Materials Science and Technology

ISSN: 0267-0836 (Print) 1743-2847 (Online) Journal homepage: http://www.tandfonline.com/loi/ymst20

Method for improving parametric creep rupture life of 2·25Cr–1Mo steel using artificial neural networks

M. Evans

To cite this article: M. Evans (1999) Method for improving parametric creep rupture life of 2·25Cr–1Mo steel using artificial neural networks, Materials Science and Technology, 15:6, 647-658, DOI: 10.1179/026708399101506391

To link to this article: https://doi.org/10.1179/026708399101506391

Published online: 19 Jul 2013.

Method for improving parametric creep rupture life of 2·25Cr–1Mo steel using artificial neural networks

M. Evans

A number of well known parametric models and a multilayer neural network are subjected to an 'acid test' of extrapolation to determine whether the latter can produce improved long term rupture life predictions for 2·25Cr–1Mo steel. Linear and non-linear least squares analysis was used to estimate the parametric models and genetic algorithms were used to identify and train the network. All the parametric models produced lifetime predictions for stresses below 60 MPa that were in error by some 20–40% on average. This reflected their tendency to overfit the data sets used for their estimation. Standard statistical measures of model inadequacy were of little use in overcoming this problem and the more non-linear models (e.g. Manson–Haferd) produced implausible extrapolations. In contrast, the optimised neural network was able to identify general patterns in the training data that were useful for extrapolation purposes and this was reflected in an average error of some 4–5%. MST/4184

The author is in the Department of Materials Engineering, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK. *Manuscript received 4 August 1998; accepted 27 November 1998.*

List of symbols

A_i, B_i, m	parameters to be estimated by least squares techniques
e	prediction error obtained from parametric or neural network
$error_k$	error in extrapolation
E_p	test for normality in log failure time
H_w	F statistic in test for heteroscedasticity
$\max(X_{i1\ldots n})$, $\min(X_{i1\ldots n})$	maximum, minimum of all n values for variable i
M	maximum number of input variables, i.e. input layer neurons
MAE	mean absolute error in predicted log failure times
MPAE	mean percentage absolute error in predicted log failure times
MSE	mean square error in predicted log failure times
n	number of values for given variable, e.g. failure time
p	parameter allowing for non-linear convergence in Manson–Brown technique
P	number of neurons in hidden layer
Q_c	activation energy
R	universal gas constant
R^2	coefficient of determination
s_{ik}	scaled value for input X_i
t_a	failure time at convergence point in Manson–Brown technique
t_j	threshold value used in neural network
t_F	time to failure
T	absolute temperature
T_a	coordinate for convergence point in Manson–Brown technique
U_{jk}	kth value for weighted sum produced by the jth hidden layer neuron
Var(e)	variance in prediction error obtained from parametric or neural network technique
w_{ij}	weight connecting input i to the jth hidden layer neuron
w_{jf}	weights connecting hidden layer neuron j to output layer neuron f
X_i	ith input variable to neural network or in parametric model
X_{ik}	kth value for input variable i
Y_{fk}	actual output used to compare with Y^p_{fk}
Y_{jk}	kth value for output produced by jth hidden layer neuron
Y^p_{fk}	kth value for output produced by fth output layer neuron
α	momentum parameter of back propagation algorithm
β_i, γ_i, δ_i, θ_i, λ_i, ϕ_i	parameters to be estimated by least squares techniques
$\dot{\varepsilon}_s$	secondary or minimum creep rate
η	learning parameter of back propagation algorithm
μ_p	percentile of failure time distribution
$\mu(\sigma)$	functional form relating log failure times to stress
ν	parameter in specification for heteroscedasticity
σ	stress

Introduction

A clear understanding of the long term creep and rupture behaviour is important for the safe design and operation of the structural components used in high temperature power plants. Fossil power plants are usually designed to have a life expectancy of about 100 000 h before creep rupture occurs. Commonly, rupture strength is then defined by the stress and temperature levels (subject to some arbitrarily selected safety factor) required to cause rupture

within the design life (100 000 h). Theoretically, this stress–temperature combination would then determine the safe operating levels for the power plant.

There are substantial difficulties in the application of this approach. The major problem is that few data up to 100 000 h are available even for well established materials. This problem is compounded by the fact that new materials are continually being developed for high temperature use and it is therefore essential to know their long term performance characteristics within a short period of time. There is therefore no alternative but to attempt to extrapolate rupture lives from data obtained in laboratories using much shorter tests (typically about 10^3–10^4 h). Such tests are usually carried out by increasing the temperatures (rather than the stresses) beyond those at which such materials would normally operate.

Various methods have been devised over the past 50 years to allow rupture lives to be extrapolated from such accelerated test data. Grant and Bucklin[1] and Glen[2] have used simple graphical techniques to relate rupture life to metallurgical changes, whereas the more empirical approaches used by Mendelsohn and Manson[3] and Brozzo[4] make use of numerical integration and recurrence relations. When information is available on the whole creep curve, as opposed to the rupture times only, θ projection techniques[5] have typically been used.

However, there is little doubt that the 'parametric' techniques have enjoyed more popularity than any of these other methods. A number of such techniques have been proposed (*see* the 'Parametric techniques' subsection below) but without exception they all impose some structure on the data that lacks a complete theoretical derivation. For example, Larson and Miller[6] suggested that a plot of log failure times versus the reciprocal of absolute temperature, for a given stress, should yield a linear line having a slope related to the activation energy. They further suggested that the slopes (and not the intercepts) of these linear lines (one for each stress level considered), were a function of stress – usually a log stress or a polynomial of stress. It has become evident over recent years that over a wide range of stresses, rupture time and stress are not log–linearly related[7] and the use of polynomial approximations can never give reliable long time extrapolations.

The resulting extrapolations are therefore based on a confidence in the universal applicability of the Arrhenius relation,[8] the assumption that the Monkman–Grant[9] relation is independent of stress and temperature (which has been questioned on a theoretical basis by Garofalo[10] and at an empirical level by Evans[11]), and an unknown functional relationship between stress and the activation energy. It is therefore hardly surprising that parametric techniques such as these have had little success in extrapolating rupture lives to 10 000 h (*see* Ref. 7).

Neural networks, by contrast, have general non-linear function mapping capability and are non-parametric data driven 'models' which impose few prior assumptions on the nature of the functional relationships which are responsible for generating the data. Many authors, such as Kang,[12] have also claimed that neural networks do not require large data sets to perform well at extrapolation. As such they are potentially very strong candidates for dealing with the types of uncertainty mentioned above that exist within relatively small creep databases.

Indeed there are increasing signs that neural networks surpass their parametric counterparts in interpolating microstructures, mechanical properties, and fatigue crack growth. Li *et al.*[13] used a neural network to predict the flow stress of a Ti–5Al–2Sn–2Zr–4Cr–4Mo (all compositions are in wt-%) alloy using input variables such as deformation temperature, equivalent strain rate, and equivalent strain. They showed such predictions to be a considerable improvement on those obtained using standard regression techniques. Badmos *et al.*[14] have recently used a Bayesian neural network to predict a variety of mechanical properties using variables such as alloying content, aging temperature, aging time, cold work, etc. as inputs. They used mechanically alloyed oxide dispersion strengthened ferritic steels and produced mechanical property interpolations that were metallurgically reasonable. Fujii *et al.*[15] have modelled fatigue crack growth rates in nickel base superalloys using neural networks.

The aim of the present paper is to ascertain whether neural networks are equally well suited to the prediction of creep rupture lives. In particular, a number of well known parametric techniques are subjected to an 'acid test' of extrapolation to failure times at stresses below 60 MPa and compared with a multilayer perceptron (MLP) neural network to ascertain whether the latter yields any improvement in performance. To meet this objective this paper is structured as follows. The section below outlines the Japanese National Research Institute for Metals (NRIM) creep database used in the present paper together with the acid test of extrapolation to which each parametric and neural network model will be subjected. This is followed by a section reviewing the structure of an MLP neural network together with a number of popular parametric techniques used to extrapolate rupture lives. Then there is a section describing the procedures used for identifying and estimating the models (either parametric or an MLP network) to be used for the extrapolation of failure times. In the penultimate section the extrapolation results are summarised and their implications discussed. Finally, conclusions will be drawn.

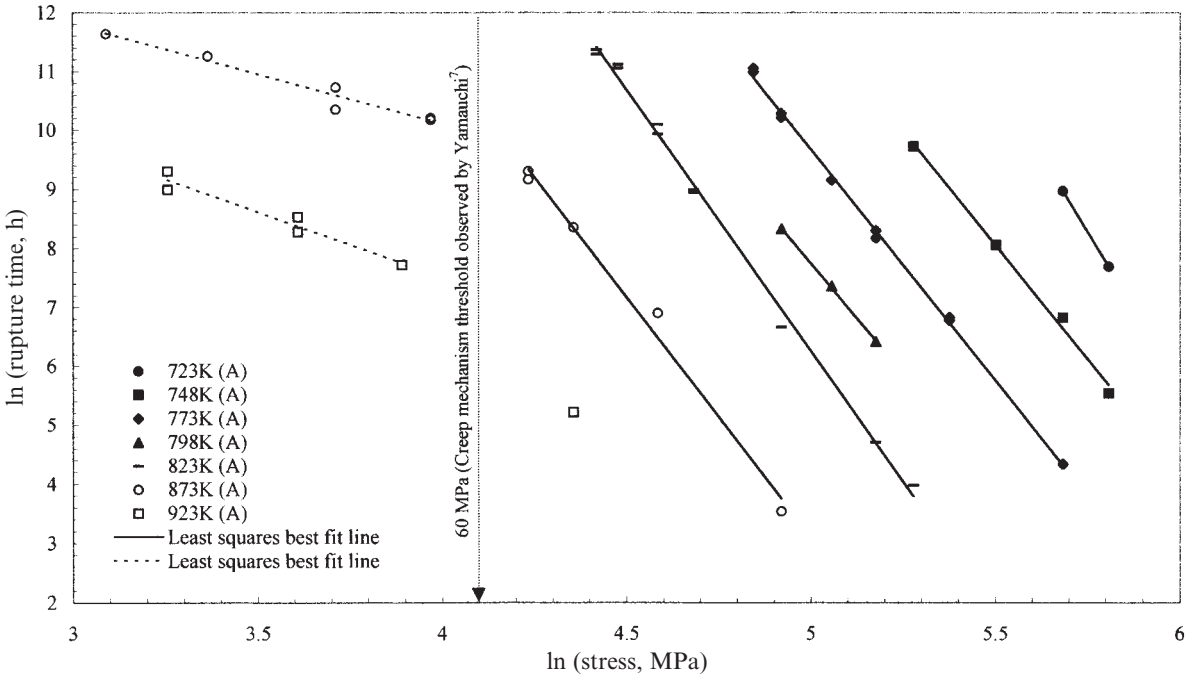
Creep database

The NRIM[16] creep data sheet no. 3B contains a large quantity of information on the elevated temperature properties of 2·25Cr–1Mo steel used for boiler and heat exchanger seamless tubes. The boiler tubes used in the test programme were sampled at random from commercial stocks that were produced in Japan using an electric arc furnace. Each tube had an outside diameter of 50·8 mm, a wall thickness of 8 mm, and a length of 5000 mm.

Creep tests were conducted on solid cylindrical specimens with a gauge length of 30 mm at constant load with a loading accuracy of $\pm 0{\cdot}5\%$. Temperatures between 373 and 873 K were controlled to ± 3 K and temperatures in excess of 873 K to ± 4 K. A number of different heat treated specimens were tested. This paper concentrates on the MAF heat batch of test specimens. For this batch, materials were obtained from rotary pierced and cold drawn ingots and heat treated for 1200 s at 1200 K followed by 7800 s at 993 K and air quenching. The material had a grain size of 6 (ASTM) and a Rockwell hardness of 78 HRB. The chemical composition of this heat batch was Fe–0·1C–0·23Si–0·43Mn–0·011P–0·009S–2·46Cr–0·94Mo–0·043Ni–0·005Al–0·07Cu–0·008N.

The test temperatures varied from 723 to 923 K and the stress varied from 14 to 333 MPa. Over these ranges some 60 specimens were tested and information on time to failure, elongation, reduction of area, and the minimum creep rate were collected. However, by the 1st April 1986, 12 of these specimens had not yet failed so that only a survival or runout time, rather than a failure time, could be recorded. The remaining 48 failure times are plotted against stress in Fig. 1 (using a log scale).

The plots for the different temperatures appear linear and almost parallel to each other as shown by the linear least squares best fit lines that were estimated separately for each temperature regime. However, the slopes of the lines are clearly different between the higher stress range



1 Stress *v.* time to rupture plots for 2·25Cr–1Mo steel (heat MAF): A actual data

above 60 MPa and the lower stress range below 60 MPa. This apparent discontinuity was observed by Yamauchi *et al.*,[7] who attributed it to differing creep mechanisms over these two stress ranges.

Any parametric technique capable of accurately predicting rupture times at stresses below 60 MPa using only the data obtained at stresses above 60 MPa can therefore be considered to have passed a rather stringent test of extrapolation. This applies especially to failure times obtained at temperatures of 923 K. For this temperature, there is only one failure time at a stress above 60 MPa from which to extrapolate to lower stresses (at this same temperature). Any technique that extrapolates well to lower stresses at this temperature has successfully identified a more general relationship existing between rupture time and all stresses and temperatures. This will be the acid test of extrapolation.

Models for extrapolation of failure times to lower stress regimes

PARAMETRIC TECHNIQUES

Parametric methods are also referred to as temperature compensated time techniques in that they attempt to remove the influence of temperature on the time to rupture. For most techniques this is achieved through the use of the Arrhenius[8] and Monkman–Grant[9] equations. Most mathematical analysis of creep rupture behaviour is based upon the Arrhenius equation that relates chemical reactions and temperature. As a consequence modelling creep rate as a process takes the form

$$\dot{\varepsilon}_s = A_0 \exp\left(-\frac{Q_c}{RT}\right) \quad \ldots \quad (1a)$$

Monkman and Grant have suggested that time to failure increases as the minimum creep rate decreases as follows

$$t_F = \frac{A_1}{\dot{\varepsilon}_s^m} \quad \ldots \quad (1b)$$

where the constants, m and A_1 purportedly differ only for different materials (m typically by a factor in the range 0·1–10%). However, in a recent paper Evans[11] has shown this assertion to be false.

The time–temperature parameter used by Dorn and Shepherd[17] and Larson and Miller[6] can be derived from equations (1a and b). As this parameter is purely a function of stress, a two-dimensional plot correlating stress with rupture life can be made. To illustrate, let $\mu(\sigma)$ be a function describing the dependency of time to failure on stress. Dorn and Shepherd assumed that the parameter A_0 in equation (1a) can be made a function of stress so that

$$A_0 = \mu(\sigma) \quad \ldots \quad (2a)$$

Substituting equations (1a) and (2a) into equation (1b) and taking natural logs gives the following model for failure time

$$\ln(t_F) = \ln(A_1) - m\ln[\mu(\sigma)] + \frac{mQ_c}{R}\left(\frac{1}{T}\right) \quad \ldots \quad (2b)$$

A plot of $\ln(t_F)$ against $1/T$ for a given stress should therefore give a line with slope $(mQ_c)/R$ and an intercept that varies in some manner with the stress level. Such a representation implies that the effects of stress and temperature on times to failure are separable, i.e. no interaction between the two is present.

Larson and Miller,[6] instead of combining stress and temperature multiplicatively, assumed the activation energy to be a function of stress. This allows stress and temperature to interact in the determination of failure time. This is supported by the observation that there is a tendency for lines of constant stress plotted in $\ln(t_F)$–$1/T$ space to diverge. That is

$$Q_c = \mu(\sigma) \quad \ldots \quad (3a)$$

Substituting equations (1a) and (3a) into equation (1b) and taking natural logs gives the following model for rupture times

$$\ln(t_F) = \ln(A_1) - m\ln(A_0) + \frac{m}{R}\left[\frac{\mu(\sigma)}{T}\right] \quad \ldots \quad (3b)$$

A plot of $\ln(t_F)$ versus $1/T$ for a given stress should therefore give a line with slope $(mQ_c)/R$ and intercept $\ln(A_1) - m\ln(A_0)$. This time however the slopes (not intercepts) will differ for different stress levels. It may be

that both A_0 and Q_c in equation (1a) are functions of stress, leading to

$$\ln(t_F) = \ln(A_1) - m\ln[\mu(\sigma)] + \frac{m}{R}\left[\frac{\mu(\sigma)}{T}\right] \quad . \quad . \quad . \quad . \quad (4)$$

It is usual for equations (2b), (3b), and (4) to be made operational by using a polynomial for the stress functions $\mu(\sigma)$ and $\ln[\mu(\sigma)]$. That is

$$\mu(\sigma) = \beta_0 + \beta_i \sum_{i=1}^{k} (\sigma)^i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5a)$$

and

$$\ln[\mu(\sigma)] = \delta_0 + \delta_i \sum_{i=1}^{k} [\ln(\sigma)]^i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5b)$$

where the constants β_0, β_i, δ_0, and δ_i are typically estimated using least squares techniques.

A rather curious parametric method is that suggested by Manson and Haferd[18] which is based upon the approximate linearity which they (and Bailey[19]) observed in plots of $\ln(t_F)$ versus T, for constant stress, and the extrapolation to a common point of intersection. The fundamental equations used by Manson and Haferd appear therefore to be equation (1b), together with

$$\dot{\varepsilon}_s = A_0 \exp\left(-\frac{Q_c}{R}T\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6a)$$

If it is further assumed that $Q_c = \mu(\sigma)$ and that the slopes of the lines of constant stress in T–$\ln(t_F)$ space converge to a specific failure time t_a at a temperature T_a, then the following familiar Manson–Haferd equation for rupture time is obtained

$$\ln(t_F) = \ln(A_1) - m\ln(A_0) + \frac{m}{R}[\mu(\sigma)(T - T_a)] \quad (6b)$$

A generalisation of this technique has been provided by Manson and Brown, Jr[20] who allow for non-linear convergence on point T_a at time t_a by introducing a further parameter p into equation (6b) as follows

$$\ln(t_F) = \ln(A_1) - m\ln(A_0) + \frac{m}{R}[\mu(\sigma)(T - T_a)^p] \quad (6c)$$

Much of this independent work was collated by Manson[21] who derived a rupture time equation encompassing the Larson–Miller, Manson–Haferd, and Manson–Brown parametric models. However, this is by no means the only method for generalising such techniques. For example, Eyring *et al.*[22] have suggested that the Arrhenius equation lacks a theoretical derivation when factors other than temperature are important in determining rate processes. The Eyring model based on chemical reaction rate theory and quantum mechanics suggests that

$$\dot{\varepsilon}_s = A_0 T^{A_2} \exp\left(-\frac{Q_c}{RT}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6d)$$

where A_2 requires estimation. Invoking the assumption that Q_c is a function of stress and that $A_2 = 0$ results in the Larson–Miller parametric technique.

More recent (but still some 15–25 years old) parametric techniques include the minimum commitment method[23] and the 'Soviet' models.[24] These are essentially just combinations of the above parametric techniques. For example, the minimum commitment method uses log stress, stress, stress squared, the reciprocal of temperature, and the ratio of stress to temperature to predict log rupture time. One variation of this technique gives

$$\ln(t_F) = \lambda_0 + \lambda_1 \ln(\sigma) + \lambda_2\sigma + \lambda_3\sigma^2 + \lambda_4 T + \lambda_5 \frac{1}{T} \quad (7a)$$

The first variant of the Soviet model (Soviet model 1) is similar to this and takes the form

$$\ln(t_F) = \lambda_0 + \lambda_1 \ln(T) + \lambda_2 \ln(\sigma) + \lambda_3 \frac{1}{T} + \lambda_4 \frac{\sigma}{T} \quad (7b)$$

The second variant (Soviet model 2) takes the form

$$\ln(t_F) = \lambda_0 + \lambda_1 \ln(T) + \lambda_2 \frac{\ln(\sigma)}{T} + \lambda_3 \frac{\sigma}{T} + \lambda_4 \frac{1}{T} \quad (7c)$$

All the above techniques relate log rupture time to transformations of stress and temperature. As such all those specimens that have not failed by the end of the testing programme must be ignored. However, such censored data are of considerable value in that they contain information that would be of use in predicting failure times. Only recently have a number of authors addressed this important issue by relating some percentile μ_p of the log failure time distribution, rather than the failure time itself, to transformations of stress and temperature. Now runout times can be included in the analysis. For example, in this type of analysis the Soviet model 2 above would take the form

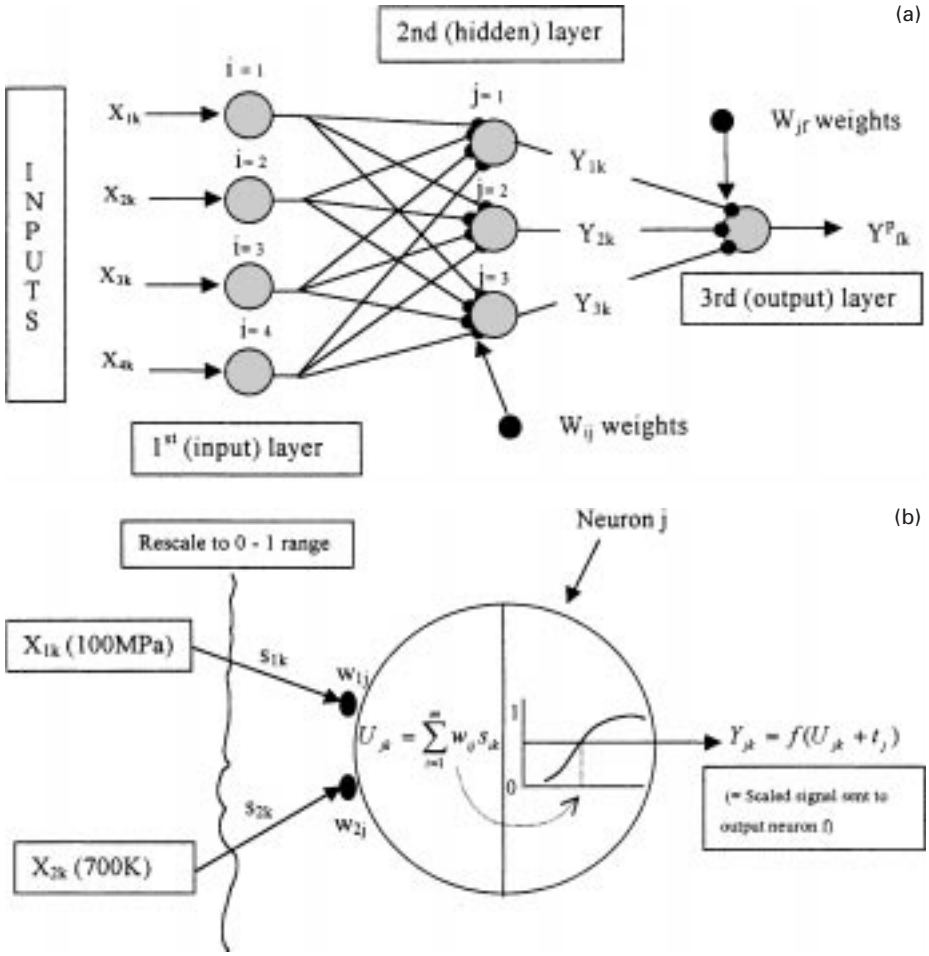
$$\mu_p = \lambda_0 + \lambda_1 \ln(T) + \lambda_2 \frac{\ln(\sigma)}{T} + \lambda_3 \frac{\sigma}{T} + \lambda_4 \frac{1}{T} \quad . \quad . \quad . \quad (7d)$$

To become operational some log failure time distribution must be specified. Authors such as Holdsworth *et al.*[25] have restricted their coverage to the lognormal, Weibull, and log logistic distributions, whereas the present author[11] has taken a more general approach by using a distribution that encompasses all of these distributions, among others.

NEURAL NETWORKS

Neural networks behave similarly to the current conception of the neurons in the human brain and so have proved highly effective in the area of pattern recognition. As such neural networks have been applied in a wide variety of fields and to many different problems (*see* 'Introduction' section for some examples in the field of engineering). The term 'neural network' is actually used to describe a variety of different network architectures. The multilayer perceptron (MLP) is one of the earliest neural networks developed and is today the most frequently used network architecture. The present paper will work with such a network (*see* Ref. 26 for an informative summary of MLPs). The reader should however be aware of the considerable developments in neural network modelling that have occurred in recent years. These include recurrent networks (*see* Ref. 27 for more details) that are mainly used to process time varying data and radial basis function networks (*see* Ref. 28 for more information).

The basic architecture of the MLP neural network is shown in Fig. 2*a*. The first layer, termed the input layer, consists of a number of input nodes termed neurons. In general there will be one such neuron for each variable used to model the output. Let M represent the maximum number of input variables so that there will be $i = 1, \ldots, M$ neurons in the input layer. Each input variable X_i then has $k = 1, \ldots, n$ values associated with it. The network then contains one or more hidden layers and each neuron in the hidden layer is connected to every neuron in the previous layer via a set of weights. Let j represent neurons in a single hidden layer with $j = 1, \ldots, P$ such hidden layer neurons. Then there will be $M \times P$ such weights w_{ij} connecting the input neurons to the hidden layer neurons. Of course MLPs can have more than one hidden layer. Following the hidden layer is the output layer which will contain a neuron for every variable that is to be modelled. Again, weights are used to connect each output neuron to every neuron in the previous hidden layer. With one output neuron and one hidden layer there will be P such weights



a basic architecture of MLP network; *b* feed forward procedure through part of network

2 Multilayer perceptron (MLP) neural network

w_{jf} connecting the hidden layer neurons to the output neuron.

The MLP is a statistical model that is built up by tuning the weights described above. These weights are used to map a series of input signals into the required output. Values for each weight are obtained via a process termed training. Such training is carried out by passing a set of inputs (a stress–temperature combination) through the MLP and adjusting the weights to minimise the error between the answer the network gives for the output Y^{p}_{fk} and the actual output value Y_{fk}. This is repeated for all input combinations and the weights are those that minimise the average of all n squared errors. Once the weights have been set the network is able to produce predictions for the output using input values not used during the training process.

Figure 2*b* shows the training process in more detail by illustrating the connection between two input neurons ($i = 1$ and $i = 2$) and a single neuron (neuron j) in the first hidden layer. A pair of inputs X_{1k} and X_{2k}, (where X_{1k} might be a stress of 100 MPa and X_{2k} a temperature of 700 K, for example), are presented to the network and a weighted sum of them formed using the relevant weights (w_{1j} and w_{2j}). This weighted sum U_{jk} is then biased using a threshold value t_j and passed through a non-linear activation function $f(U_{jk} + t_j)$ to produce an output Y_{jk} in the range 0 to 1. This threshold value is analogous to the constant term in a linear regression analysis. It is because of this activation function that the MLP is ideal for modelling non-linear relationships such as that between stress and time to failure. This output is then sent to a neuron in the next hidden layer or to an output neuron where a similar weighting and normalisation procedure takes place. The result will be an output prediction in the range 0 to 1. To allow comparison to be made, the output data (failure times) must be rescaled to this range before analysis. It is also advisable to rescale all inputs into a similar range so that the weighted sums described above are often calculated on the scaled inputs.

Clearly there are a number of unknowns to be specified before a MLP can be used for prediction. These include the number of hidden layers, the number of neurons in each hidden layer, the number of input variables to use, and the method for optimising the weights mentioned above. These will be discussed in the following section and are analogous to estimation considerations when using the least squares technique on parametric models.

Estimation and network training

PARAMETRIC TECHNIQUES

As considered above, none of the failure times obtained below 60 MPa should be used in estimating the parametric models that will be used to predict these failure times. For this reason, and to aid extrapolation, the NRIM data shown in Fig. 1 are divided as follows. Two interpolative data sets and one extrapolative data set are used. The first interpolative data set consists of all specimens tested at a stress above 60 MPa that actually failed. There are 37 such specimens comprising this set, termed training data set 1. The second interpolative data set consists of all specimens (including those that remained unfailed by the end of the

test programme) tested at a stress above 60 MPa – of which there are 47 such specimens. These comprise training data set 2. The extrapolative data set comprises all those specimens in Fig. 1 tested at a stress level below 60 MPa that actually failed – of which there are 11. This will be termed the test data set.

Each parametric model described above, except the more recent ones proposed by Holdsworth *et al.*[25] and by Evans,[11] will therefore be estimated by linear or non-linear least squares using only the 37 failure times above 60 MPa that constitute training data set 1. The models described by Evans and by Holdsworth *et al.* will be estimated using maximum likelihoood techniques and because they use unfailed specimens, training data set 2 will be used for this estimation.

For each technique a simplified model will be used for extrapolative purposes. That is, all terms with an insignificant Student t statistic will be omitted from the model provided that the residuals remain normally distributed with a variance that is independent of stress and temperature, i.e. homoscedastic. The simplified model is then re-estimated using the interpolative data set and used to extrapolate to lower stresses.

Normality in the residuals is tested using the omnibus test suggested by Doornik and Hansen.[29] Suppose e is used to represent the estimated residuals from any of the parametric techniques. For example, in the Larson–Miller[6] equation (3b) these residuals are computed from the formula

$$e = \ln(t_F) - \left\{ [\ln(A_1) - m\ln(A_0)] + \frac{m}{R}\left[\frac{\mu(\sigma)}{T}\right] \right\} \qquad (8)$$

where the parameters A_1, etc. are estimated by linear least squares. The omnibus test for normality has as its basis the observation that if e has a standard normal distribution (and therefore $\ln(t_F)$ is normally distributed) then the degree of skewness (asymmetry around the mean of zero) should be zero and the degree of kurtosis (peakeness around the mean) should be 3. The resulting test statistic E_p has a chi square distribution with two degrees of freedom under the null hypothesis that the residuals are normally distributed. Like many tests for normality, their effectiveness depends upon using large samples.

A test attributed to White[30] involves the following specification for heteroscedasticity

$$\mathrm{Var}(e) = \phi_0 + \phi_i \sum_{i=1}^{M} X_i + \gamma_i \sum_{i=1}^{M} X_i^2 + \theta_i \sum_{i \neq j} X_i X_j + v \qquad (9a)$$

where v is assumed normally distributed with mean zero and constant variance. This suggests a particular type of heteroscedasticity where the variability in e and therefore $\ln(t_F)$ depends on all or some of the predictor variables, their squares, and their cross products. Note that for constant variance, only ϕ_0 should be statistically significant. Hence a significant student t statistic for ϕ_i, γ_i, or θ_i is an indication of heteroscedasticity. Alternatively, a standard F statistic can be constructed to test the joint hypothesis that all ϕ_i, γ_i, and θ_i (except ϕ_0) are equal to zero. Let H_w be this F statistic. Thus a value for H_w in excess of its critical value suggests that heteroscedasticity is present. White[30] has shown that for practical purposes Var(e) can be replaced with e^2 without invalidating the test. Hence the test requires a least squares regression of the following

$$e^2 = \phi_0 + \phi_i \sum_{i=1}^{M} X_i + \gamma_i \sum_{i=1}^{M} X_i^2 + \theta_i \sum_{i \neq j} X_i X_j + v \qquad (9b)$$

NETWORKS

The MLP neural network requires training in a learning process before it can be used to interpolate and extrapolate rupture times. The training procedure used in this present work had eight tasks as described below.

Task 1

This determines the architecture of the neural network, i.e. the number of hidden layers, the number of neurons in each hidden layer, and the number of input variables to feed into the network. Values for these three unknowns were obtained using a simple genetic algorithm. Davis[31] provides an informative overview of the workings of genetic algorithms and in the context of neural networks the algorithm functions as follows. The genetic algorithm is basically a stochastic numerical optimisation procedure. As such some 35 different architectures were randomly selected within the limits of no more than three hidden layers, no more than 35 neurons in the first hidden layer, 15 in the second, and five in the third, and no more than two input variables, namely, stress and temperature. Tasks 2–6 below are then carried out on each architecture.

Task 2

For each of the 35 architectures the weights and thresholds (w_{ij}, w_{jf}, and t_j) are initially assigned a random value between -1 and $+1$.

Task 3

Each architecture is then provided with the same input and output vectors. The first input vector is composed of all stresses above 60 MPa at which a rupture test was carried out. The second input vector is composed of all the temperatures corresponding to the stresses above 60 MPa at which a rupture test was carried out and the output vector is the log of rupture time associated with the test conditions shown in the input vectors. That is, training data set 1 is used. Each input vector is scaled so that each term is within the range 0 to 1 for the transformed value s_{ik}

$$s_{ik} = \frac{X_{ik} - \min(X_{i1\ldots n})}{\max(X_{i1\ldots n}) - \min(X_{i1\ldots n})} \qquad (10)$$

The output variable, log failure time, is tranformed in the same manner.

For the present study Fig. 1 shows that there are $n = 48$ failure times, where 11 of these are held back from the network during training so that extrapolations made by the network can be assessed. Thus $n - 11 = 37$ scaled stress–temperature combinations (some of which are duplicates) are presented to the network.

Task 4

These scaled values are presented to each of the 35 architectures and the outputs produced by the neurons of the successive layer in each architecture are described by Y_{jk}, which then forms the input to the next hidden layer or the output layers

$$Y_{jk} = \frac{1}{1 + \exp(-U_{jk} + t_j)} \qquad (11a)$$

where

$$U_{jk} = \sum_{i=1}^{M} w_{ij} s_{ik} \qquad (11b)$$

The above procedure of weighting and normalisation then continues. It should be noted that a simple logistic function has been chosen for the normalisation procedure but others are of course feasible.

Task 5

The weights and thresholds of all 35 architectures are now updated using the error back propagation algorithm. This is a standard deterministic numerical optimisation procedure, but stochastic procedures such as simulated annealing could also be used (*see* Masters[32]). For example, in the single hidden layer architecture shown in Fig. 2*a*

above, the hidden to output layer weights w_{jf} are adjusted using the formula

$$\Delta w_{jf} = -\eta Y_{jk}(Y_{fk} - Y^{p}_{fk})Y_{fk}(1 - Y_{fk}) \quad . \quad . \quad . \quad . \quad . \quad (12a)$$

The input to hidden layer weights w_{ij} are then adjusted using the formula

$$\Delta w_{ij} = -\eta Y_{jk}(Y_{fk} - Y^{p}_{fk})Y_{fk}(1 - Y_{fk})w_{jf}Y_{jk}(1 - Y_{jk})s_{ik} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12b)$$

In equations (12a and b) η is termed the learning parameter and its value aids in determining smooth convergence. Such smooth convergence is also accelerated or retarded by using a momentum parameter α which is a proportion of the difference between the current w_{if} (and w_{ij}) weight and its value on the previous iteration. The parameter α can then be added to the weight adjustments given by equations (12a and b).

Each of the 35 architectures is assigned a different learning and momentum parameter. For each architecture the first stress–temperature pairing (X_{11} and X_{21}) is presented to the 35 architectures and Y^{p}_{f1} computed for each. With this, the weights are modified using equations (12a and b). Then the second input pairing is presented to obtain Y^{p}_{f2} and the weights again updated using equations (12a and b). This is repeated until all 37 scaled stress–temperature combinations have been presented to all 35 architectures.

Task 6

Once all $k = 1, \ldots, (48 - 11)$ observations on the two input variables have been presented to the 35 architectures and the weights adjusted, an epoch is defined as having occurred. The resulting output predictions from each architecture Y^{p}_{fk} obtained at the end of the first epoch can be compared with the actual outputs Y_{fk} (log failure time) and the mean square error (MSE) computed using the equation

$$\mathrm{MSE} = \frac{\sum_{k=1}^{n=37}(Y_{fk} - Y^{p}_{fk})^2}{\text{no. of data points in training set 1}} = \frac{\sum_{k=1}^{n=37}(Y_{fk} - Y^{p}_{fk})^2}{48 - 11} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Thus each of the 35 architectures will have a MSE associated with it. A total of 500 epochs are carried out by repeating tasks 4–6 this number of times. By this time the weights of each network architecture should have converged to give a minimum MSE for each type of architecture.

Task 7

Once task 6 has been completed the worst performing network architectures (i.e. those with the highest MSEs after 500 epochs) are dropped from the population of 35 and replaced with randomly generated new architectures, with the remaining architectures being crossbred (using a double point crossover technique) and mutated on a random basis (this is set to occur 10% of the time) to generate 35 new architectures. These 35 new architectures will now have different learning and momentum parameters and different numbers of hidden layers and numbers of neurons as a result of this crossbreeding.

Task 8

Tasks 4–7 are now repeated until 25 generations of 35 architectures have been evolved, by which time the average performance of the 35 architectures will have converged on the best architecture in the population which represents the optimal network structure, i.e. the architecture that minimises the average squared interpolation error.

Once the optimal network architecture has been selected and trained using only those results obtained above 60 MPa, the weights associated with it are used to predict the log rupture times obtained at stresses below 60 MPa. Thus no information at stresses below 60 MPa is used by the network to obtain such predictions.

EXTRAPOLATION

Once the various parametric models have been estimated using training data sets 1 and 2 and simplified and an optimal MLP network trained using training data set 1, they are used to predict failure times at stresses below 60 MPa. Such failure times form the extrapolative data set as defined above. The error in extrapolation is defined as

$$error_k = Y^{p}_{fk} - Y_{fk} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14a)$$

where Y_{fk} is one of the 11 log failure times constituting the extrapolative data set and Y^{p}_{fk} is the prediction made for that log failure time using one of the parametric techniques or the MPL network. The performance of each model can now be assessed using these errors through three statistics. The mean square error over the failure times constituting the extrapolative data set is defined as

$$\mathrm{MSE} = \frac{\sum_{k=1}^{11}(Y^{p}_{fk} - Y_{fk})^2}{\text{no. of test data points}} = \frac{\sum_{k=1}^{11}(Y^{p}_{fk} - Y_{fk})^2}{11} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14b)$$

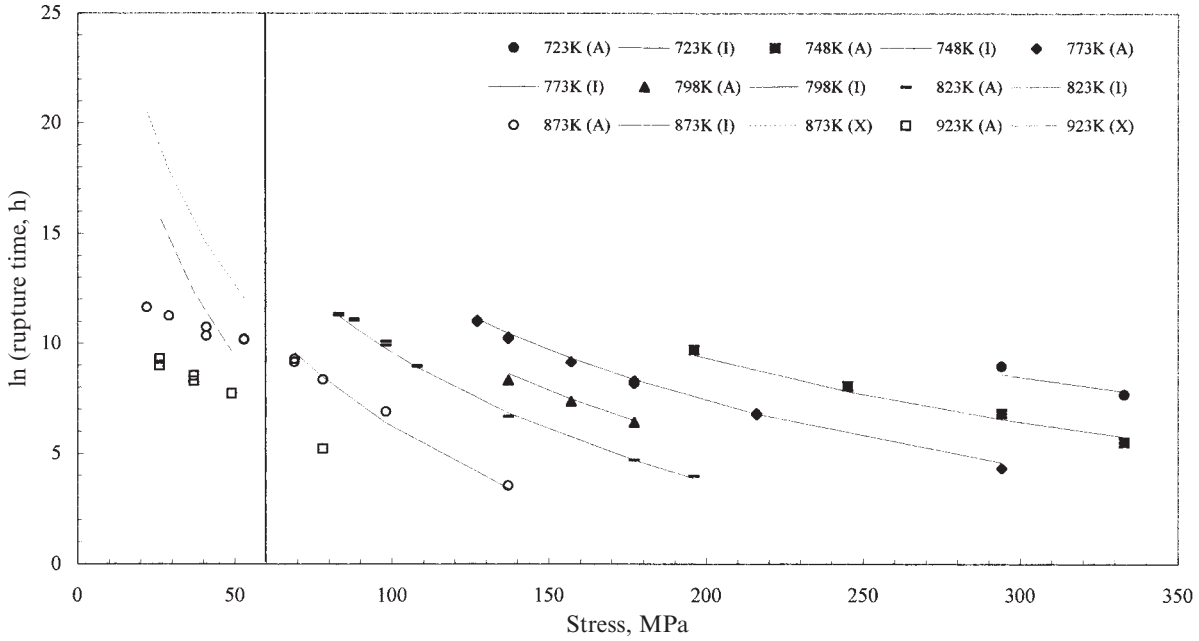
The mean absolute error over the failure times constituting the extrapolative data set is defined as

$$\mathrm{MAE} = \frac{\sum_{k=1}^{11}|Y^{p}_{fk} - Y_{fk}|}{\text{no. of test data points}} = \frac{\sum_{k=1}^{11}|Y^{p}_{fk} - Y_{fk}|}{11} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14c)$$

Table 1 Least squares parameter estimates* and diagnostic values for Soviet 1 and Larson–Miller models for log rupture life

	Variable							Diagnostic†		
Model	Constant	$1/T$	$\ln(T)$	$\ln(\sigma)$	σ/T	σ^2/T	σ^3/T	R^2, %	E_p	H_w
Soviet 1, full	342·79 (1·66)	96·684 (0·44)	−44·742 (−1·68)	−9·897 (−24·11)	8·082 (4·14)	…	…	98·97	2·01 [5·99]	1·01 [2·31]
Soviet 1, simplified	433·63 (51·62)	…	−56·471 (−48·22)	−9·988 (−28·55)	8·638 (5·89)	…	…	98·96	1·80 [5·99]	1·01 [2·32]
Larson–Miller, $i = 1$	−36·483 (−7·04)	407·27 (8·98)	…	…	−30·614 (−10·08)	…	…	75·23	3·35 [5·99]	1·69 [2·53]
Larson–Miller, $i = 2$	−47·699 (−24·18)	552·34 (29·70)	…	…	−94·071 (−22·15)	0·1486 (15·44)	…	96·99	0·69 [5·99]	2·85 [2·36]
Larson–Miller, $i = 3$	−50·253 (−49·35)	614·09 (55·06)	…	…	−171·72 (−21·35)	0·5826 (13·35)	−0·0007 (−10·01)	99·17	12·2 [5·99]	0·94 [2·31]

* Values in parentheses are Student t values.
† Values in brackets are critical values at 5% significance level.



3 Interpolative and extrapolative capabilities of Soviet model 1: A actual data, I interpolative prediction, X extrapolated prediction

The mean percentage absolute error over the failure times constituting the extrapolative data set is defined as

$$\text{MPAE} = \frac{\left(\sum_{k=1}^{11} |Y_{fk}^{p} - Y_{fk}|/Y_{fk}\right) \times 100}{\text{no. of test data points}} = \frac{\left(\sum_{k=1}^{11} |Y_{fk}^{p} - Y_{fk}|/Y_{fk}\right) \times 100}{11} \quad \ldots \quad (14d)$$

Results

Table 1 gives the least squares estimates for the Soviet model 1 and the Larson–Miller model using training data set 1. The full Soviet model 1 estimates all the parameters shown in equation (7b), whereas the simplified Soviet model 1 has removed from it all the variables shown to be statistically insignificant by the Student t statistics (shown in parentheses below the parameter estimates). For the Larson–Miller model three variants corresponding to $i = 1$, $i = 2$, and $i = 3$ are shown and these correspond to the term i in equation (5a). Thus when $i = 1$, the Larson–Miller model contains the term σ/T, but when $i = 2$ it also contains the quadratic term σ^2/T.

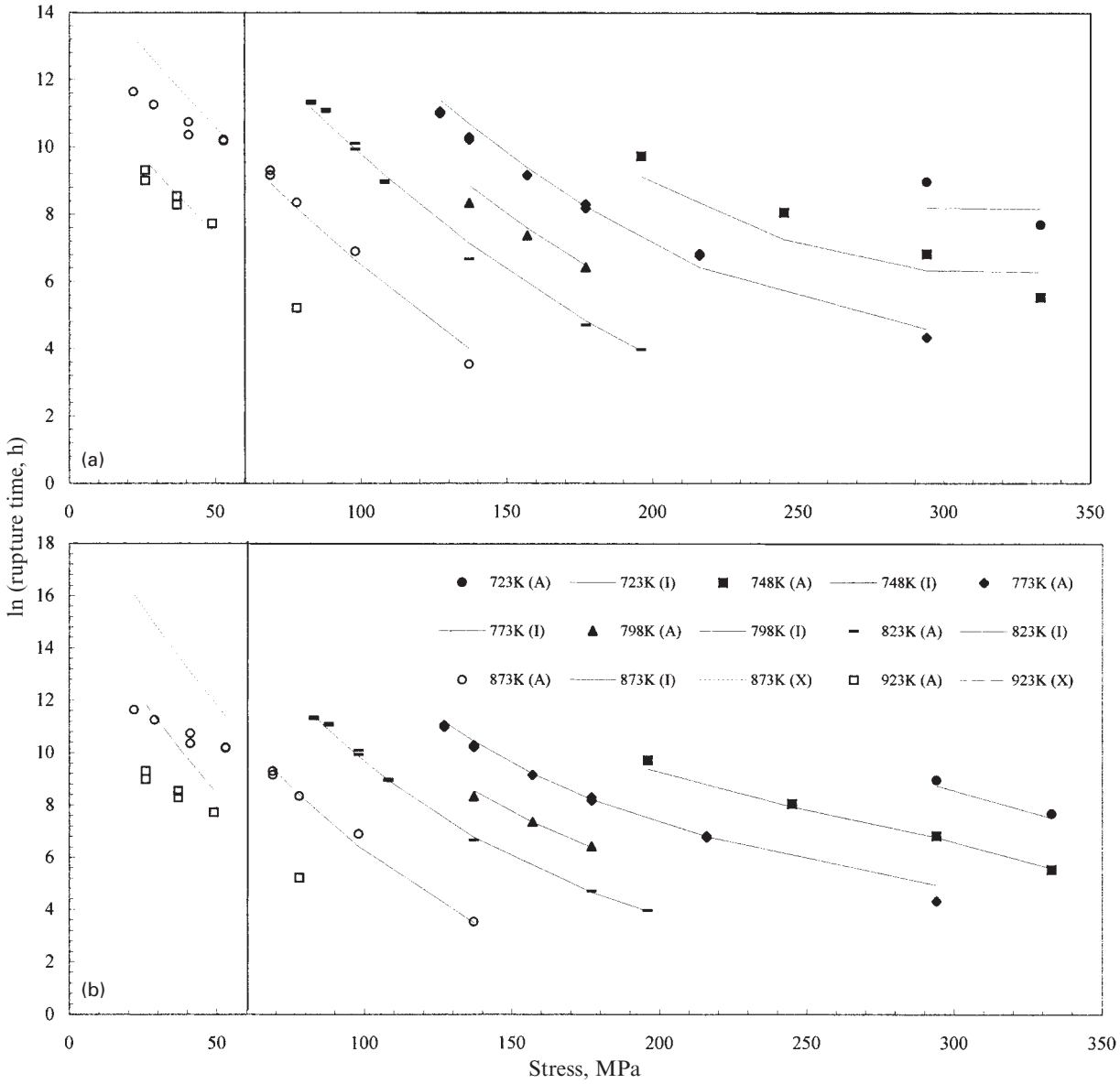
Presented alongside the parameter estimates in Table 1 are three diagnostic tests for model adequacy. The coefficient of determination R^2 shows the percentage variation in log failure time explained by all the variables included in the model. Under the null hypothesis of normality the Doornik and Hansen[29] test for normality in log failure time E_p follows a chi distribution with two degrees of freedom. Critical values at the 5% significance level are given in brackets. Under the null of homoscedasticity, the White[30] test for heteroscedasticity H_w (equation (9a)) using the squares and cross products of independent variables follows an F distribution. Critical values at the 5% significance level are given in brackets.

Table 2 gives the errors obtained in extrapolating log failure times to the stresses and temperatures given at the

Table 2 Errors in extrapolation using various parametric and multilayer perceptron (MLP) network techniques

		Error* in extrapolation of rupture time, ln(h), at T/σ of										
		873 K				923 K						
Model	Eqn	53 MPa	49 MPa	41 MPa	37 MPa	37 MPa	29 MPa	26 MPa	22 MPa	MSE†	MAE†	MPAE†, %
Soviet 1	7b	−1·8964	−1·9273	−4·1669	−4·0643	−3·8085	−6·6029	−6·772	−8·9108	25·8256	4·5691	46·4631
(simplified)		−1·8672		−3·7851				−6·4584				
Soviet 2	7c	−1·8151	−1·6106	−4·1247	−3·6456	−3·3898	−6·6502	−6·2588	−9·0518	24·1418	4·3655	44·0943
(simplified)		−1·7858		−3·743				−5·9452				
Min. commitment	7a	−1·8544	−2·1061	−4·0886	−4·1964	−3·9406	−6·4737	−6·8457	−8·7395	25·6928	4·5736	46·6579
(simplified)		−1·8252		−3·7069				−6·5321				
Dorn	2b	−1·1661	−1·2064	−2·4923	−2·1263	−1·8705	−2·859	−2·3898	−2·7067	4·3923	2·0128	20·647
($i = 3$)		−1·1368		−2·1106				−2·0762				
Larson–Miller	3b	−1·2248	−0·8541	−2·7242	−1·9317	−1·6759	−3·6571	−2·8533	−4·4249	6·423	2·3112	23·2239
($i = 3$)		−1·1955		−2·3424				−2·5397				
Dorn–Larson–	4	−0·382	−0·0271	−0·725	0·3839	0·6397	0·7584	2·2304	2·7098	1·8942	1·0088	10·188
Miller ($i = 3$)		−0·3528		−0·3433				2·544				
Manson–Haferd	6b	−0·8337	−0·6979	−1·5719	−0·9075	−0·6517	−0·6216	0·4225	1·0862	0·8411	0·8658	8·8592
($i = 3$)		−0·8045		−1·1902				0·7361				
MLP network	...	−0·286	−0·4301	−0·6431	−1·0744	−0·8186	−0·0351	−1·2279	0·2442	0·4585	0·5629	6·1829
		−0·2568		−0·2613				−0·9143				

*Each error term is difference between actual log rupture time and that predicted by specified model at given T and σ (equation (14a)).
† *See* equations (14b–d) for definitions.



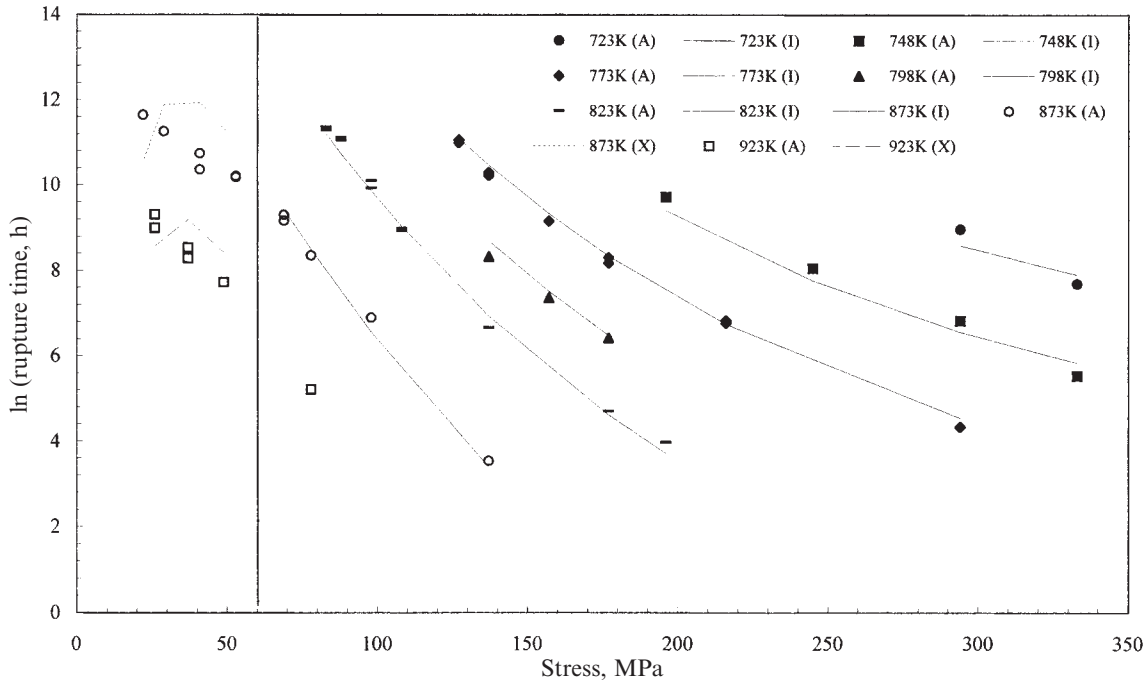
4 Interpolative and extrapolative capabilities of Larson–Miller technique using *a* quadratic and *b* cubic stress functions: A actual data, I interpolative prediction, X extrapolated prediction

head of the table. Each error term is the difference between the actual log rupture time and that predicted by the specified model at the stated stresses and temperatures (equation (14a)). The models presented include all the Soviet models, but in their simplified forms (equations (7b) and (7c)), that is, with all statistically insignificant variables removed. It also includes the minimum commitment model as given by equation (7a) but again with all insignificant variables removed.

The other parametric methods presented in Table 2 are the Dorn model as given by equation (2b), the Larson–Miller model as given by equation (3b), the Dorn–Larson–Miller model as given by equation (4), and the Manson–Haferd model as given by equation (6b). For these models the stress function is taken to be a cubic ($i = 3$). In the final column of Table 2 the errors obtained from extrapolating the neural network are shown.

The first half of Table 1 gives the least squares estimates for the Soviet model 1 (equation (7b)) using training data set 1. There are no signs of mis-specification as shown by a high R^2 value and normally distributed residuals having a variance that is homoscedastic. However, the variable $1/T$ appears to be statistically insignificant and when deleted for the simplified model, all remaining variables appear highly significant. The values for E_p and H_w again suggest that this simplified model is correctly specified. With $R^2 = 98{\cdot}96\%$ the model fits the interpolative data set well. This is confirmed by Fig. 3, where the solid curves are the predictions made by the simplified model for log rupture time above 60 MPa, i.e. within the range of data used to estimate the model.

When the Soviet model presented in Table 1 is used to predict log rupture lives below 60 MPa a completely different impression emerges. The broken curves in Fig. 3 show the extrapolations made for these lower stresses at 873 and 923 K. These extrapolations are consistent with the interpolated predictions but unfortunately are very different from the actual log rupture times. The extrapolations are always too high and deteriorate with lower stresses. In the first row of Table 2 the actual size of these extrapolation errors is presented. On average the extrapolative predictions made by the Soviet model 1 differ from the actual log rupture times by more then 46%. In addition, there does not appear to be any scope to improve upon this within the parametric framework. For example, Table 3 gives the maximum likelihood estimates for the Soviet



5 Interpolative and extrapolative capabilities of Manson–Haferd method: A actual data, I interpolative prediction, X extrapolated prediction

model 1 which also uses all unfailed specimens and assumes a Weibull rather than a log–normal distribution. The average absolute percentage error in the extrapolations decreases only to 42·26%. Unfailed specimens help improve the forecasting ability of parametric techniques but not to any substantial extent.

Evidently Soviet model 1 is highly effective in modelling the rupture times presented to it for estimation purposes, but is totally inadequate for predicting data points not used in its estimation. The model is not sufficiently parsimonious to identify general relations that are of use in extrapolation. This inability to generalise, or tendency to overfit the interpolative data set, is a characteristic of all the parametric techniques considered in the present paper.

This problem reflects a conflict that seems to exist between attempting a best fit to the interpolative data and attempting to extrapolate successfully. For example, consider the Larson–Miller model presented in the second half of Table 1. The inclusion of σ^2/T results in a greatly improved fit to the interpolative data set as shown by the increase in R^2 from 75·23% to 96·99%. The errors remain normally distributed, although heteroscedasticity can only be rejected at the 1% significance level. The extrapolations produced by this quadratic version of the Larson–Miller model are shown in Fig. 4*a*. They are fairly accurate (especially at 923 K). The average absolute percentage error in these extrapolations is 5·62%. Compared with the Soviet model 1 in Fig. 3, however, there is much more scatter in the data points around the interpolated prediction curves.

The situation is worse however in that without the foresight of data points held back for extrapolative tests, there is no guarantee that this version of the Larson–Miller model would have been selected for extrapolation. For example, the variable σ^3/T is shown to be statistically significant in Table 1, the R^2 value increases to 99·17%, and any indication of homoscedasticity in the residuals is removed (although the residuals are no longer normally distributed). Figure 4*b* shows the results of using this version of the Larson–Miller model to extrapolate below 60 MPa. The extrapolated predictions have deteriorated substantially, with an average absolute percentage error of 23·22% (*see* Table 2). However, the interpolative fit is much improved.

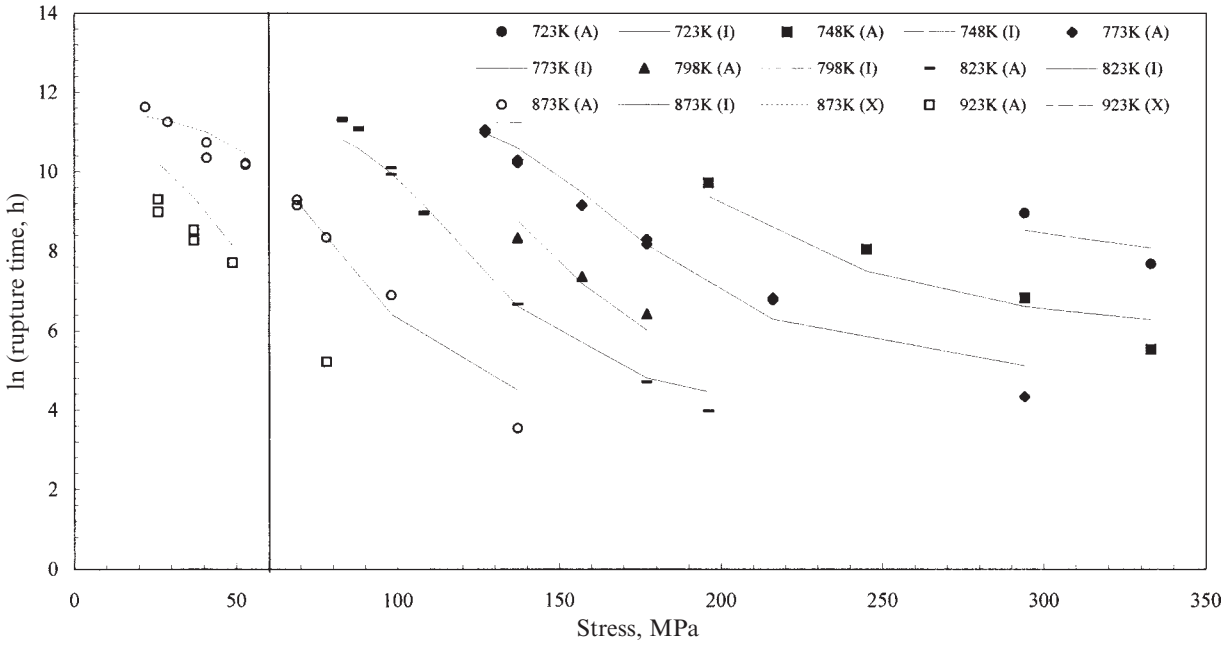
Consider one further parametric technique, namely, the Manson–Haferd model (equation (6b)) given in Table 2. This appears to produce the best forecasts of all the parametric techniques, with an average absolute percentage error in these extrapolations of 8·86%. However, these figures are deceptive. Consider Fig. 5, which shows the extrapolations obtained using this model. The extrapolations are good in that they 'wrap around' the failure times below 60 MPa. The predictions fail to identify the trend of increasing failure times with lower stresses, and actually predict lower failure times at the lower stresses. Evidently using this model to predict at even lower stresses would lead to ridiculous and misleading predictions.

In summary, parametric techniques tend to overfit the interpolative data and consequently are not sufficiently parsimonious to produce accurate extrapolations. In addition, the standard diagnostic tests for model adequacy are not sufficiently good to prevent this overfitting. Worse still are the unusual extrapolations produced by some parametric variants.

Contrast these observations now with the predictions obtained from the MLP network. Using stress and temperature from training data set 1 as the inputs, the genetic algorithm identified a network with five neurons in one hidden layer as the optimal structure. This architecture was optimised using a learning rate of 0·92 and a momentum value of 0·9 (again determined by the genetic algorithm).

Table 3 Maximum likelihood estimates for full Soviet model 1 using Weibull distribution

Variable					Diagnostic		
Constant	$1/T$	$\ln(T)$	$\ln(\sigma)$	σ/T	MSE	MAE	MPAE, %
394·62	97·57	−51·31	−9·05	7·40	20·7863	4·1422	42·2579



6 Interpolative and extrapolative capabilities of genetically tuned MPL network: A actual data, I interpolative prediction, X extrapolated prediction

The mean square error for the trained network was 0·0012. The last row of Table 2 gives the predictions made by this trained network for the failure times below 60 MPa (which were not used in training). The average absolute percentage error in these extrapolations is 6·18%, which is far superior to any of the parametric techniques.

These extrapolations are shown in Fig. 6. Not only is the average percentage error low but the direction of the extrapolations is good. For the failure times below 60 MPa and at a temperature of 873 K the extrapolation is truly impressive. The extrapolation at 923 K starts well but tends to overpredict at lower stresses. The degree of overprediction is however much less than those obtained via the parametric techniques (*see* Fig. 3 for a comparison). Remember also that at the highest temperature there is only one failure time at a stress above 60 MPa from which the network could identify a pattern relating failure times to stress, and the extrapolations shown at this temperature in Fig. 6 are therefore all the more remarkable. This tends to suggest that the network has identified some relationships between stress and rupture life at temperatures below 923 K that are of use in extrapolating to lower stresses at 923 K. With more test data at this temperature there is evidently potential for improvements on the present extrapolations.

Conclusions

Old and more recent parametric techniques used for predicting failure times have been compared with a traditional MPL neural network. The parametric techniques were shown to produce poor extrapolations because of their tendency to overfit the data sets used for their estimation. Standard statistical measures of model inadequacy were of little use in overcoming this problem. Some of the more non-linear techniques such as the Manson–Haferd model produced implausible extrapolations.

In contrast, a genetically optimised MLP network was able to extrapolate well outside the range of data used to train it. The average extrapolation error was shown to be about 6%, compared with some 45% for the Soviet models and 20% for the Larson–Miller based techniques. The network appears to be able to identify general patterns in the training set that are useful for extrapolation purposes and with carefully selected training data sets, improved long term predictions appear possible.

References

1. N. J. GRANT and A. G. BUCKLIN: in 'Deformation and fracture at elevated temperatures', (ed. N. J. Grant and A. W. Mullendore); 1963, Boston, MA, MIT Press.
2. J. GLEN: *J. Metall. Club*, 1956–57, (9), 15–18.
3. A. MENDELSOHN and S. S. MANSON: *J. Basic. Eng.*, 1960, **82**.
4. P. BROZZO: Proc. ASME/IMechE Joint Conf. on 'Creep'; 1963, New York/London, Institution of Mechanical Engineers.
5. R. W. EVANS: *Mater. Sci. Technol.*, 1989, **5**, 699–707.
6. F. R. LARSON and J. MILLER: *Trans. ASME*, 1952, **174**, (5).
7. M. YAMAUCHI, Y. CHUMAN, N. NISHIMURA, and F. MASUYAMA: in Proc. 6th Int. Conf. on 'Creep and fatigue', 511–519; 1996, London, The Institution of Mechanical Engineers.
8. R. W. EVANS and B. WILSHIRE: 'Introduction to creep', Chap. 4; 1993, London, The Institute of Materials.
9. F. C. MONKMAN and N. J. GRANT: *Proc. ASTM*, 1956, **56**, 593.
10. F. GAROFALO: *Proc. ASTM*, 1956, **56**, 612.
11. M. EVANS: *Mater. Sci. Technol.*, 1999, **15**, (1), 91–100.
12. S. KANG: 'An investigation of the use of feed forward neural networks for forecasting', PhD thesis, Kent State University, Kent, OH 44240, USA, 1991.
13. M. LI, X. LIU, S. WU, and X. ZHANG: *Mater. Sci. Technol.*, 1998, **14**, 136–138.
14. A. Y. BADMOS, H. K. D. H. BHADESHIA, and D. J. C. MACKAY: *Mater. Sci. Technol.*, 1998, **14**, 793–809.
15. H. FUJII, D. J. C. MACKAY, and H. K. D. H. BHADESHIA: *ISIJ Int.*, 1996, **36**, 1373–1382.
16. NRIM Creep Data Sheet No. 3B, National Research Institute for Metals, Tokyo, Japan, 1986.
17. J. E. DORN and L. A. SHEPHERD: Proc. Symp. on 'The effect of cyclical heating and stressing on metals at elevated temperatures', STP 165; 1954, Philadelphia, PA, ASTM.
18. S. S. MANSON and A. M. HAFERD: 'A linear time–temperature relation for extrapolation of creep and stress–rupture data', NACA, TN, 1953, **2890**.
19. R. W. BAILEY: *Proc. Inst. Mech. Eng.*, 1935, **131**, 131.
20. S. S. MANSON and W. F. BROWN, Jr: *Proc. ASTM*, 1953, **53**.
21. S. S. MANSON: Proc. ASME/ASTM/IMechE Joint Conf. on 'Creep'; 1963, New York/London, Institution of Mechanical Engineers.

22. H. EYRING, S. GLASSTONE, and K. J. LAIDLER: 'The theory of rate processes'; 1941, New York, McGraw-Hill.
23. S. S. MANSON and U. MURALDIHAN: 'Analysis of creep rupture data in five multi heat alloys by minimum commitment method using double heat term centring techniques', Research Project 638–1, EPRI Cs–3171, Palo Alto, CA, USA, July 1983.
24. I. I. TRUNIN, N. G. GOLOBOVA, and E. A. LOGINOV: Proc. 4th Int. Symp. on 'Heat resistant metallic materials', Mala Fatra, CSSR, 1971, Vol. 168.
25. S. R. HOLDSWORTH: 'Review of WG1 evaluation of creep rupture data measurement methods recommendation validation', Appendix C1, European Creep Collaborative Committee, ERA Technology Ltd, Surrey, 1995.
26. K. SWINGLER: 'Applying neural networks: a practical guide'; 1996, London, Academic Press.
27. J. L. ELMAN: *Cogn. Sci.*, 1990, **14**, 179–211.
28. L. TARASSENKO: 'A guide to neural computing applications'; 1998, New York, Wiley.
29. J. A. DOORNIK and H. HANSEN: 'An omnibus test for univariate and multivariate normality', Discussion paper, Nuffield College, Oxford, 1994.
30. H. WHITE: *Econometrica*, 1980, **48**, 817–838.
31. L. DAVIS: 'Handbook of genetic algorithms'; 1991, New York, Van Nostrand Reinhold.
32. T. MASTERS: 'Neural, novel and hybrid algorithms for time series prediction', Chap. 9; 1995, New York, John Wiley.