\gamma > 1$ the function y takes a convex shape. A typical Power model is shown in Fig. 4.



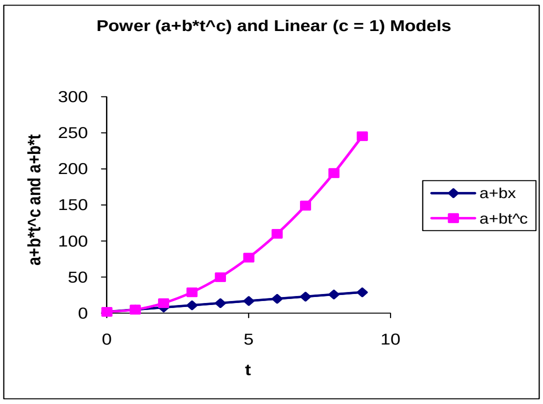


Fig. 4: A Typical Power Model

2.4 The Gompertz Model

The Gompertz (1825) model describes a trend in which the growth increment of the logarithms is declining by a constant percentage. Thus, the natural value of trend would show a declining ratio of increment, but the ratio does not decrease by either a constant amount or a constant percentage. The equation for the Gompertz curve is

$$y = ab^{c^t}$$

i.e., $\log y = \log a + (\log b)\, c^t$

The Gompertz and logistic curves are similar in that they both can be used to describe an increasing series, which is increasing by a decreasing percentage of growth, or a decreasing series, which is decreasing, by a decreasing percentage of declines. They differ in that the Gompertz curve involves a constant ratio of successive first differences of the log y values, while the logistic curve entails a constant ratio of successive first differences of 1/y values.

2.5 Ware's Model

The other equally well-known model is Ware's model. It is represented as

$$y = \delta\,(1-\phi^{-t}) \qquad \delta, \phi > 1.$$

The Figure 5 shows a typical Ware's model.

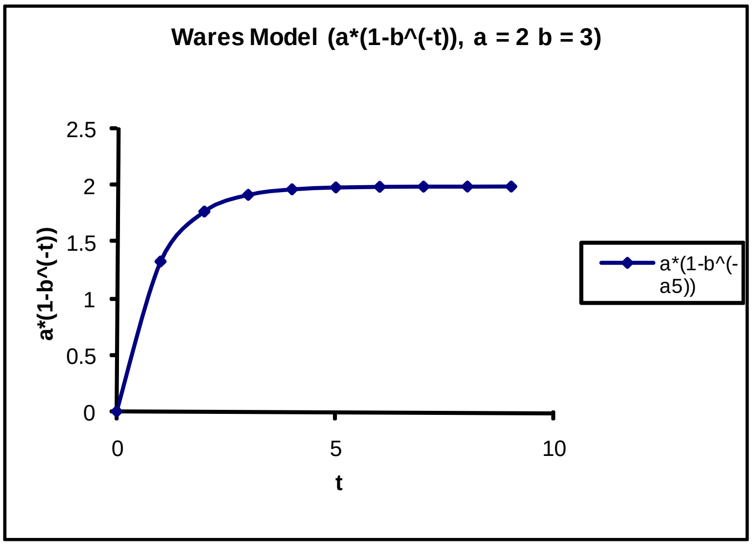


Fig. 5: A Typical Ware's Model

3. Selecting a Trade Type

There are many trend types. It is difficult to decide which one to use. Following are the simple guidelines to select the appropriate model (Croxton and Cowden):

i. If the approximate trend, when plotted on semi-logarithmic paper is strait line or if the first differences of logarithms are constant, use an exponential model.
ii. If the first differences resemble a skewed frequency curve, or if the first differences of logarithms are changing by a constant percentage, use a Gompertz model.
iii. If the first difference resembles a normal frequency curve or if the first differences of reciprocals are changing by a constant percentage, use logistic model. Also, it the approximate trend value (or the original data), when expressed, as percentage of a selected asymptote, appears linear on arithmetic probability paper, use a logistic model.

Egghe and Rao (1992) have suggested an innovative methodology for identification and classification of growth models. They have classified the growth models based on growth functions, i.e. α_1 and α_2. They have denoted the growth function as C(t) (theoretical or concrete data). These growth rate functions may be defines as:

$$\alpha_1(t) = C(t+1)/C(t) \quad \text{and} \quad \alpha_2(t) = C(2t)/C(t)$$
$$t = 1,2,3, \ldots.$$

α_1 is called the first growth rate function and α_2 is called the second growth rate function. The basic idea here is that the graph of α_1 and α_2 are much more different than the corresponding graphs of other growth models. The theoretical relationship between α_1 and α_2 has been worked out to be:

$$\alpha_2(t) = \alpha_1(2t\text{-}1)\, \alpha_1(2t\text{-}2) \ldots. \alpha_1(t)$$

According to them, the method of determining the growth goes back to the intrinsic growth rate properties of the data for a good understanding of what is really going on. The authors have also suggested that in order to get a simple clue for the selection of the best model, the plot of two growth rate functions for different mathematical models (namely; exponential, power, linear, logistic and Gompertz) may be drawn and visualized. These graphs can be classified as: Type 1 – increasing, Type 2 – Constant, type 3 – decreasing, Type 4 – increasing and then decreasing, as shown in Table 2.

Types of model	**Growth rate function**	
	α_1	α_2
(1)	**(2)**	**(3)**
Exponential	Type 2	Type 1
Logistic or Gompertz (0<b, c<1)	Type 3	Type 4
Gompertz (b, c>1)	Type 1	Type 1
Power (α >0, 0< $\gamma \leq 1$)	Type 3	Type 1
Power (α >0, γ >1)	Type 4	Type 1
Power (α > 0)	Type 3	Type 2

Table 2: Classification of growth models using growth rate function

Wolfram et.al (1990) explored the Linear, Exponential and Power model to the growth of publications in a period of 20 years, as reflected in the databases belonging to science, technology, social science and humanities. They found that, in most cases, the mathematical model that provided the best fit to the observed data was a power model, rather than an exponential, logistic or a linear model; and they concluded, "The breakdown in exponential growth is well underway. The power model was in particular best, because it has the advantages of modeling the growth behavior of both the linear and exponential models.

Egghe and Rao (1992) clarified the formal distinctions between the four models that Wolfram et al. examined, pointing out that any linear model should more properly be recognized as a power model of a special kind, and introducing two other comparable models, the Gompertz and Ware functions, that Wolfram et al. did not consider. Revisiting the data collected in the earlier study, Egghe and Rao observed that an exponential model was never the best fit. Indeed, they have shown that such a model could never have been expected to provide the best fit, given that the rate of growth in every database declined steadily over the years studied. They also found that a power model fitted best in cases of convex growth and that a Gompertz model generally fitted best in cases of S-shaped growth. Egghe and Rao's findings suggested that, in modeling the growth of literature, the choice between an exponential and a logistic function may have always been a false one, and that we should instead be asking whether growth is best described by a power law or a Gompertz function.

3.1 Relationship between Growth of Sources (Journals) and Items (Articles)

Egghe (2005) observed that growth rates of sources (journals) usually are different from growth rates of items (articles); further he argued, "The references in publications grow with a rate that is different (usually higher) from the growth rate of the publications themselves." His study showed that Naranan's model (exponential model: y=act) hardly fits the empirical data. He showed that the "simple" 2-dimensional informetrics models of source-item relations are not able to explain this. He has further shown that a linear 3-dimensional informetrics (i.e. adding a new source set) is capable to model disproportionate growth. The explanation consists of "defining" a set of "super sources" which produce the original sources but which also attach the items into the original sources. In this way, disproportionate growth of references versus articles can be explained by looking at authors. In the same way, disproportionate growth of articles versus journals (a new dataset is compiled from the database Econlit) can be explained by considering journal publishers. Formulae of such different growth rates are presented using Lotkaian informetrics and new and existing data sets are presented and interpreted in terms of the used linear 3-dimensional mode.

Sahoo (2006) in his thesis compared the growth of the journals with that of the articles, in the area of software studies. He observed that the correlation coefficient (r) between the number of journals and the articles is 0.9811. That is, 96% of the variation in y (articles) is due to the variation in x (journals) – **higher the number of journals, the higher the number of articles**. However, for the case of World literature he has observed that the correlation coefficient between number of journals and the number of articles published is only 0.8517. Unlike India, the correlation is not so high for the world literature. Figures 6 and 7 shows the growth rate curves for journals and articles published by the journals for India and World literature respectively from 1990 to 2003. It has been observed that the growth rates of journals have been decreased with the decrease of growth rate of articles. However, for the 'India data for the years 1995 and 2001, it has been observed that when there is a negative growth in journals, it does not show any negative growth for the article. From these observations, Sahoo concluded that only in some cases the growth of journals affect the growth of the number of articles. The growth rate of journals and articles are not same; for the Indian literature average growth rate is 9% for both journals and the articles and for the world literature the average growth rate of journals is 3% and average growth rate of articles is 9%.

In another study, Ravichandra Rao and Divya (2010) studied the growth of literature in Malaria Research. They also studied the relation among journals, articles and authors. Their study suggests that the journals, articles, and authors increase approximately exponentially. The number of articles has increased from 3,996 to 57, 627 from 55-65 to 96-05. Also the number of journals has been increased 503 to 3,072 from 55-65 to 96-05. The R^2 value for the trend for journals, articles, and authors are 0.9502, 0.9475 and 0.9651 respectively; the low R^2 value are perhaps due to the less number of data sets; the Figures 8-10 are much more convincing that the data on journals, articles, and authors increase exponentially. Under the assumption that the data confirm to exponential model, the growth rates have been computed; the growth rates of the journals, articles and authors are 5.31%, 7.38%, and 10.06% respectively. The most important observation is that the number of least productive journal has been increased to 2,951 from 463. This perhaps due to

- Interdisciplinary nature of research in Malaria and related topics
- May be an incomplete bibliography
- High growth rates (exponential in nature!) of journals and articles

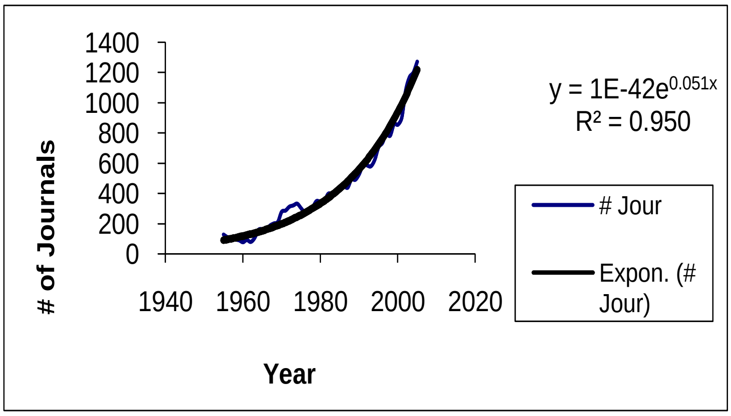


Fig. 8: Growth curve of journals

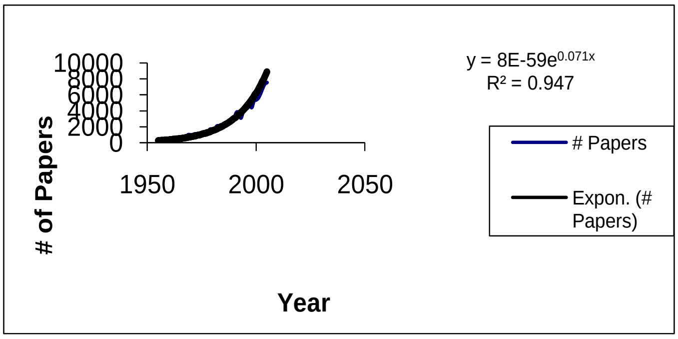


Fig. 9: Growth curve of articles

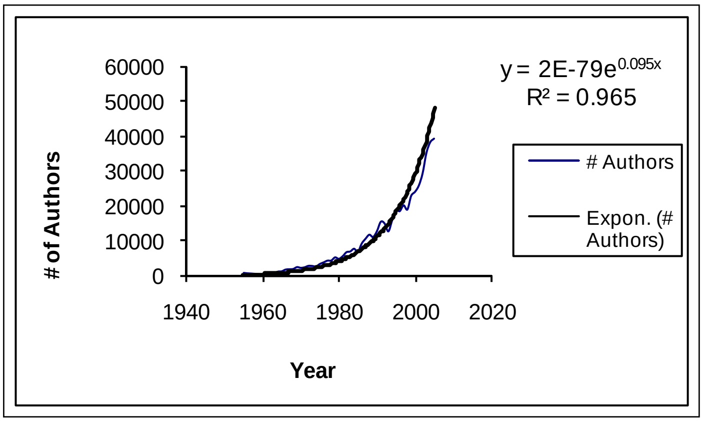


Fig. 10: Distribution of author

An attempt was also made to analyze the chemical literature based on the data from Chemical Abstracts. The Fig. 11 shows the growth trend with its R^2. The growth rate is 5.98.

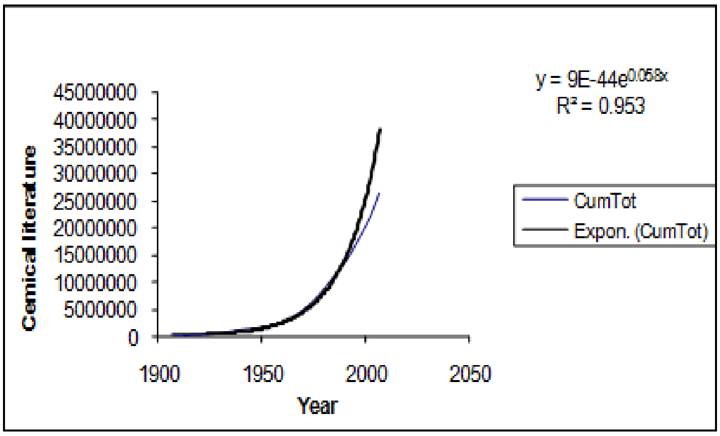


Fig. 11: Growth of chemical literature

4. Obsolescence of Literature

'Obsolete' generally means out of date or no longer in use. The process of being obsolete is known as obsolescence. It is also often referred to as 'phenomenon of replacement.' The term obsolescence is used for the first time by Gross and Gross in 1927. The authors analyzed the references in the 1926 volume of the journal of chemical literature and observed that the number of references falls to one-half in

fifteen years. Obsolescence is thus a characteristic of scientific and technical literature. Burton and Kebler (1960) are the first to use the term 'half-life' in 1960. It is defined as 'the time during which one-half of all the currently active literature published.' It is the period of time needed to account for one-half all the citations received by a group of publications. The concept of half-life is always discussed in the context of diachronous studies. More precisely, Line and Sandison (1974) refer to diachronous studies in those that follow the use of particular items through successive observations at different points in time, whereas synchronous studies are concerned with the plotting the age distribution of material used at one point in time.

However, there is no reason suppose that the half-life for some subject is the same as the median citation age in that subject. Half-life in the context of synchronous data is referred to as median age of the citations / references. The use of literature may decline much faster with data of ephemeral relevance, if it is in the form of reports, thesis, advance communication or pre-print and in the context of advancing technology. However, the use of literature may decline slowly when it is descriptive (e.g., taxonomic botany) and critical (e.g., literary criticism); it may also decline if it deals with concepts (e.g., philosophy).

Brookes in one of his articles (1970) argued that if growth rates of literature and contributors are equal then the obsolescence rate remain constant. In this sense growth and obsolescence are related. Ravichandra Rao and Meera (1991) have studied the relation between growth and obsolescence of literature, particularly in mathematics. Gupta (1998) studied the relationship between growth rates and obsolescence rates and half-life of theoretical population genetics literature. He explored the application of lognormal distribution to the age distribution of citations over a period of time.

In the analysis of obsolescence, Brookes argued that the geometric distribution expresses the idea that when a reference is made to particular periodical of age t years. The geometric distribution is given by

$$(1-a)\, a^{t-1} ,$$

'a (< 1)' is a parameter – the annual aging factor; it is assumed to be constant over all values of t.

$$\text{Let } U = 1 + a^2 + a^3 + a^4 + \ldots. + a^t + \ldots.$$

$$\text{i.e., } U = 1/(1-a).$$

Similarly if $U(t) = a^t + a^{t+1} + a^{t+1} + a^{t+2} + \ldots\ldots = a^t\,(U(0),$

then $U(t)/U(0) = a^t$.

Using this relation, by graphical method, we can compute half-life as well as 'a'.

4.1 Relation between Growth rate and Obsolescence rate

If we assume the literature is growing exponentially at an annual rate of g, we then have

$$R(T) = R(0)e^{gT},$$

where R(T) is the number of references made to the literature during the year T.

We also have

$$U(0) = R(0)/(1-a_0) \text{ and } U(T) = R(T))/(1-a_T)$$

where a_0 and a_T are the annual aging factors corresponding to the years 0 and T respectively. Under the assumption that utility remains constant (U(0) = U(T)) , we then have

$$R(0)/(1-a_0) = R(T))/(1-a_T)$$

By substituting the value of R(T), we thus have a relation between the growth and the obsolescence:

$$\mathbf{e^{gT} = (1-a_T)/(1- a_0)}$$

However, Egghe and Ravichandra Rao (1992) showed that the obsolescence factors (aging factors) 'a' is not a constant, but merely a function of time. The authors have also shown that the function 'a' has a minimum which is obtained at a time t later than the time at which the maximum of the number of citations is reached.

Egghe (1993) also developed a model to study influence of growth on obsolescence. He found different results for the synchronous and diachronus study. He argued that for an increase of growth implies an increase of the obsolescence for the synchronous case and for the diachronous case, it is quite opposite. In order to derive the relation, he also assumed the exponential models for growth as well as for obsolescence. In another paper, for the diachronous aging distribution and based on a decreasing exponential model, Egghe derived first citation distribution. In his study he assumed the distribution of the total number of citations received confirms to a classical Lotka's function. The first citation distribution is given by

$$\phi(t_1) = \gamma (1- a^{t}{}_1)^{\alpha-1}$$

where $\gamma$the fraction of papers is that eventually get cited; t_1 is the time of the citation, 'a' is the aging rate and α is Lotka's exponent. Egghe and Ravichandra Rao in their study in 2002 observed that the cumulative distribution of the age of the most recent reference distribution is the dual variant of the first citation distribution. This model is different from the first citation distribution. In another study, Egghe and Rao have shown the general relation between the first citation distribution and the general citation age distribution. They have shown that if Lotka's exponent $\alpha = 2$, both distributions are the same. In the same study, they have argued that the distribution of

n^{th} citation is similar to that of the first citation distribution. Egghe, Rao and Rousseau (1995) studied the influence of production on utilization function. Assuming an increasing exponential function for production and a decreasing one for aging, the authors have shown that in the synchronous case, the larger the increase in production, the larger the obsolescence; however, for the diachronous case it is quite opposite. This proof is different from the earlier one derived by Egghe (1993).

5. Summary

Most of the studies on growth of literature or obsolescence of literature are empirical in nature. Once we identify a suitable model to explain the empirical data on growth of literature or obsolescence of literature, we can apply regression analysis (to fit either the linear or non-linear model to the observed data). Then, by applying the appropriate statistical tests (usually, by considering the R^2 value), we may choose a model to explain the data.

Any study in scientometrics must be based on certain guidelines/ methodologies; in order to accept its validity as well as its generalization of the result. The general guidelines are:

- Identify the general problem(s)
- Conduct literature search
- Decide the design methodology
- Collect the data either for the population or for a sample
- Analyze the data
- Report the result
- Refine the hypotheses

In Step 5, to analyze the data we may adopt one of the models which are discussed in this Unit. Scientometric techniques have evolved over time and are continuing to do so. The counting of papers with attribution (by country, by institution, by author, etc.) is a part of the scientometrics.

It has been argued that the growth of entities (such as journals, articles, authors, etc.) follows exponential model; however, it has been observed that it is not always true; so there is need for to identify a suitable model to explain the data related to growth of scientometric entities in general and also to study the data on obsolescence of literature.

6. References

1. Brookes, B.C. (1970) "Obsolescence of special library periodicals: Sampling Errors and Utility contours". Journal of the American Society for Information Science, 21: 320-9.

2. Burton, R.E. and Kebler, R.W. (1960) The half-life of some scientific and technical literature. American Documentation, 11: 18-22.

3. Croxton, F.E. and Cowden, D.J. (1966) “Applied general statistics”. Prentice-Hall of India (Private) Ltd. New Delhi.

4. “Definition of Gross domestic product”. http://www.wordiq.com/definition/Gross_ domestic_product.

5. Donella, H. Meadows; Dennis L. Meadows; Jorgen Randers and William W. Behrens III. (1972) The Limits to Growth. Club of Rome.

6. Egghe, L. (1993) “On the Influence of Growth on Obsolescence”. Scientometrics, 27(2); 195-214.

7. Egghe, L. (2005) “An explanation of disproportionate growth using linear 3-dimensional informetrics and its relation with the fractal dimension”. Scientometrics. 63(2); 277-296.

8. Egghe, L. and Rao, I.K.R. (1992) “Classification of growth models based on growth rates and its applications”. Scientometrics, 25: 5-46.

9. Egghe, L. and Ravichandra Rao, I.K. (1992) “Citation age data and the obsolescence function: fits and explanations”. Information Processing and Management, 28(2): 201-17.

10. Egghe, L. and Ravichandra Rao, I.K. (1995) “On the influence of production on utilization functions: Obsolescence or increased use?” Scientometrics, 34(2): 285-315)

11. Egghe, L. and Ravichandra Rao, I.K. (2002) “Theory and Experimentation on the most recent reference distribution”. Scientometrics. 53(3): 371-387.

12. Exponential Model. http://www.ento.vt.edu/~sharov/ PopEcol/lec5/exp.html.

13. Gompertz, B. (1825) “On the Nature of the Function Expressive of the Law of Human Mortality and on a New Mode of Determining the Value of Life Contingencies”. Philosophical Transactions of the Royal Society of London, 115: 513-585.

14. Gross, P.L.K. and Gross, E.M. (1927) "College Libraries and chemical education". Science, 66: 1229-34.

15. Gupta, B.M. (1998) "Growth and Obsolescence off Literature in Theoretical Population Geneticsss…" Scientometrics, 42(3):335-347.

16. Line (Maurice B.) and Sandison. A. (1974) Obsolescence and change in the use of literature with time. Journal of Documentation. 30(3): 283-350.

17. Logistic Functions. http://www.wmueller.com/precalculus/families/1_81.html.

18. Mabe, Michael. (2003) "The growth and number of journals". Serials. 16(2): 191-7.

19. Malta, Kaushik; et al. (2005) "Scaling phenomena in the growth dynamics of scientific output". Journal of the American Society for Information Science and Technology, 56(9): 893-902.

20. Meadows, A.J. (1993) In Woodward and Pilling. The International Serials Industry. Aldershot, Gower, pp 24-7.

21. Meadows, A.J. (1998) "Communicating Research. Academic Press. London and San Diego Press. p. 15-6.

22. Menard, H.W. (1974) Science: growth and change. Harvard University Press, Cambridge, Mass.

23. Naranan, S. (1970.) "Bradford's law of bibliography of science: an interpretation". Nature. 227(5258): 631-632.

24. Neelameghan, A. (1963) Documentation of the History of Medicine in India. Ann Lib Sc. Doc., 10(3/4): 116-42.

25. Price Derek De Solla. (1963) "Little Science, Big Science". Columbia University Press. New York (GS201.)

26. Ravichandra Rao, I.K. and Meera, B.M. (1991) "Growth and obsolescence of literature: an empirical study". **In** I.K.R. Rao ed. Informetrics – 91. Sarada Ranganatahan Endowment for library. Bangalore. p. 377-394.

27. Ravichandra Rao I.K. and Divya Srivastava (2010). 'Growth of Journals, Articla and authors". Journal of Informetrics, 4(3): 249-256.

28. Robert Malthus, Thomas. (1798) "An essay on the Principle of Population".

29. Sahoo, Bibhuti Bhusan. (2006) "Scientometric Study of Literature in Software Studies in India with a comparison to the World Literature". Thesis Submitted to the Dept. of Library and Information Science of the Pune University for the Degree of Doctor of Philosophy in Library and Information Science. Guide: I. K. Ravichandra Rao, Documentation Research and Training Center, Indian Statistical Institute, Bangalore.

30. Wolfram D; Chu, C.M. and Liu, Xin (1990). "Growth of Knowledge: Bibliometric analysis using online databases data in L Egghe and R Rouseau, Eds. Informetrics 89/90. 355-72.

31. Yamazaki, Shigeaki. (1987) "A critical review of some ecological studies on scientific journals [in Japanese]". Journal of Information Processing and Management, 29(10): 863-870.