DESIDOC Bulletin of Information Technology, Vol. 22, No. 2, March 2002, pp. 17-25

Application of Growth Models to Science and Technology Literature in Research Specialities

Praveen Sharma, B.M. Gupta, Suresh Kumar

Abstract

The paper discusses the application of three well-known growth models, namely exponential, logistic, and power to the growth of world literature in physics, chemistry, and electrical & electronic engineering research, as reflected in international databases in these three fields from 1907 to 1994. The growth models are classified and characterised on the basis of two growth-rate functions as suggested by Egghe and Rao[9]. The methodology suggested by Egghe and Rao in identifying a growth model has been reviewed with the help of case studies on the growth of literature in the three major fields. The results obtained through the analysis are not found to be in full agreement with the results expected using Egghe & Rao methodology. The present study emphasises the need to undertake more rigorous research on the time series data on publications growth of various fields to test the utility of these growth rate functions.

1. INTRODUCTION

Cole and Eales [1] published their first study in 1917 on the development of comparative anatomy over three centuries, covering the period since 1550, with disaggregation of the data by country and topical sub-area within the field. They were among the first to utilise the published literature for constructing the quantitative profile and indicators of a research field. This pioneering effort was subsequently strengthened by two agricultural scientists, Wilson and Fred [2] who published a study in 1935 of the growth and development of research activities on the nitrogen fixation of plants. After a gap of two decades, interest in quantitative studies of science again resurged. Derek Solla de Price[3] a historian of science and physicist by training, strongly advocated the idea that, while we might not be able to read all the scientific publications that appear, we might learn something about science by merely counting them. He actively suggested studying scientific developments with quantitative methods using demographic and bibliometric data. Price[4] studied the growth of physics literature by counting the number of abstracts in *Physics Abstracts* from 1900 to 1950. He found that, except for interruptions during the two world wars, physics publications have been increasing exponentially, with a doubling time period of about 12 years. Along with this, Price also took different types of statistical data on science and technology, such as journals, manpower, expenditure, etc., and has shown that an exponential growth model fits in all types of growth of such statistical data.

The proliferation of scientific literature appearing throughout the world has quite naturally promoted a truism on the part of observers and specialists, that the escalation in growth of scientific literature could not follow an exponential function indefinitely. This very high level that has been reached for sometime led them to think that inevitable slow down would take a form of logistic curve at some point of time in the growth of a research field.

The pioneering efforts of Price's work brought legitimacy to the scholarly examination of the literature of specialities, partly because his conclusions conceded with, and indeed reinforced, the widely held belief that the proliferation of science seemingly knew no bounds. In recent years, studies of emerging fields of science and technology, using bibliometric data, have fairly become common. The nature of this research is quite diverse and there is no single research method that has come to the forefront. However, in general, there are two basic approaches. The first approach derived from the works of Price and few other scholars, seeks to model the growth of a research field by measuring its annual growth of publications and authors. The second approach is quite different, in that it uses the citation as a unit of analysis. Citation-based studies seek to understand the interlocking nature of citation patterns as a means of observing the growth and development of clusters of scholars who may ultimately come to form the basis of a new field. Several scholars undertake both type of studies, the attention to the latter has been increasing more recently in the field.

2. LITERATURE REVIEW

Many scholars have studied the application of selected growth models to the literature growth in research specialities. May [5] explored the growth of mathematical literature since 1868. He found that the number of mathematics publications per year (measured by the counts of titles abstracted) has grown from 800 to 13000, at an average compound interest of about 2.5% per year, doubling about four times a century. Similarly, he found that cumulated growth of mathematics titles grew from 41,000 in 1867 to 41,900 by the end of 1965, and followed an exponential model, with few deviations only during world wars. Jean Tague [6], *et al* studied the application of linear, exponential, and logistic models in the growth of literature as listed and covered in indexing and abstracting services in science and technology and social sciences. Efthomidis [7] analysed the growth of OPAC literature from 1970 to 1985, and showed that it could be a good example of logistic growth. Wolfram [8] et al tried linear, exponential and power models to the growth of 20 years publications in several databases belonging to science and technology, social sciences and humanities. According to them, the popular models of exponential growth is no longer appropriate to represent literature growth of last twenty years. Instead, linear and power model performed quite well in such literature growth. The latter growth model, in particular was found to perform best, because it has the advantage of modeling the growth behavior of both linear and exponential models. Egghe and Rao [9] went a step further and reanalysed the Wolfram data and explored the applications of few growth models in all the data sets of Wolfram and observed that the power model explains very well the growth of science and technology literature. The Gompertz model, on the other hand, explains very well the growth of social sciences literature. They suggested a novel method of classifying various growth models, on the basis of two new growth rate functions, which also provided a clue in the selection of growth model, which might fit in any given observed data.

3. CRITERIA FOR SELECTION OF MODELS

Once we have broadly defined the nature of growth models available in the literature, the next question arises as how to identify a particular model or models that may possibly fit in a given growth data of a discipline or speciality. Egghe and Rao [9] have suggested a methodology by which one can get a clue in this direction. They have defined two new growth rate functions. These two functions are defined as:

$\alpha_1(t) = f(t+1)/f(t)$

$\alpha_2(t) = f(2t)/f(t)$ for $t = 1,2,3,4...$

$\alpha_1(t)$ or α_1 is the first growth rate function and α_2 is the second growth rate function. The function α_2 in fact compares the growth after double time period (from t to $2t$). The relation between α_1 and α_2 is given by the following equation:

$\alpha_2(t) = \alpha_1(2t-1)\ \alpha_1(2t-2)..\ \alpha_1(t).....$

If there were in total N observations (i.e. t=0, 1, 2, N-1), there would be N-1 values for α_1 and N/2 values for α_2. Egghe and Rao[9] have also derived and plotted graphs of α_1 and α_2 for the following growth models: exponential, logistic, power, Ware, and Gompertz. They have also given plots of α_1 and α_2 of the various models in terms of the following classification: (1) increasing, (2) constant, (3) decreasing, (4) increasing and then decreasing.

Using the criteria suggested above and the trends obtained through plotting the first growth rate functions of the observed data sets, we have selected three models, namely exponential, logistics, and power model for our analysis and studied their application in the growth of world research papers in physics, chemistry and electrical & electronics engineering.

4. NATURE OF GROWTH MODELS

Many scholars have tried to fit different type of growth curves to data on publications and authors. Few models, which are likely to fit in the observed data and their characteristics, are briefly introduced as follows:

4.1 Exponential Growth Model

Exponential growth represents an increase with a fixed proportion of total population for each unit of time, expressed in percentages or doubling time. It also assumes a constant growth rate. Mathematically this function is represented as follows. A graphical representation of this function is given in figure 1 (Figures 1 to 5 are taken from source: Egghe, L & Ravichandra Rao, I.K. Classification of growth models *Scientometrics*, 1992, **25**, 5-46).

$f(t) = a_2 * e^{-a1t}$ (Note: a_1 is written as a1 also)

$df(t)/dt = a_1 * f(t)$

where f(t) is cumulated publications at time t and $a_1 > 0$ is a constant. The term a_2 represents number of publications at time t = 0. The constant growth rate from time t to t+1 is e^{a1}.

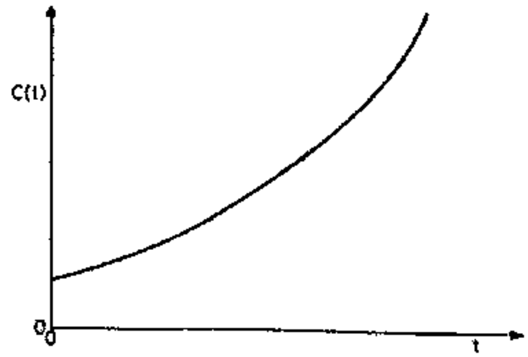

Figure 1. Exponential growth

4.2 Logistic Growth Model

Logistic growth assumes that the growth rate is proportional to the product of present size and future growth. This cumulative curve takes a S-shaped pattern. Compared to the exponential growth, here an upper limit to the growth curve exists. It follows a slow inception, followed by steep growth and then a long period of saturation. The non-cumulative curve shape is symmetrical about its point of infection. A graphical representation of this function is given in figure 2. Mathematically this function is represented by the following equations:

$f(t) = F/[1+\{F/f(0)-1\}e^{a1\ t}]$

Where F is an upper limit or asymptotic value to the growth curve and $a_1 > 0$.

The differential equation for the above can be written as:

$df(t)/dt = a_1\ [1 - f(t)/F\ ]*.\ f(t)$

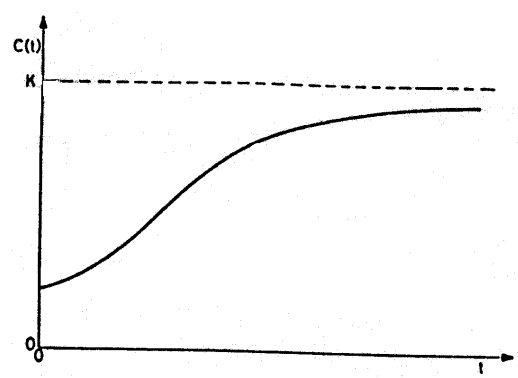

Figure 2. Logistic growth

4.3 Power Model

This model is represented by the equation:

$f(t) = a_0 + a_1 * t^{a2}$ (Note: a_2 is written as a2 also)

Where a_0, a_1 > 0. The parameter a_2 defines the shape of the curve. Three different types of curves are formed depending on the following values of a_2:

For $0 < a_2 < 1$, the curve is concave,

For $a_2 = 1$, the curve is linear,

For $a_2 > 1$, the curve is convex.

A graphical presentation of this function is given in figures 3 to 5. The differential equation for the power model is:

$$df(t)/dt = [(a_2/t) * f(t)] - (a_0 * a_2)/t$$

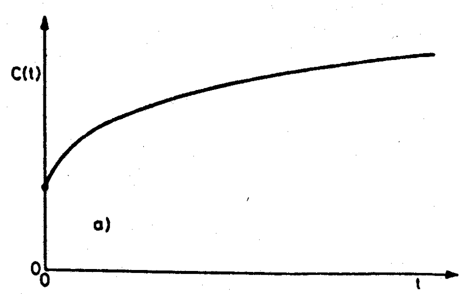

Figure 3. Power model ($0<\gamma<1$)

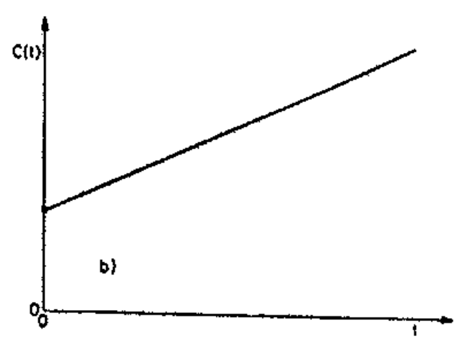

Figure 4. Linear model ($\gamma=1$)

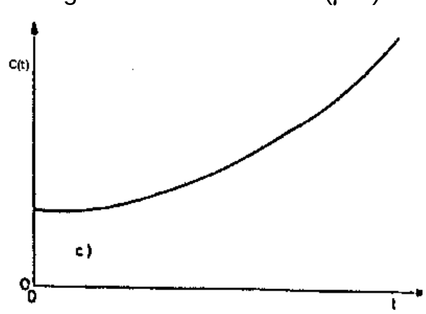

Figure 5. Power model ($\gamma>1$)

5. DATABASES AND METHODOLOGY USED

For studying the growth of publications in a discipline, the sources normally used for getting data are printed bibliographies, indexing and abstracting services, and computerised databases (available in CD-ROM media). There are, however, few sources available, which have a long history of coverage of developments of a particular discipline or speciality.

It is necessary to have a long time series data on the growth of development of research field to study the application of growth models in their growth data effectively. Keeping this in mind, we have selected the following three research fields for analysis:

(a) Physics research output as reflected in *Physics Abstracts,* from 1907 to 1994 (referred as Phy or Physics).

(b) Chemical sciences research output as reflected in *Chemical Abstracts*, from 1907 to 1994 (referred as Chem or Chemistry) and

(c) Electrical & Electronic Engineering research output as reflected in *Electrical and Electronic Abstracts*, from 1907 to 1994 (referred as Engg or Engineering).

Few characteristics of these source databases are given in Table 1. The methodology used in the paper is described in the following steps:

(a) The two growth rate functions have been calculated from the annual growth data of publications of the three research fields under consideration.

(b) The growth rate functions are derived, plotted and matched with the growth rate functions of the known growth models. The nature of growth model, which possibly may fit in the cumulative growth data on publications of the three research fields, is identified.

(c) The actual fitting of the different growth models is explored in the cumulative growth data of publications for the three research fields.

(d) The fitting of the specific growth models is matched with the results suggested by growth rate functions. To what extent there are similarities or dissimilarities in the two approaches are then discussed.

Table 1. Characteristics of the databases studied

Subject	Initial Publications	Cumulative Publications	Data Span
Phy	2132	174237	1907-1994
Chem	11847	15517854	1907-1994
Engg	1474	187725	1907-1994

6. ANALYSIS

The results obtained from the application of the three models in the growth of physics, chemistry, and electrical and electronics literature are presented and discussed as follows:

6.1 Exponential Model

The model was applied to the growth of literature in the field of physics, chemical sciences, and electrical and electronics engineering. The parameter values obtained from the application of this model are given in Table 2. A graphical fit of the observed and fitted values obtained from the application of this model in these three data files are shown in figures 6 to 8.

Table 2. Parameter values obtained from application of Exponential Model

Subject	Model Parameters	
	F	a_1
Phy	9.61(0.72)	0.068(0.001)
Chem	192.10(7.20)	0.050(0.000)
Engg	5.67(0.36)	0.067(0.001)

The fit statistics obtained mainly in terms of R^2 and F values from the application of the model in the three subjects are shown in the following Table 3.

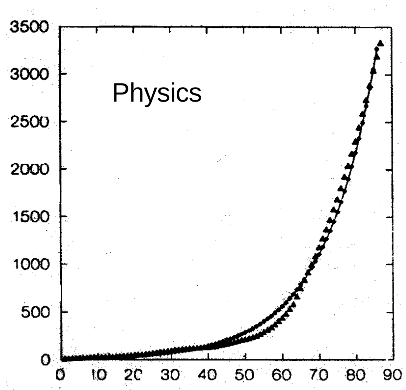

Figure 6.

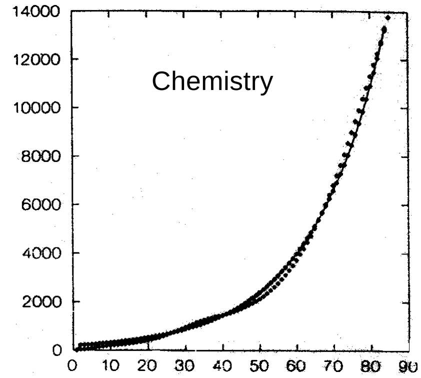

Figure 7.

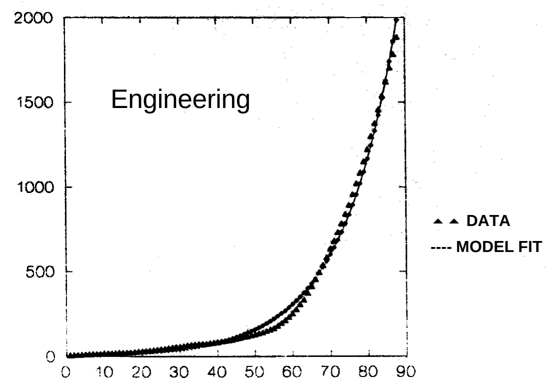

Figure 8.

Table 3. Fit statistics obtained from the application of Exponential Model

Subject	Fit statistics	
	R^2	F
Phy	0.994	9806.53
Chem	0.996	21382.67
Engg	0.995	13732.08

6.2 Logistic Model

The model was applied to the growth of literature in the field of physics, chemical sciences, and electrical & electronic engineering. The parameter values obtained from the application of this model are given in Table 4. A graphical fit of the observed and expected values obtained from the application of this model in these three data files are shown in figures 9 to 11.

Table 4. Parameter values obtained from the application of Logistic Model

Subject	Model Parameters	
	F	a_1
Phy	6796.00(149.4)	0.092(0.000)
Chem	18839.67(608.8)	0.098(0.001)
Engg	3907.70(100.3)	0.089(0.000)

The fit statistics in terms of R^2 and F values from the application of the model in the three subject fields are shown in Table 5.

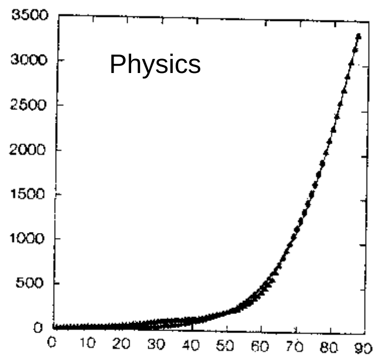

Figure 9.

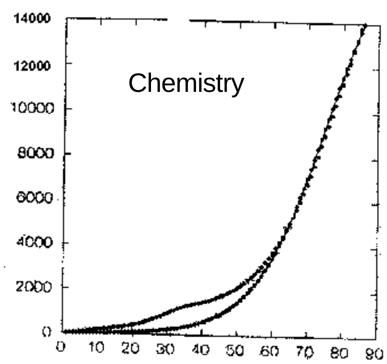

Figure 10.

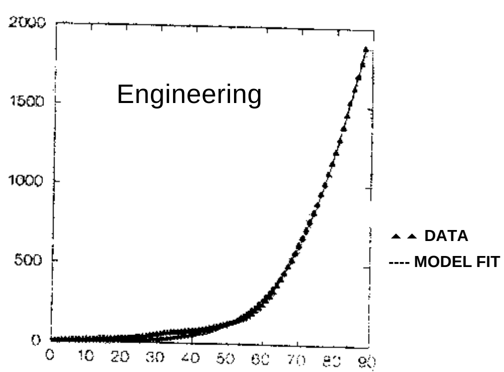

Figure 11.

Table 5. Fit statistics obtained from the application of Logistic Model

Subject	Fit Statistics	
	R^2	F
Phy	0.999	45511.46
Chem	0.986	5289.17
Engg	0.999	43116.07

6.3 Power Model

The model was applied to the growth of literature in the physics, chemistry, and electrical & electronics engineering. The parameter values obtained from the application of this model are given in table 6. A graphical fit of the observed and fitted values the models in these data files are shown in figures 12 to 14.

Table 6. Parameter values obtained from application of Power Model

Subject	Model Parameters		
	a_0	a_1	a_2
Phy	30.00(7.4)	0.34*10-6	5.15(.06)
Chem	396.24(39.8)	0.63*10-3	3.8(0.05)
Engg	23.41(3.4)	0.21*10-6	5.1(0.05)

The fit statistics obtained mainly in terms of R^2 and F values from the application of the model are shown in Table 7.

Table 7. Fit statistics obtained from application of Power Model

Subject	Fit statistics	
	R^2	F
Phy	0.997	15806.02
Chem	0.986	18176.54
Engg	0.999	21306.87

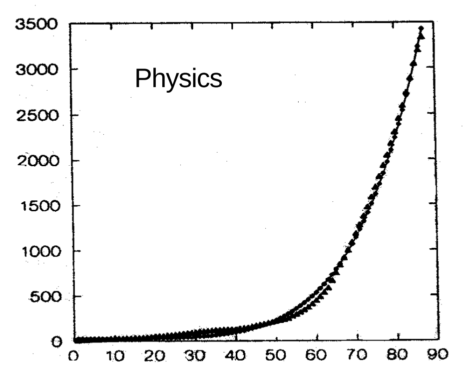

Figure 12.

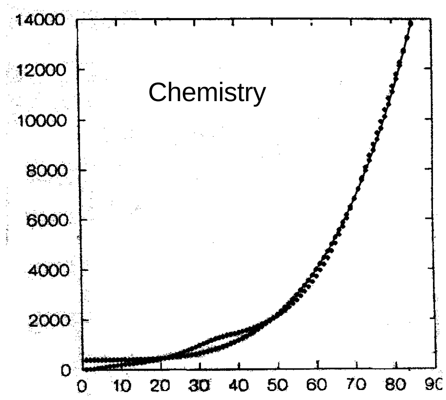

Figure 13.

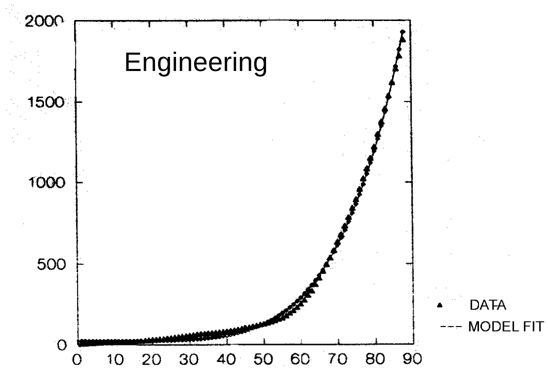

Figure 14.

7. DISCUSSION AND RESULTS

When the first growth rate function α_1 was plotted for the physics, chemistry, and electrical & electronics engineering data files (figure 15 to 17), it showed a decreasing trend for all three data files, giving an indication that power or logistic model may fit to the data. The plots of second growth rate function α_2 (figure 18 to 20) showed an increasing trend for all three cases, indicating that either of the two models namely exponential or power model ($a_2 > 1$, $a_0 > 0$) may fit to the data. No combination of α_1 and α_2 from the list given in the paper could be matched to the corresponding model. Since α_1 is decreasing in all cases and α_2 is increasing in all of them, the power model with parameters $a_0 > 0$, $0 < a_2 < 1$ should be the preferred one if one goes by growth rate formulations. But in the present case, the selected (power) model gave $a_2 > 1$ and not $a_2 < 1$.

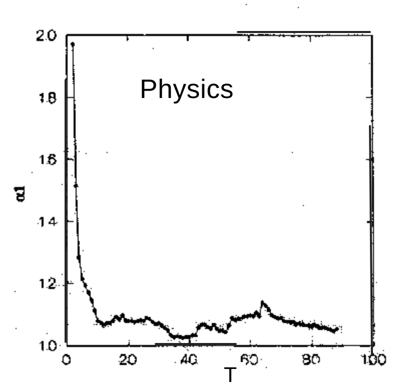

Figure 15. α_1 for Physics data file

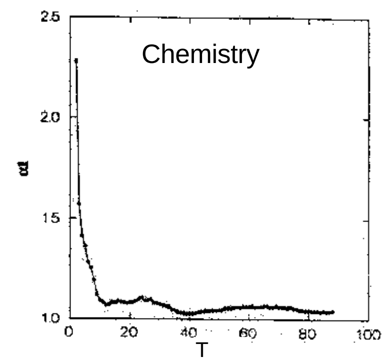

Figure 16. α_1 for Chemistry data file

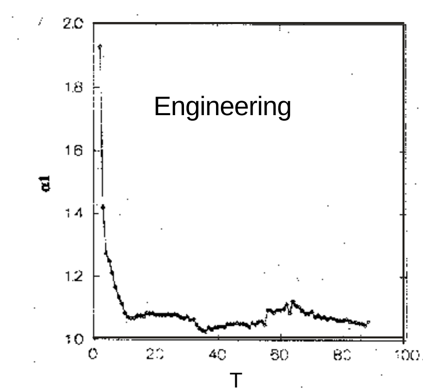

Figure 17. α_1 for Engineering data file

The coefficient of determination R^2 of most model fits is comparable for Physics, Engineering and Chemistry. Exponential model gives slightly better fit. The results obtained in terms of fit statistics from the application of the three models in physics data file indicate that logistic model

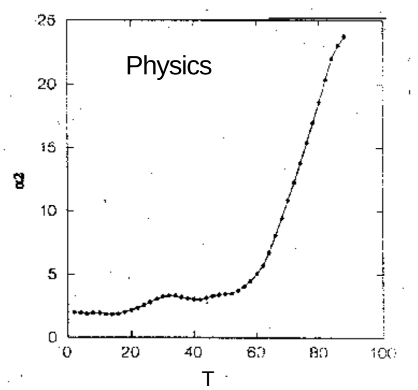

Fig 18. α_2 for Physics data file

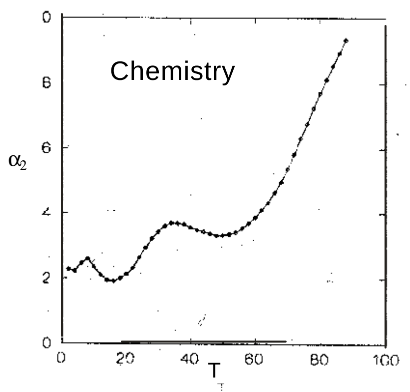

Fig 19. α_2 for Chemistry data file

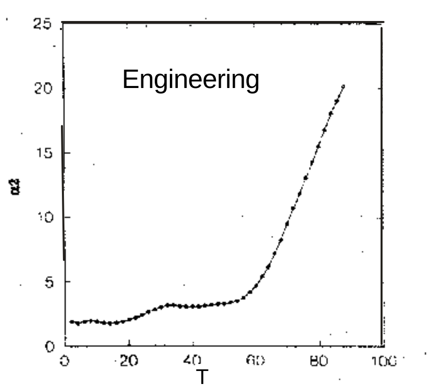

Fig 20. α_2 for Engineering data file

(R^2=0.999), power model (R^2=0.997), and exponential model (R^2=0.994) have all shown the good fits in the data. The good performance obtained in case of power model is in full agreement with the indications or clues given by both growth rate functions. However, it is quite difficult to say why the logistic model (where only the first growth rate function has given a clue) given best performance and the exponential also shown good performance (where the second growth rate function has given an indication). This is in partial agreement with the suggested indicators given by growth rate functions.

According to the results obtained in terms of fit statistics from the application of the three models in chemistry data file, exponential model (R^2=0.996) is observed to fit best to the data file. The good performance obtained in case of exponential model is in agreement with the indications obtained through the only second growth rate function α_2.

As can be seen from the results of fit statistics of the three models in engineering data file, the power model (R^2=0.999) and logistic model (R^2=0.999) are observed to fit best in the data file. The results obtained in case of power model are in full agreement with the indications (given by both growth rate functions). In case of logistic model, the results are in partial agreement with the indications given (by only first growth rate function α_1).

8. CONCLUSIONS

According to Egghe and Rao, the two growth rate functions α_1 and α_2 proposed by them should give sufficient indication of the type of model which should fit best in a given data set. It is also implied that the best model that fits to a given data will be that model where both the growth rate functions indicate the same model selection. In our analysis, where we have applied three models, namely exponential, logistic and power models, the following conclusions are made:

(i) The power model fits best to all three data files. This is in full agreement with the best model suggested by both the growth rate functions but for $\alpha_2(t)$ the trend is not increasing and then decreasing. We have a decreasing trend of $\alpha_2(t)$ for all three data files which suggests power model with $a_2<1$ but we obtained $a_2>1$ in all three data files when power model was applied;

(ii) All three models fit well to the Engineering data file. These results are in full agreement in case of power model and in partial agreement in case of logistic model; and

(iii) Power and the exponential model fit best in the chemistry data file. This is in partial agreement with the best model suggested by the growth rate functions.

Based on the analysis and observations made, it can be concluded that the results obtained in our analysis are not found to be in full agreement with the expected results, which should come through the methodology suggested, by Egghe and Rao. We therefore suggest that a more rigorous research on growth of publications of different research fields should be carried out to test the utility of these two growth rate functions in selecting the right growth model which may best fit in a given data.

REFERENCES

1. Cole, F.J. The history of comparative anatomy, Pt 1. Statistical analysis of the literature. *Science Progress*, 1917, **11**, 578-596.
2. Wilson, P.W. & Fred, E.B. The growth curves of scientific literature: Nitrogen fixation by plants. *Scientific Monthly*, 1935, **41**, 241-50.
3. Solla de Price, Derek. Science since Babylon. Yale University Press, New Haven. 1975.
4. Solla de Price, Derek. Little science, big science. Columbia University Press, New York, 1966.
5. May, K.O. Quantitative growth of the mathematical literature. *Science*, 1966, **154**, 1672-73.
6. Tague, Jean; Beheshti, Jamshid & Rees-Potter, Lorana. The law of exponential growth: Evidence, implications and forecasts. *Library Trends*, 1981, **30**, 125-49.
7. Efthomidis, E.N. The growth of the OPAC literature. *Journal of the American Society for Information Science*, 1990, **41,** 342-47.
8. Wolfram, Dietmar; Chu, Clara M. & Xin, Liu. Growth of knowledge: Bibliometric analysis using online databases data. *In* Informetrics, *edited by* L. Egghe & R Rousseau. Elsevier, Amsterdam, 1990, 355-72.
9. Egghe, L & Ravichandra Rao, I.K. Classification of growth models based on growth rates and its applications. *Scientometrics*, 1992**, 25**, 5-46.

Contributors: **Mrs Praveen Sharma** is Technical Officer at National Institute of Science, Technology and Development Studies (NISTADS), Dr K S Krishnan Marg, New Delhi 110 012.

Shri B. M. Gupta is Scientist F at NISTADS, New Delhi 110 012.

Shri Suresh Kumar is Technical Officer at NISTADS, New Delhi 110 012.