A. MENDELSON
S. S. MANSON
Lewis Research Center,
NASA, Cleveland, Ohio

The Extrapolation of Families of Curves by Recurrence Relations, With Application to Creep-Rupture Data

A method using finite-difference recurrence relations is presented for direct extrapolation of families of curves. The method is illustrated by applications to creep-rupture data for several materials and it is shown that good results can be obtained without the necessity for any of the usual parameter concepts.

Introduction

THE PROBLEM of extrapolating sets of data finds many applications in engineering practice. An example which will be treated in detail in this report is the extrapolation of short-time creep-rupture data to predict long time life.

The present day techniques for the extrapolation of creep-rupture data are based primarily on the time-temperature parameter concept. It is generally assumed that all creep-rupture data for a given material can be correlated to produce a single master curve using as one of the co-ordinates a parameter involving a combination of time and temperature. This master curve which can be obtained using primarily short time data could then be used for extrapolation to long times. The three best known parameter methods are the Larson-Miller, Manson-Haferd, and Dorn parameters [1 to 3].[1]

These parameter methods have the great advantage that, at least in theory, they require only a relatively small number of data to establish the required master curve. In particular, the development of the objective least squares method in reference [4] makes it possible to obtain the optimum constants contained in the parameters described in references [1 to 3] without plotting and crossplotting the data and with the use of a minimum of judgment on the part of the analyst.

On the other hand, an alternative to the parameter methods would be advantageous for several reasons. First, these parameter methods presuppose a knowledge of the existence of some form of functional relationship between the variables. Thus it is assumed that, at a constant nominal (initial) stress, the logarithm of rupture time is linearly related to the temperature, for the Manson-Haferd parameter, and to the reciprocal of the absolute temperature for the Larson-Miller and Dorn parameters. For most materials it has been found that such relationships are valid in the range of variables investigated. In particular, the Manson-Haferd parameter has been shown [5 and 6] to give best results in the majority of cases. However, occasionally materials are encountered for which none of the parameters are valid. For example, plots of the logarithm of rupture time against temperature or reciprocal of absolute temperature at constant nominal stress will not result in straight lines for crucible 422 steel [7]. In these cases, modification of the parameter methods is necessary [7].

Secondly, the parameter methods in their usual form give equal weight to all the data, whether they be low time or high time data. For extrapolating to long times more weight should be given long time data. The method to be presented attempts to overcome these objections as well as providing an independent check on the parameter methods.

The method presented in this paper enables one to extrapolate directly a family of curves such as creep-rupture curves without assuming any particular parametric form. The method is based on the assumption that a family of curves such as a set of isothermal curves can be represented approximately by a finite-difference recurrence relation. Once the coefficients of this recurrence relation are determined, it is a simple matter to extrapolate each member of the family individually. The method has the advantage that it does not require an explicit knowledge or assumption as to the analytic character of the curves or as to what parametric form to use to correlate the data, although implicitly it assumes any of the broad class of functions which satisfy linear finite-difference equations. It is thus much more general in nature than the parametric methods previously discussed and can be used by itself to extrapolate creep-rupture data or as an independent check on any of the parameter methods. In addition, the method treats the data in such a way as to give more weight to the higher time data than to the lower time data.

Several examples of the use of the method are described in detail in the body of the paper, and the results of this type of extrapolation on the creep-rupture curves of several materials are presented. The use of this method is, of course, not restricted to creep-rupture data, and several examples of other type curves such as Bessel functions, Legendre polynomials, and power functions, are presented.

Method

The usual method for extrapolating a curve is to assume some explicit functional relation between the ordinate y and the abscissa x. The present method assumes instead a linear relation between successive equally spaced ordinates without directly introducing the abscissa x. Thus consider a section of a curve such as the solid line from 0 to 60 deg in Fig. 1. This section is actually part of a sine curve; however, assume that this is not known and that it is desired to extrapolate the curve. This can be done in several ways and, as previously mentioned, the one that is generally used is to fit the known curve by some assumed function and to then use this function for extrapolation. For example, a polynomial such as a parabola or cubic can be fitted to the curve. While assumptions of this type usually yield very good values when interpolating they are usually poor in extrapolating more than a short distance beyond the end of the curve.

[1] Numbers in brackets designate References at end of paper.

Contributed by the Metals Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 7, 1959. Paper No. 59—A-155.

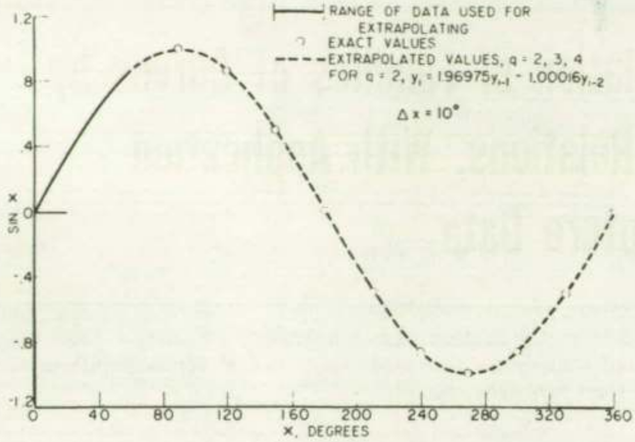

Fig. 1 Extrapolation of sin *x*

Table 1 Extrapolation of sin *x*

x, deg	sin x, extrapolated	sin x, exact
0		0
10		0.1736
20		0.3420
30		0.5000
40		0.6428
50		0.7660
60		0.8660
70	0.9397	0.9397
80	0.9840	0.9848
90	1.000	1.000
100	0.9848	0.9848
110	0.9396	0.9397
120	0.8658	0.8660
130	0.7657	0.7660
140	0.6423	0.6428
150	0.4993	0.5000
160	0.3411	0.3420
170	0.1725	0.1736
180	−0.0014	0
200	−0.3439	−0.3420
220	−0.6451	−0.6428
240	−0.8685	−0.8660
260	−0.9871	−0.9848
280	−0.9864	−0.9848
300	−0.8666	−0.8660
320	−0.6421	−0.6428
340	−0.3401	−0.3420
360	0.0031	0

If, however, the sine function were chosen to fit this curve, then in the foregoing case it would obviously give excellent results. Thus, although for interpolating purposes it is possible to fit a curve by many different functions, the extrapolated values will, in general, be very sensitive to the function used.

The present method avoids this difficulty by working directly with successive values of the ordinates of the curve rather than assuming any functional relation between the co-ordinates. A linear relation is sought between successive ordinates on the curve of the following form

$$y_i = A_1 y_{i-1} + A_2 y_{i-2} + \ldots + A_q y_{i-q} \tag{1}$$

where y_i represents the value of the dependent variable at the end of the ith interval, and the A_j are constants to be determined. Equation (1) gives the value of the ordinate at the ith station in terms of the values at the q previous stations. Once the value at the ith station is computed, this value is used together with the $q - 1$ values preceding it, to calculate the value at the $(i + 1)$st station. In this way successive ordinates can be computed in terms of the previous ordinates.

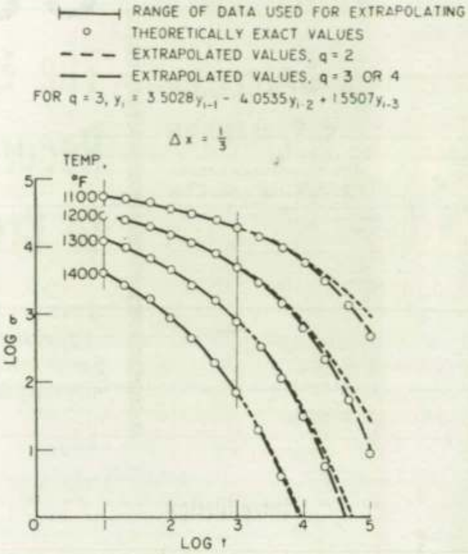

Fig. 2 Results for theoretical stress-rupture curves given by

$$\log \sigma = 5.131 - \left[\frac{T - 600}{81.8(9.5 - \log t)}\right]^{2.9326}$$

Equation (1) can, for example, be obtained from a differential equation. Thus, consider the differential equation associated with the sine function

$$\frac{d^2y}{dx^2} + y = 0 \tag{2}$$

writing (2) in finite difference form gives

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2} + y_i = 0$$

or

$$y_{i+1} = [2 - (\Delta x)^2]y_i - y_{i-1} \tag{3}$$

which is an equation of the form of (1) relating the ordinates of the sine curve to each other. In the general case, of course, the differential equation for the curve is not known. The coefficients A_j appearing in (1) can, however, be obtained from the ordinates of the curve by least squares as will be subsequently described.

One additional assumption is necessary in order to apply the present method to a family of curves. Just as equation (3) is valid for any one of an infinite family of sine curves of the same frequency (the members of the family can have different amplitudes or different phase angles), so it will be assumed in general that all the members of the family of curves to be extrapolated are represented by the same recurrence relation of the form of (1). A more extensive discussion of the origin and validity of equation (1) is given in the Appendix.

In order to determine the coefficients A_j of equation (1) use is made of least squares. Consider a family of curves to be extrapolated as for example in Fig. 2. In general, these will be obtained by fairing by eye the best curves through the data points. Choose n equally spaced stations for each member of the family. (These stations need not be chosen at the same values of the abscissas for every member of the family, but the spacing must be the same.) Choose the order q of the recurrence relation to be used (usually, q = 3 or 4). For each member of the family write a recurrence relation of the form of (1), as follows

$$
\left.
\begin{aligned}
y_k^{(1)} &= A_1 y_{k-1}^{(1)} + A_2 y_{k-2}^{(1)} + A_3 y_{k-3}^{(1)} + \ldots + A_q y_{k-q}^{(1)} \\
y_k^{(2)} &= A_1 y_{k-1}^{(2)} + A_2 y_{k-2}^{(2)} + A_3 y_{k-3}^{(2)} + \ldots + A_q y_{k-q}^{(2)} \\
&\vdots \\
y_k^{(m)} &= A_1 y_{k-1}^{(m)} + A_2 y_{k-2}^{(m)} + A_3 y_{k-3}^{(m)} + \ldots \\
&\qquad + A_q y_{k-q}^{(m)}
\end{aligned}
\right\} \quad (4)
$$

where $y_k^{(p)}$ represents the value of the dependent variable at the end of the kth interval and the superscript represents the member of the family under consideration. It should be noted that q cannot be greater than the number of intervals chosen.

The A_j are now chosen so that the recurrence relation (4) will best represent the data in the given range. To do this the usual least squares method of minimizing the sum of the squares of the residuals is used. Thus, the following sum is here minimized

$$
S = \sum_{p=1}^{m} \sum_{k=q+1}^{n} (y_k^{(p)} - A_1 y_{k-1}^{(p)} - A_2 y_{k-2}^{(p)} - \ldots - A_q y_{k-q}^{(p)})^2 \quad (5)
$$

Differentiating S with respect to each of the A_j and setting the results equal to zero, to find the optimum values of the A_j results in the following set of simultaneous equations for the coefficients.

$$
\begin{aligned}
r_{11}A_1 + r_{12}A_2 + \ldots + r_{1q}A_q &= r_{10} \\
r_{21}A_1 + r_{22}A_2 + \ldots + r_{2q}A_q &= r_{20} \\
&\vdots \\
r_{q1}A_1 + r_{q2}A_2 + \ldots + r_{qq}A_q &= r_{q0}
\end{aligned}
\quad (6)
$$

where the coefficients r_{ij} are given by

$$
r_{ij} = \sum_{p=1}^{m} \sum_{k=q+1}^{n} y_{k-i}^{(p)} y_{k-j}^{(p)} = r_{ji} \quad (7)
$$

Once the A_j are determined from (6), each member of the family can be extrapolated using (4), as previously explained.

Examples

As a first illustration consider the simple case of extrapolating a single sine curve as shown in Fig. 1. The values of the sine function are assumed to be given between 0 and 60 deg and it is desired to extrapolate by the method presented herein. Choose seven equally spaced stations so that

$$
\begin{aligned}
y_1 &= \sin 0° = 0 & y_5 &= \sin 40° = 0.6428 \\
y_2 &= \sin 10° = 0.1736 & y_6 &= \sin 50° = 0.7660 \\
y_3 &= \sin 20° = 0.3420 & y_7 &= \sin 60° = 0.8660 \\
y_4 &= \sin 30° = 0.5000
\end{aligned}
$$

Let $q = 2$ then from (1)

$$
y_i = A_1 y_{i-1} + A_2 y_{i-2}
$$

and from (7)

$$
r_{11} = \sum_{k=3}^{7} y_{k-1}^2 = y_2^2 + y_3^2 + y_4^2 + y_5^2 + y_6^2
$$

$$
r_{12} = r_{21} = \sum_{k=3}^{7} y_{k-1} y_{k-2} = y_2 y_1 + y_3 y_2 + y_4 y_3 + y_5 y_4 + y_6 y_5
$$

$$
r_{22} = \sum_{k=3}^{7} y_{k-2}^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2
$$

$$
r_{10} = \sum_{k=3}^{7} y_{k-1} y_k = y_2 y_3 + y_3 y_4 + y_4 y_5 + y_5 y_6 + y_6 y_7
$$

$$
r_{20} = \sum_{k=3}^{7} y_{k-2} y_k = y_1 y_3 + y_2 y_4 + y_3 y_5 + y_4 y_6 + y_5 y_7
$$

Substituting the values for y_1 through y_7 gives

$$
\begin{aligned}
r_{11} &= 1.3970488 \\
r_{12} = r_{21} &= 1.0441560 \\
r_{22} &= 0.81029280 \\
r_{10} &= 1.7075120 \\
r_{20} &= 1.2463024
\end{aligned}
$$

Substituting into equation (6) and solving gives

$$
\begin{aligned}
A_1 &= 1.96975 \\
A_2 &= -1.00016
\end{aligned}
$$

Therefore

$$
y_i = 1.96975 y_{i-1} - 1.00016 y_{i-2}
$$

Thus at 70 deg

$$
\begin{aligned}
y_8 &= 1.96975 y_7 - 1.00016 y_6 \\
&= 1.96975(0.8660) - 1.00016(0.7660) \\
&= 0.9397
\end{aligned}
$$

continuing this way the values of $y = \sin x$ have been computed up to 360 deg and compared with the exact values in Table 1 and Fig. 1.

As a second example, a family of curves was chosen simulating a set of stress-rupture data. It was assumed that an equation can be written for this family similar in form to those proposed in reference [7].

$$
y \equiv \log \sigma = 5.131 - \left[\frac{T - 600}{81.8(9.5 - \log t)}\right]^{2.9326} \quad (8)
$$

The curves represented by equation (8) are plotted in Fig. 2 for four temperatures. Again take seven stations ($n = 7$) equally spaced from $\log t = 1$ to $\log t = 3$ as shown. The values of $y_j^{(p)}$ for these stations as computed from equation (8) are shown in Table 2. The r_{ij} are now computed from equation (7). In this case we have a double summation. The products for all four curves ($m = 4$) must be added together. Thus, for example, taking $q = 2$:

$$
\begin{aligned}
r_{12} = {} & y_2^{(1)}y_1^{(1)} + y_3^{(1)}y_2^{(1)} + y_4^{(1)}y_3^{(1)} + y_5^{(1)}y_4^{(1)} + y_6^{(1)}y_5^{(1)} + \\
& y_2^{(2)}y_1^{(2)} + y_3^{(2)}y_2^{(2)} + y_4^{(2)}y_3^{(2)} + y_5^{(2)}y_4^{(2)} + y_6^{(2)}y_5^{(2)} + \\
& y_2^{(3)}y_1^{(3)} + y_3^{(3)}y_2^{(3)} + y_4^{(3)}y_3^{(3)} + y_5^{(3)}y_4^{(3)} + y_6^{(3)}y_5^{(3)} + \\
& y_2^{(4)}y_1^{(4)} + y_3^{(4)}y_2^{(4)} + y_4^{(4)}y_3^{(4)} + y_5^{(4)}y_4^{(4)} + y_6^{(4)}y_5^{(4)}
\end{aligned}
$$

Table 2 Values of $\log \sigma = 5.131 - \left[\frac{T - 600}{81.8(9.5 - \log t)}\right]^{2.9326}$

$\log t$ \ T	1100°	1200°	1300°	1400°
1.0	4.751	4.482	4.111	3.622
1.333	4.703	4.401	3.984	3.434
1.667	4.648	4.306	3.835	3.214
2.000	4.582	4.194	3.659	2.953
2.333	4.504	4.061	3.449	2.642
2.667	4.410	3.900	3.197	2.269
3.000	4.296	3.706	2.891	1.817

Once the r_{ij} are computed, equation (6) is solved for the A_j and the extrapolation is then carried out for each curve from the relations

$$y_i^{(1)} = A_1 y_{i-1}^{(1)} + A_2 y_{i-2}^{(1)}$$
$$y_i^{(2)} = A_1 y_{i-1}^{(2)} + A_2 y_{i-2}^{(2)}$$
$$y_i^{(3)} = A_1 y_{i-1}^{(3)} + A_2 y_{i-2}^{(3)}$$
$$y_i^{(4)} = A_1 y_{i-1}^{(4)} + A_2 y_{i-2}^{(4)}$$

The values obtained for this case were: $A_1 = 2.2052$, $A_2 = -1.2048$. The results for this calculation are shown in Fig. 2.

Results and Discussion

Fig. 1 and Table 1 show the results of extrapolating sin x by the method presented herein. It is seen that excellent results are obtained using only a second-order recurrence relation ($q = 2$). As a matter of fact, little or no improvement can be obtained by going to higher values. However, this is to be expected since the governing differential equation is of the second order. It is seen from equations (2) and (3) that a recurrence relation with q equal to 2 is sufficient to describe the sine function. As a matter of fact, taking Δx equal to 10 deg or 0.174533 radians gives from equation (3)

$$y_{i+1} = 1.96954 y_i - y_{i-1}$$

which is seen to be very close to the values previously obtained by the method presented herein.

Fig. 2 shows the results obtained for the theoretical set of

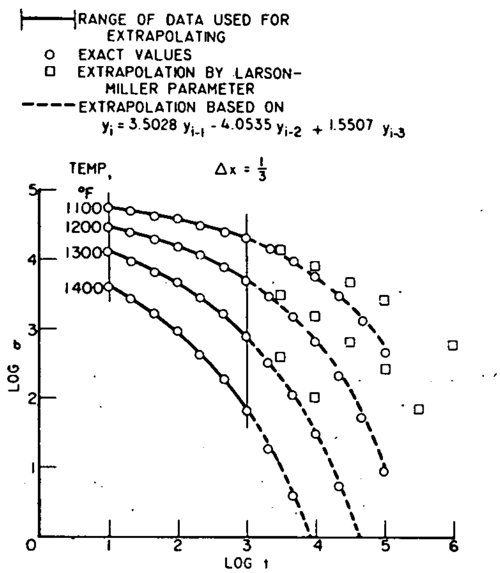

Fig. 3 Results of extrapolating stress-rupture curves based on Manson-Haferd parameter by recurrence relation and Larson-Miller parameter

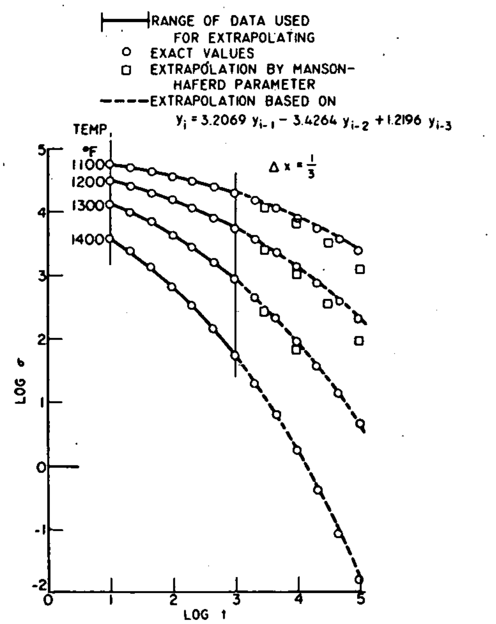

Fig. 4 Results of extrapolating stress-rupture curves based on Larson-Miller parameter by recurrence relation and Manson-Haferd parameter

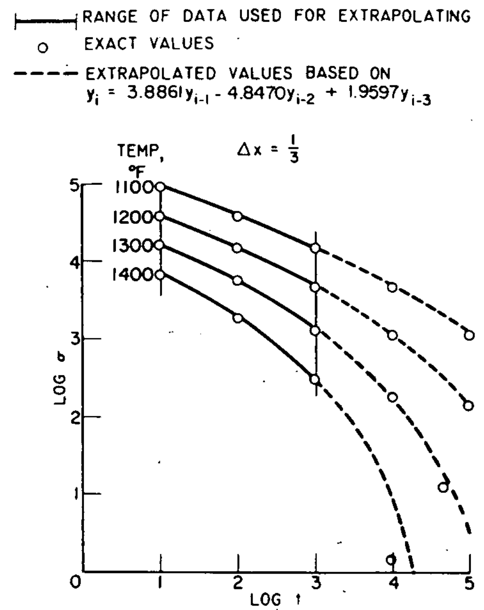

Fig. 5 Extrapolation of curves based on Dorn equation

$$e^{0.34(\log \sigma)^{2.5}} = 10^{-10} \frac{e^{\frac{60,000}{T+460}}}{t}$$

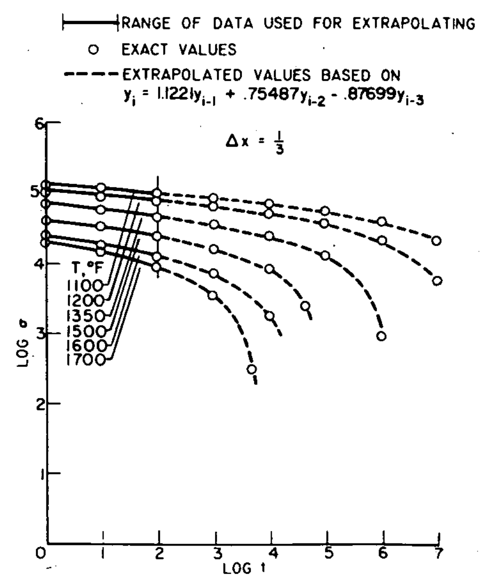

Fig. 6 Extrapolation of curves based on Conrad equation

$$t = 3.24 \times 10^{-9}\, e^{\frac{60,600}{T+460}}\, e^{-\frac{\sigma}{\sigma_0(T)}}$$

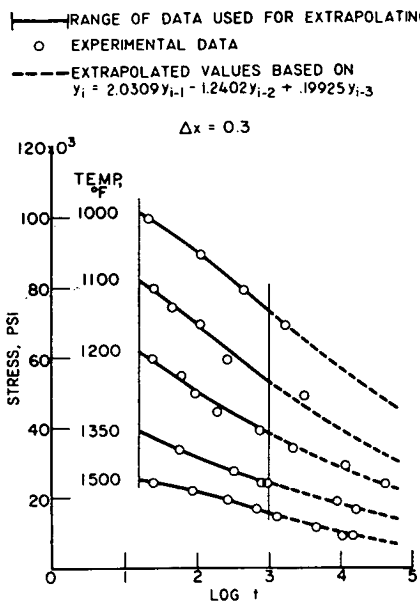

Fig. 7 Results of extrapolating alloy S-590. Data from reference [6].

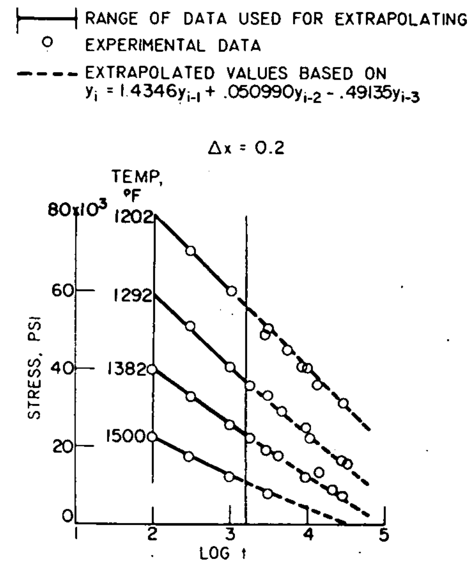

Fig. 8 Results of extrapolating alloy Nimonic 80A. Data from reference [6].

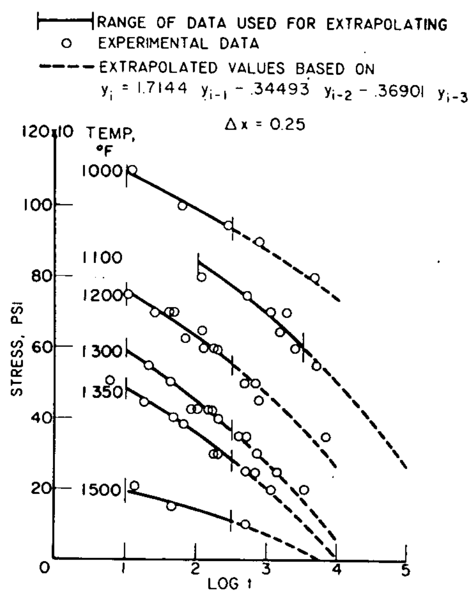

Fig. 9 Results of extrapolating alloy A-286. Data from reference [6].

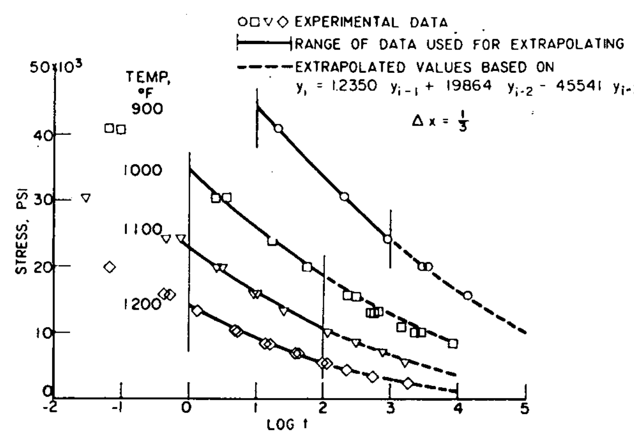

Fig. 10 Results of extrapolation for killed carbon steel. Data from reference [10].

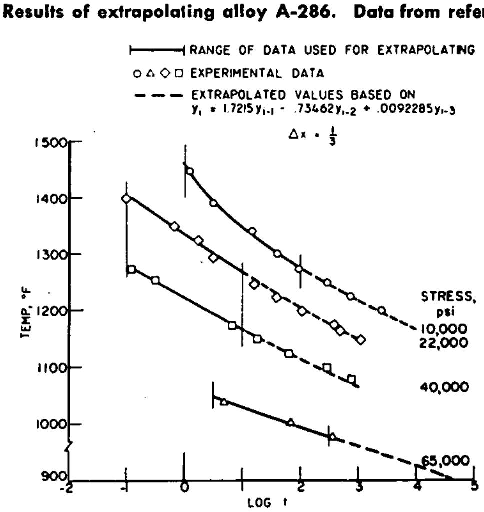

Fig. 11 Results for crucible 422 steel. Data from reference [7].

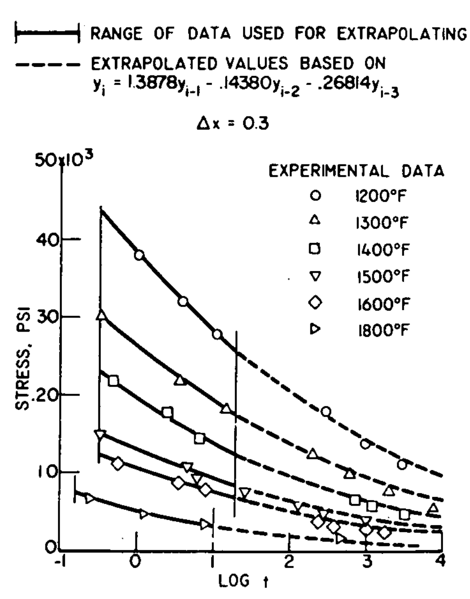

Fig. 12 Results of 18-8 stainless. Data from reference [2].

creep-rupture curves as given by equation (8). The extrapolations were made with q equal to 2, 3, and 4, and it is seen that there is no significant difference between the results obtained for q equal 3 and those obtained for q equal 4. In both cases the extrapolated values up to a hundred thousand hours shows excellent agreement with the theoretically exact values.

The generality of the present method for creep-rupture application was tested by trying it for several cases where one or another of the parameter methods could not be expected to give good results. Thus if one took a set of data given exactly by the Manson-Haferd parameter equation (8) and tried to correlate and extrapolate these data by the Larson-Miller parameter, one would not expect to obtain good results. This was done as shown in Fig. 3. Using the data exactly satisfying the Manson-Haferd parameter equation (8) in 10–1000 hour range, the best value of the Larson-Miller parameter C was computed by the method of reference [4]. The value of C was determined as 18, and a master curve was then constructed and the isothermals extrapolated as shown in Fig. 3. It is seen that as expected the extrapolation is very poor. However, using the same data between 10 and 1000 hours, the extrapolation obtained by the recurrence relation method gives excellent results as previously shown in Fig. 2 and repeated in Fig. 3.

A second calculation was then made using the reverse procedure. A set of data exactly satisfying the Larson-Miller parameter was generated by the equation

$$(\log t + 18)(T + 460) = 33{,}500(5.13 - \log \sigma)^{0.126} \quad (9)$$

The data obtained by this equation are very close to those obtained from equation (8) in the range between 10 and 1000 hours. These data were then used to obtain the best values of the constants for the Manson-Haferd parameter and to construct a master curve which was then used to extrapolate the original isothermals. The results are shown in Fig. 4 and as expected the extrapolation by the Manson-Haferd parameter for this case was not very good. However, again as can be seen the extrapolation obtained by the recurrence relation, using the same data, is excellent.

The calculations shown in Figs. 3 and 4 indicate the greater generality of the proposed method. The recurrence relation technique can be used whether the data are best fitted by the Larson-Miller parameter or the Manson-Haferd parameter. As a further test the method was applied to sets of data generated by use of the Dorn parameter and the recently proposed equation by Conrad [8]. The results are shown in Figs. 5 and 6, and it is seen that the agreement is very good.

In the previous examples the values of the ordinates required to calculate the r_{ij} by equation (7) were precisely known, since they were obtained from exact mathematical expressions such as equations (8) and (9). It might be expected therefore that these input data would reflect implicitly the shape of the curves even at a distance relatively far from the initial range of the data used. In practice, however, the data are given experimentally, and the input data for equation (7) are therefore much cruder than in the examples given. The extrapolation in this case might therefore not be quite so good as previously indicated, particularly if the curves have a large amount of curvature in the extrapolated range.

The method was therefore tried out using the actual creep-rupture data for six different materials. The actual data were plotted in the range to be used for extrapolating, and curves drawn by eye through these data. The values of the ordinates at seven equally spaced values of the abscissa were then read from these curves, and the coefficients of the finite difference recurrence relations obtained. The curves were then extrapolated and the results are shown in Figs. 7 through 12. It is seen that good extrapolation is obtained for all six materials.

The materials in Figs. 7, 8, and 9 were analyzed in reference [6], where a comparison was made between the Larson-Miller, Manson-Haferd, and Dorn parameters. In evaluating the various parameters, reference [6] uses the standard deviation defined by

$$D = \sqrt{\frac{1}{N} \Sigma d^2}$$

where D is the standard deviation, d is the deviation of each data point, and N is the total number of data points. Part of the results comparing the various parameters are reproduced in Table 3.

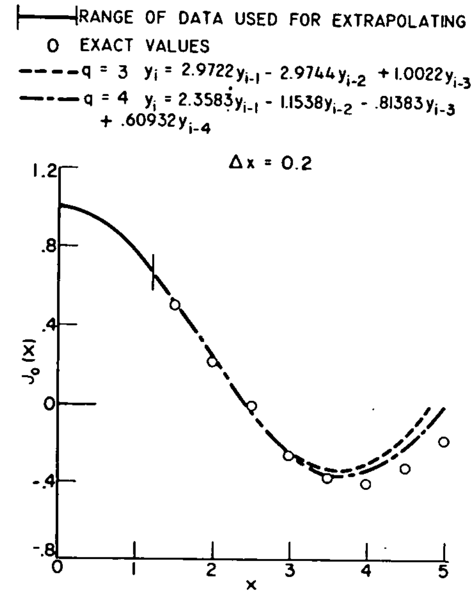

Fig. 13 Extrapolation of Bessel function, $J_0(x)$

Table 3 Standard deviations

Material	Larson-Miller	Dorn	Manson-Haferd	Recurrence relation
S-590	0.351	0.430	0.324	0.133
Nimonic 80A	0.385	0.235	0.091	0.045
A-286	0.207	0.172	0.134	0.148

In addition, the standard deviations using the method presented herein for extrapolation were also computed and added to the table. It is seen that the present method gives better results than any of the parameters shown.

Inspection of Figs. 7 through 12 shows that this method generally gives good results. However, a note of caution is in order. In order to use this method a sufficient number of low time data are needed for adequate definition of the basic curves to permit proper fairing. If such data are not available then the present method cannot be used. It should also be noted that the value of q or the number of terms appearing in the recurrence relation should not be made too large since it has been found that the larger the number of simultaneous equations to be solved the greater is the loss in significant figures in carrying out the calculations. However, the results for the examples and materials shown indicate that in general using a value of q equal to 3 is adequate and that very little can be gained by going to higher values of q.

It is obvious from the foregoing that this method is not restricted to families of creep-rupture curves. The sine function used as an example illustrates that other type curves can be extrapolated with equal ease. For example, Fig. 13 shows the results of applying this method to a Bessel function, Fig. 14 shows the results for a Legendre polynomial, and Fig. 15 shows a family given by $y = x^n$ with n equal to $^1/_2$, $^1/_3$, and $^1/_4$. In all these

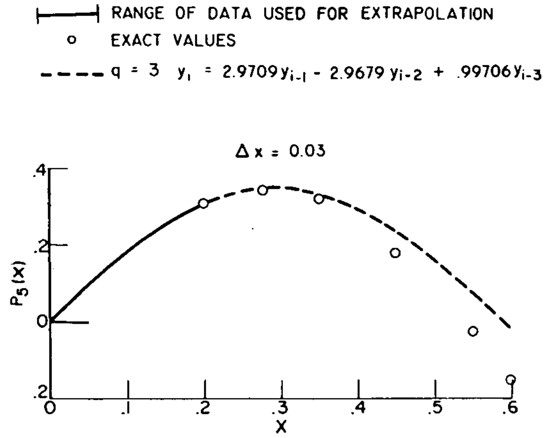

Fig. 14 Extrapolation of Legendre polynomial, $P_5(x)$

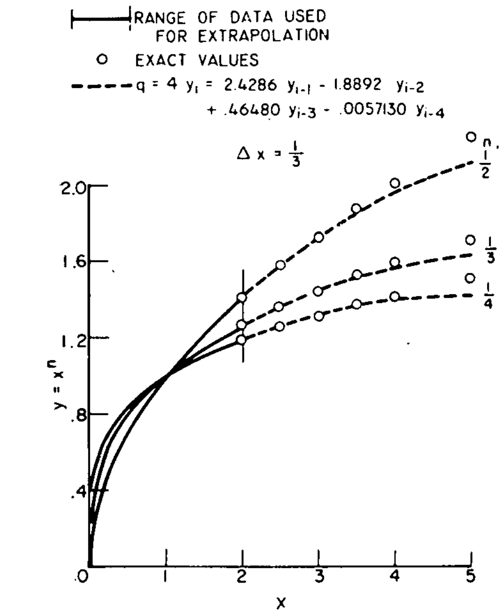

Fig. 15 Extrapolation of family of curves given by $y = x^n$

cases good results are obtained, even though as indicated in the Appendix one would not expect that these functions could be represented by finite-difference recurrence relations with constant coefficients. It is evident, therefore, that the applications of this method can be manifold.

APPENDIX

Origins of Finite-Difference Recurrence Relations

Consider any nth-order ordinary linear differential equation with constant coefficients. It is apparent that if each of the derivatives appearing in the differential equation is replaced by its finite difference equivalence, a finite difference recurrence relation is obtained relating $(n + 1)$ stations to each other. This has already been illustrated by the sine function discussed earlier. As another simple example consider the differential equation

$$\frac{d^3y}{dx^3} = 0 \tag{10}$$

in finite difference form this becomes

$$\frac{y_{i+1} - 3y_i + 3y_{i-1} - y_{i-2}}{(\Delta x)^3} = 0$$

or

$$y_{i+1} = 3y_i - 3y_{i-1} + y_{i-2} \tag{11}$$

which is the recurrence relation for a family of parabolas, the particular member of the family being determined by the starting values. In this particular case the relation (11) is exact since it is independent of the step size Δx.

It is apparent that, for every family of curves which are derivable as the solution of a linear differential equation with constant coefficients, there exists a corresponding recurrence relation. In the general case the recurrence relation can be said to approximate the actual function by a function which satisfies a linear differential equation with constant coefficients. Moreover, since the solution of an ordinary linear differential equation with constant coefficients or of the corresponding difference equation is generally given in terms of exponentials, it can be said that the actual function is being approximated by a sum of exponentials.

There is, however, a great difference between the use of the finite difference recurrence relation proposed herein and trying to fit the curves by a sum of exponentials. Thus consider this recurrence relation

$$y_i = A_1y_{i-1} + A_2y_{i-2} \tag{12}$$

and the corresponding exponential solution

$$y = B_1e^{\alpha_1 x} + B_2e^{\alpha_2 x} \tag{13}$$

Equations such as (13) have been used for interpolating purposes as, for example, in the Prony method [11]. Equation (12) contains two unknown constants which can be evaluated from four successive ordinates by writing down two such equations as (12) for two successive values of i. Equation (13), however, requires the evaluation of four constants: B_1, B_2, α_1, and α_2. Moreover, these constants will in general be complex numbers. The generation of the curve represented by (12) of course requires four values, A_1, A_2, y_{i-1}, and y_{i-2}, just as for (13).

This advantage of the use of (12) over (13) becomes even more evident if a family of curves is considered. Equation (12) remains the same with just two constants required, each member of the family being generated from its two starting values. The use of (13), however, requires the evaluation of a different set of constants B_1 and B_2 for each member of the family.

The major advantage of (12) over (13), however, lies in the ease with which least squares can be used to determine the unknown coefficients. Using (12) one can find the best values of the constants to fit all the data by conventional application of least squares procedures. This case is brought about by the fact that the least squares equations are linear in the unknown constants. An attempt to find the best fit for all the data using (13) leads to a set of nonlinear equations in the constants to be determined which are difficult to solve. Furthermore, it has been shown [12] that the unknown constants in (13) are extremely sensitive to the accuracy of the input data.

Another way of looking at these recurrence relations is in terms of weighting functions rather than as difference equations. Consider, for example, the problem of automatic fire control where it is necessary to point a gun not at where the target is, but where it is likely to be by the time the shell arrives. On the basis of the knowledge of the past path of the target, it is necessary to predict or extrapolate its future path. This problem has been handled in a very sophisticated way by Wiener [9] making use of autocorrelation functions and the modern techniques of operational calculus. Basically what is involved, however, is to obtain the approximate output of a given time series from a linear combination of the input and a number of its past values. The output, thus, is given approximately by a weighted sum of a number of past values of the input. The coefficients in the recurrence relation can then be considered as weighting coefficients.

To put it another way, in the linear theory of prediction an attempt is made to express $f(t + h)$ as a linear combination of values of $f(t - \tau)$ where $f(t)$ is the function one wishes to predict, h is an interval of time in the future, t is the present time, and $\tau \geq 0$. One way of doing this would be to select several values of τ, $\tau_i \geq 0$ and choose coefficients A_i, so that

$$\sum_0^n A_i f(t - \tau_i)$$

gives an optimum prediction. This is precisely what has been done here.

References

1 F. R. Larson and J. Miller, "A Time Temperature Relationship for Rupture and Creep Stress," TRANS. ASME, vol. 74, 1952, pp. 765–771.

2 S. S. Manson and A. M. Haferd, "A Linear Time Temperature Relation for Extrapolation of Creep and Stress Rupture Data," NACA TN 2890, March, 1953.

3 R. L. Orr, O. D. Sherby, and J. E. Dorn, "Correlations of Rupture Data for Metals at Elevated Temperatures," *Trans. ASM*, vol. 46, 1954, pp. 113–128.

4 S. S. Manson and A. Mendelson, "Optimization of Parametric Constants for Creep-Rupture Data by Means of Least Squares," NASA MEMO 3-1059E, 1959.

5 W. Betteridge, "The Extrapolation of the Stress-Rupture Properties of the Nimonic Alloys," *Journal, Institute of Metals*, vol. 86, no. 5, 1958.

6 R. M. Goldhoff, "Comparison of Parameter Methods for Extrapolating High Temperature Data," TRANS, ASME, Series D, JOURNAL OF BASIC ENGINEERING, vol. 81, 1959, p. 629.

7 S. S. Manson and W. F. Brown, Jr., "Time-Temperature Relations for the Correlation and Extrapolation of Stress-Rupture Data," *Proceedings, ASTM*, vol. 53, 1953, pp. 693–719.

8 Hans Conrad, "Correlation of High Temperature Creep and Rupture Data," TRANS. ASME, Series D, JOURNAL OF BASIC ENGINEERING, vol. 81, 1959, p. 617.

9 Norbert Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," The Technical Press and John Wiley & Sons, Inc., New York, N. Y., 1949.

10 P. N. Randall, "Constant Stress, Creep-Rupture Tests of a Milled Carbon Steel," *Proceedings, ASTM*, vol. 57, 1957, p. 854.

11 R. A. Buckingham, "Numerical Methods," Sir Isaac Pitman & Sons, Ltd., London, England, 1957.

12 Cornelius Lanczos, "Applied Analysis," Prentice-Hall, Inc., Englewood Cliffs, N. J., 1956.

DISCUSSION

Burt M. Rosenbaum[2]

The authors are to be commended for devising a method of extrapolation which is both simple and widely applicable. As pointed out by the authors in comparing equations (12) and (13), this ease of application stems from the fact that the resulting equations for the unknown constants A_1 and A_2 are linear in these constants whereas the Prony method of approximating the curve by a sum of exponentials gives rise to nonlinear equations. It is evident that the recurrence relation gives exactly the same extrapolated curve as would be obtained if the exponential solution, as determined by the more complicated Prony method, were used for extrapolation.

However, the Prony method does yield additional information in that it gives the entire curve over the range of the independent variable and not just points at intervals along the range. The Prony method actually determines the constants entering into the differential equation best fitting the data points. As an illustration, the curve given by equation (13) satisfies the second-order linear differential equation

$$\frac{d^2y}{dx^2} - (\alpha_1 + \alpha_2)\frac{dy}{dx} + \alpha_1\alpha_2 = 0$$

For a second-order recurrence relation, the coefficients of the corresponding differential equation may be found from the values of A_1 and A_2 by the equations

$$\left.\begin{aligned} \alpha_1 + \alpha_2 &= \frac{\ln(-A_2)}{(\Delta x)} \\ \alpha_1\alpha_2 &= \frac{1}{(\Delta x)^2}\left[\ln\left(\frac{A_1 + \sqrt{A_1^2 + 4A_2}}{2}\right) \cdot \ln\left(\frac{A_1 - \sqrt{A_1^2 + 4A_2}}{2}\right)\right] \end{aligned}\right\}$$

The determination of the constant-coefficient differential equation corresponding to the recurrence relation is desirable in some cases. Because the data are to be physically significant, one of the requirements of the differential equation in order that it can properly represent the data is that it possess a real solution throughout the range of the independent variable. This requirement implies that the coefficients of the differential equation must be real. As far as the recurrence relation is concerned, the A_i's always are real. However, the differential equation corresponding to these real A_i's may possess complex coefficients so that the solution of the differential equation is complex in general, turning real only at the stations originally chosen. A recurrence relation that corresponds to a complex-coefficient differential equation has only questionable physical significance and extrapolation by the use of such a recurrence relation must be viewed with suspicion.

It is possible to determine the recurrence coefficients for half-intervals from a knowledge of the coefficients for the original intervals. Let the order of the recurrence relation be two and let the relation be given by

$$y_i = A_1 y_{i-1} + A_2 y_{i-2} \tag{14}$$

Similarly, the half-interval recurrence relation is given by

$$\begin{aligned} y_i &= a_1 y_{i-1/2} + a_2 y_{i-1} \\ &= a_1(a_1 y_{i-1} + a_2 y_{i-3/2}) + a_2 y_{i-1} \end{aligned} \tag{15}$$

Solving for $y_{i-3/2}$ by means of

$$y_{i-1} = a_1 y_{i-3/2} + a_2 y_{i-2}$$

and substituting in equation (15), one obtains the desired result

$$y_i = (a_1^2 + 2a_2)y_{i-1} - a_2^2 y_{i-2} \tag{16}$$

A comparison of equations (14) and (16) yields

$$\left.\begin{aligned} A_1 &= a_1^2 + 2a_2 \\ A_2 &= -a_2^2 \end{aligned}\right\} \tag{17}$$

The above process can be continued indefinitely, each step giving the recurrence coefficients for double the number of stations. If either $A_2 > 0$ or $A_1 + 2\sqrt{-A_2} < 0$, the new recurrence coefficients are complex indicating that the interpolated values of the ordinate, that is, the values of the ordinate at the new stations, are not real. In order to insure that the recurrence coefficients stay real at each step, which is equivalent to requiring that the differential equation possess a real solution, the inequalities $A_2 < 0$ and $A_1 + 2\sqrt{-A_2} > 0$ must be satisfied. Thus, conditions on the recurrence coefficients may be found which are

[2] Nuclear Reactor Division, National Aeronautics and Space Administration, Lewis Research Center, Cleveland, Ohio.

equivalent to requiring that the coefficients of the corresponding differential equation be real without actually finding these latter coefficients.

The full-interval recurrence coefficients are given in terms of the half-interval coefficients by

$$\left.\begin{aligned} A_1 &= a_1^2 + 2a_2 \\ A_2 &= 2a_1a_3 - a_2^2 \\ A_3 &= a_3^2 \end{aligned}\right\}$$

for a three-term recurrence relation and

$$\left.\begin{aligned} A_1 &= a_1^2 + 2a_2 \\ A_2 &= 2a_1a_3 + 2a_4 - a_2^2 \\ A_3 &= a_3^2 - 2a_2a_4 \\ A_4 &= -a_4^2 \end{aligned}\right\}$$

for a four-term relation. Note that a necessary condition for an nth order recurrence formula to yield a real curve is apparently

$$(-1)^n A_n < 0.$$

As a final observation, data points are inaccurate to some degree and in addition the recurrence relation that is established does not fit the data points exactly. But in order to extrapolate to the next station by the use of a recurrence relation of order q, it is necessary to know the values of the ordinates at the last q stations of the curve which as mentioned are somewhat in error. Would the authors care to discuss more fully how the extrapolated curve may be obtained once the recurrence coefficients have been determined?

Authors' Closure

The authors wish to thank Dr. Rosenbaum for his illuminating comments. In particular, the equation relating the coefficients of the finite-difference recurrence relation to those of the corresponding differential equation and the conditions for the latter coefficients to be real are of great interest.

The significance of the fact that the coefficients of the differential equation may turn out to be complex is not very clear. In any recurrence relation, values of the desired function can only be obtained at discrete points. Nothing at all can be said about the function in the region between successive points. The reason for replacing the continuous solution by one which is valid at only discrete points is to simplify the solution. The fact that the corresponding differential equation would give nonreal solutions between stations does not seem to be relevant. Thus, for example, examination of Figs. 6, 8, 9, 10, and 12 indicates that, although one of the conditions given by Dr. Rosenbaum for the coefficients of the differential equation to be real, namely, $(-1)^n A_n < 0$, is violated in these cases, the extrapolations obtained were very good.

For a qth order recurrence relation, the extrapolated curves will in general be sensitive to the last q ordinates of the original data curves. This is due to the fact that once the A_i are determined, the last q values of the ordinates are used directly to begin the extrapolation. It is therefore important that these last q values be obtained in the best possible manner.

If there is a sufficient number of data points in this region, with little scatter, then one can have confidence in the originally faired curve. If, however, such is not the case, it might be wise to investigate further the validity of the extrapolation. Thus, one could extrapolate to the desired point using the last q values as described. One could then successively repeat the extrapolation each time beginning at one station earlier. In this way a set of extrapolated values can be obtained. If all these values agree with each other fairly well, one could have some confidence in the final result. If, however, widely differing values are obtained for the extrapolated point, care must be taken in deciding which value to use. One could, for example, average the different values obtained, using an appropriate weighting factor for each one.

In deciding on appropriate weighting constants, at least three important factors should be considered: (1) the number of stations between the extrapolated point and the starting values on which they are based, (2) how well the faired curve fits the data in the region where the extrapolation is started, and (3) the number of actual data points in this region. Thus it is obvious that the closer the extrapolated point is to the region where the extrapolation is started, the greater the number of data points in this region and the closer the faired curve fits the data in this region, the more weight should be attached to the resultant extrapolated value. Thus one could tentatively assume the weighting factor to be approximately inversely proportional to the number of intervals extrapolated, directly proportional to the number of data points in the region one is extrapolating from, and proportional to some function of the inverse of the deviation of these data from the faired curve. (A direct inverse proportion cannot be used for the deviation from the curve since this would lead to infinite weight for a data point right on the curve.) A detailed analysis of the type of weighting factors to be used could form an interesting future investigation.