Tensile properties of mechanically alloyed oxide dispersion strengthened iron alloys Part 1 – Neural network models

A. Y. Badmos, H. K. D. H. Bhadeshia, and D. J. C. MacKay

A neural network technique trained within a Bayesian framework has been applied to the analysis of the yield strength, ultimate tensile strength, and percentage elongation of mechanically alloyed oxide dispersion strengthened ferritic steels. The database used was compiled using information from the published literature, consisting of variables known to be important in influencing mechanical properties. The analysis has produced patterns which are metallurgically reasonable, and which permit the quantitative estimation of mechanical properties together with an indication of confidence limits. *MST/3791*

At the time the work was carried out Dr Badmos and Dr Bhadeshia were in the Department of Materials Science and Metallurgy, University of Cambridge. Dr Badmos is now in the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada. Dr MacKay is in the Cavendish Laboratory, University of Cambridge. *Manuscript received 26 March 1997; in final form 15 May 1997.*
© 1998 The Institute of Materials.

Introduction

Mechanical alloying provides a technique for the production of dispersion strengthened alloys having interesting chemical compositions and unique properties. The process consists of blending elemental powders with master alloy powders in an attritor, and ball milling the mixture under a dry and protective atmosphere. The resulting alloyed powder is highly cold worked and can be homogeneous in its microstructure.[1] On completion of mechanical alloying, the powder is canned, extruded, and hot rolled to bars or hot and cold rolled to sheet.

Mechanically alloyed oxide dispersion strengthened (MA-ODS) alloys are already available in commercial quantities. The technique is nevertheless somewhat unusual and a large number of important phenomena are not yet understood. Experimental measurements of the mechanical properties of the alloys, for example, have not been systematically coordinated to reveal a definite pattern with respect to the numerous variables thought to be important in understanding the service behaviour of the alloys. The objective of the present work was to investigate whether an artificial neural network[2] could be trained to predict the ultimate tensile strength, yield strength, and elongation of MA-ODS ferritic steels as a non-linear function of these variables. In Table 1 the chemical compositions are given of some commercial MA-ODS steels which are most represented in the data used for the present analysis.[3–11]

This paper is presented first with an introduction to the method, followed by a description of the process by which the optimum model was obtained in each instance. The final section is concerned with the use of the models to ensure that the perceived relationships are metallurgically significant.

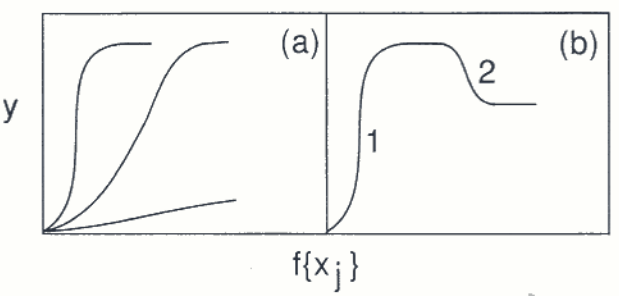
Technique

Most workers are familiar with regression analysis, where data are 'best fitted' to a specified relationship which is usually linear. The result is an equation in which each of the inputs x_j is multiplied by a weight w_j. The sum of all such products and a constant θ then gives an estimate of the output $y = \Sigma_j w_j x_j + \theta$. It is well understoood that there are dangers in using such relationships beyond the range of fitted data.

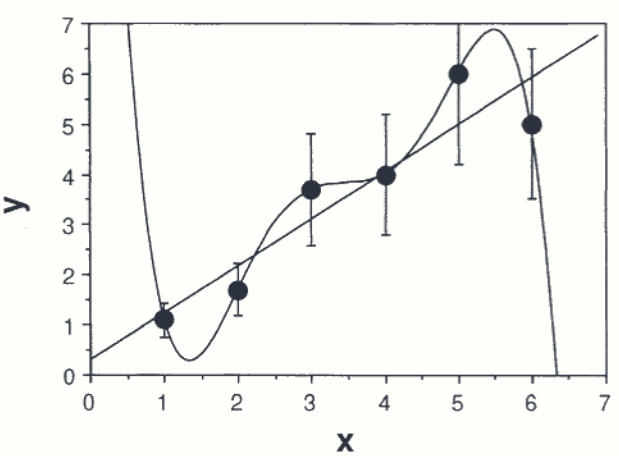
A neural network is a more general method of regression analysis. As before, the input data x_j are multiplied by weights, but the sum of all these products forms the argument of a hyperbolic tangent. The output y is therefore a non-linear function of x_j, the function usually selected being the hyperbolic tangent because of its flexibility. The exact shape of the hyperbolic tangent can be varied by altering the weights (Fig. 1*a*). Further degrees of non-

Table 1 Chemical composition of some commercial MAODS steels, wt-%

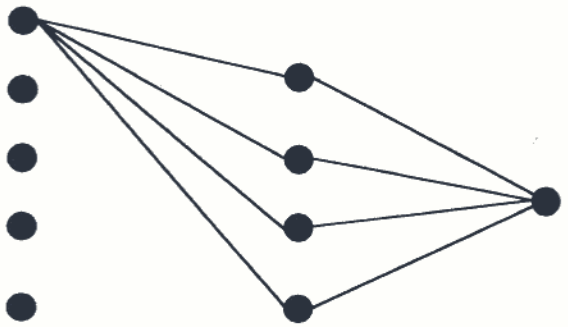
Steel	Cr	Al	Mo	Ti	Y_2O_3	Fe
MA956	20·0	4·5	...	0·5	0·5	Bal.
MA957	14·0	...	0·3	1·0	0·27	Bal.
DY (DT2203Y05)	13·0	...	1·5	2·2	0·5	Bal.
DT (DT2906)	13·0	...	1·5	2·9	...	Bal.



1 *a* three different hyperbolic tangent functions – 'strength' of each depends on weight, and *b* combination of two hyperbolic tangents to produce more complex model



2 Complex model shown may fit data, but in this instance a linear relationship may be all that is justified by noise in data



3 Typical network used in analysis: connections originating from only one input unit are illustrated and two bias units are not illustrated

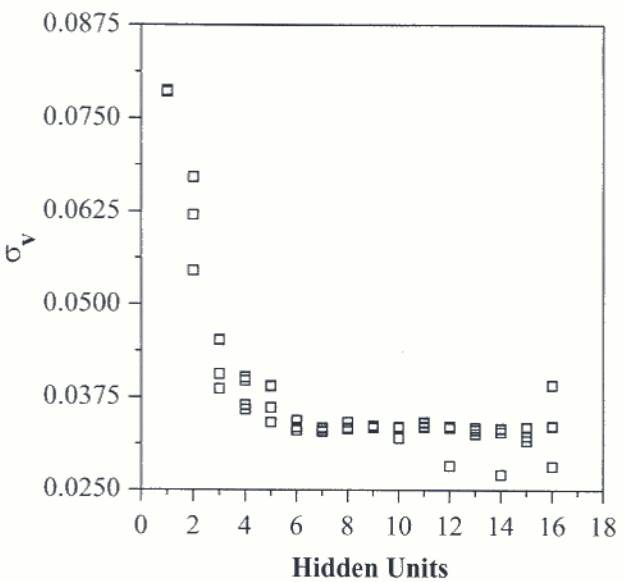
linearity can be introduced by combining several of these hyperbolic tangents (Fig. 1*b*), so that the neural network method is able to replicate almost arbitrarily non-linear relationships. It is well known that the effect of chromium on the microstructure of steels is markedly different at high concentrations than in dilute alloys. Ordinary regression analysis cannot cope with such changes in the form of relationships.

A neural network is 'trained' using a set of examples of input and output data. The outcome of the training is a set of coefficients (weights) and a specification of the functions which in combination with the weights relate the input to the output. The training process involves a search for the optimum non-linear relationship between the input and the output data and is computer intensive. Once the network is trained, estimation of the outputs for any given inputs is very rapid.

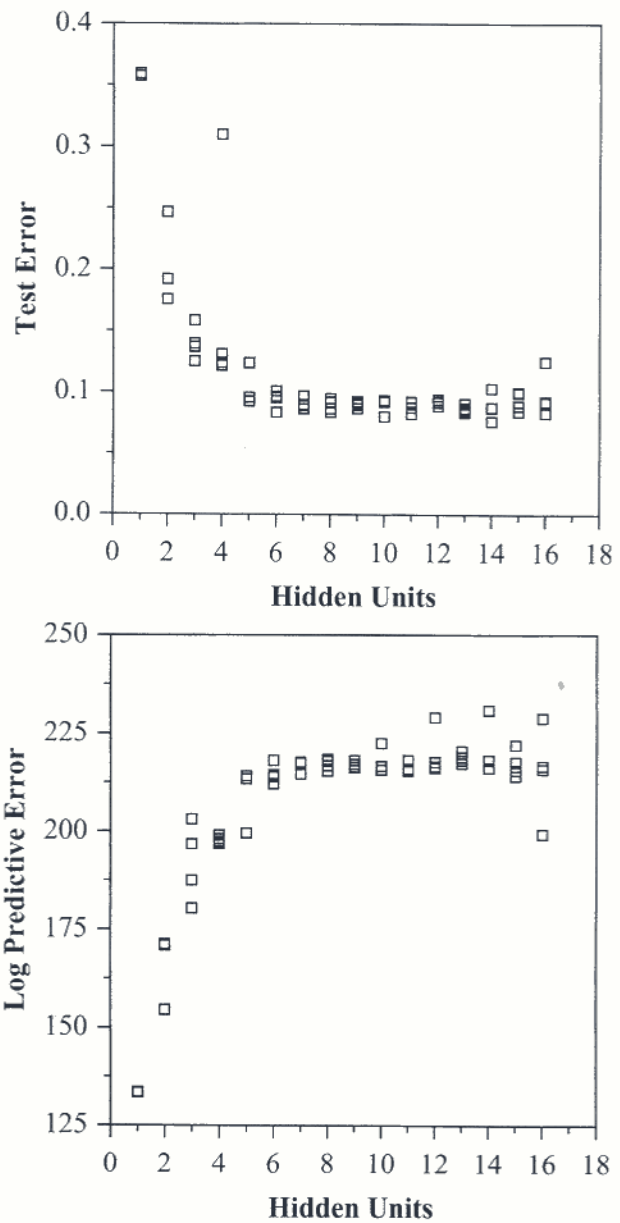
One of the difficulties associated with blind data modelling is that of 'overfitting', in which spurious details and noise in the training data are overfitted by the model (Fig. 2). This results in solutions that generalise poorly. MacKay[12–16] has developed a Bayesian framework for neural networks in which the appropriate model complexity is inferred from the data.

The Bayesian framework for neural networks has two further advantages. First, the significance of the input variables is automatically quantified. Consequently the significance, perceived by the model, of each input variable can be compared against metallurgical theory. Second, the predictions of the network are accompanied by error bars which depend on the specific position in input space. These quantify the certainty of the model regarding its predictions.

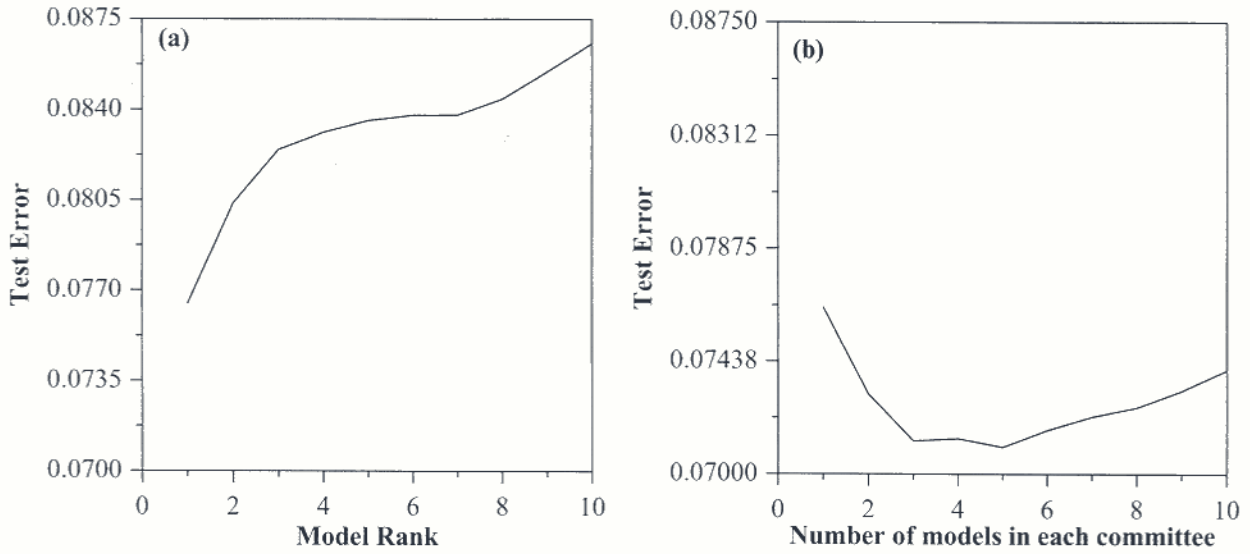
The neural network method has recently been applied to many materials problems, for example: the impact toughness of C–Mn steel arc welds by Bhadeshia *et al.*;[17] an analysis of the strength of nickel base superalloys by Jones and MacKay;[18] austenite formation in steels by Gavard *et al.*;[19] yield and ultimate tensile strength of steel welds by Cool *et al.*;[20] fatigue crack growth rate in nickel base superalloys by Fujii *et al.*;[21] mechanical properties in the heat affected zone of power plant steels by Cool and Bhadeshia;[22] prediction of martensite start temperature by



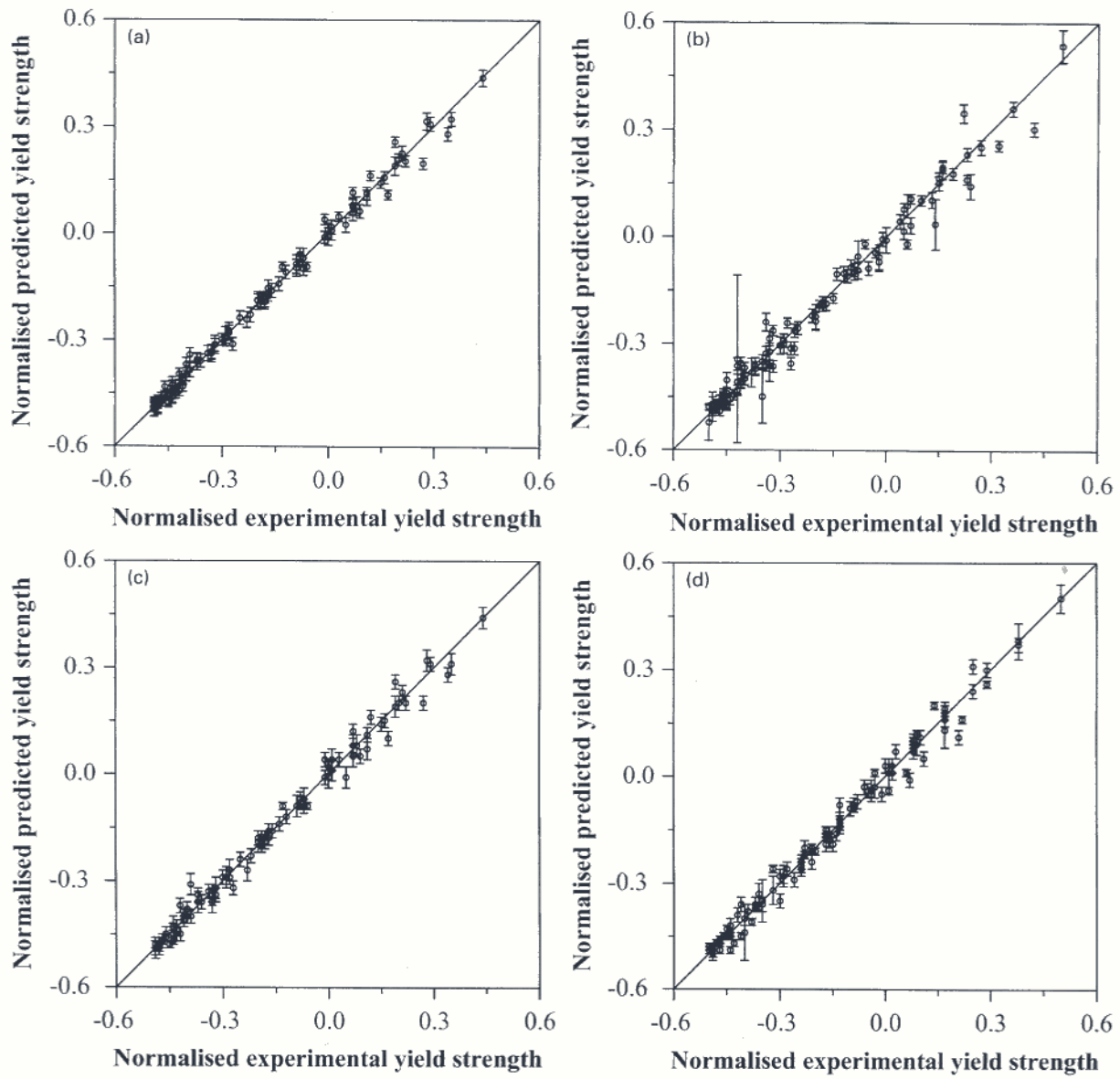
4 Yield strength analysis: variation in model perceived noise level σ_v as function of number of hidden units (i.e. complexity of model) – several values are presented for each set of hidden units because training for each network was started with various random seeds



5 Yield strength analysis: variation in test error and log predictive error as functions of number of hidden units – note that larger log predictive error indicates superior model



6 *a* test errors of top 10 yield strength models, and *b* corresponding test errors for committee models



a single model, training data set; *b* single model, test data set; *c* committee, training data set; *d* committee, test data set

7 Normalised predicted yield strength versus normalised experimental results using *a*, *b* single best model and *c*, *d* optimum committee

Vermeulen *et al.*;[23] prediction of the continuous cooling transformation diagram of some selected steels by Vermeulen *et al.*;[24] and prediction of the measured temperature after the last finishing stand in hot rolling by Vermeulen *et al.*[25]

Analysis

Both the input and output variables were normalised within the range +0·5 to −0·5. The normalisation is obtained through a procedure which is expressed quantitatively as

$$x_N = \frac{x - x_{min}}{x_{max} - x_{min}} - 0{\cdot}5 \quad . . . . . (1)$$

where x_N is the normalised value of x, which has the minimum and maximum values given by x_{min} and x_{max} respectively. The normalisation is not necessary for the analysis but it facilitates the subsequent comparison of the significance of each of the variables.

Figure 3 shows a typical network. Each network consists of input nodes (one for each variable x), a number of hidden nodes, and an output node. Linear functions of the inputs x_j are operated on by a hyperbolic tangent transfer function

$$h_i = \tanh\left(\sum_j w_{ij}^{(1)} x_j + \theta_i^{(1)}\right) \quad . . . . . (2)$$

so that each input contributes to every hidden unit. The bias is designated θ_i and is analogous to the constant that appears in linear regression analysis. The strength of the transfer function is in each case determined by the weight w_{ij}. The transfer to the output y is linear

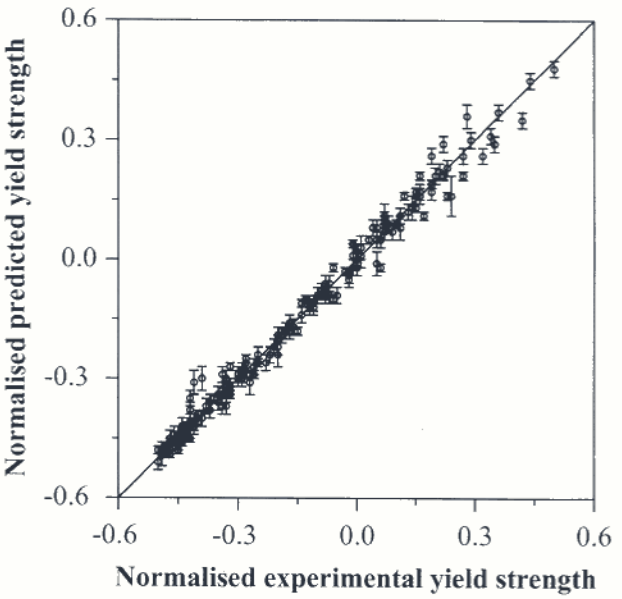
$$y = \sum_i w_{ij}^{(2)} h_i + \theta^{(2)} \quad . . . . . (3)$$

The specification of the network structure, together with the set of weights, is a complete description of the formula relating the input to the output. The weights are determined by training the network. The training is carried out using a data set $D = \{x^{(m)}, t^{(m)}\}$ by adjusting the weights w to minimise an error function, e.g.

$$E_D(w) = \frac{1}{2}\sum_m \sum_i [t_i^{(m)} - y_i(x^{(m)}; w)]^2 \quad . . . . . (4)$$

This objective function is a sum of terms, one for each input–target pair $\{x, t\}$, measuring the degree of correlation between the output $y\{x; w\}$ and the target t (Ref. 16). The parameter m denotes each input–output pair.

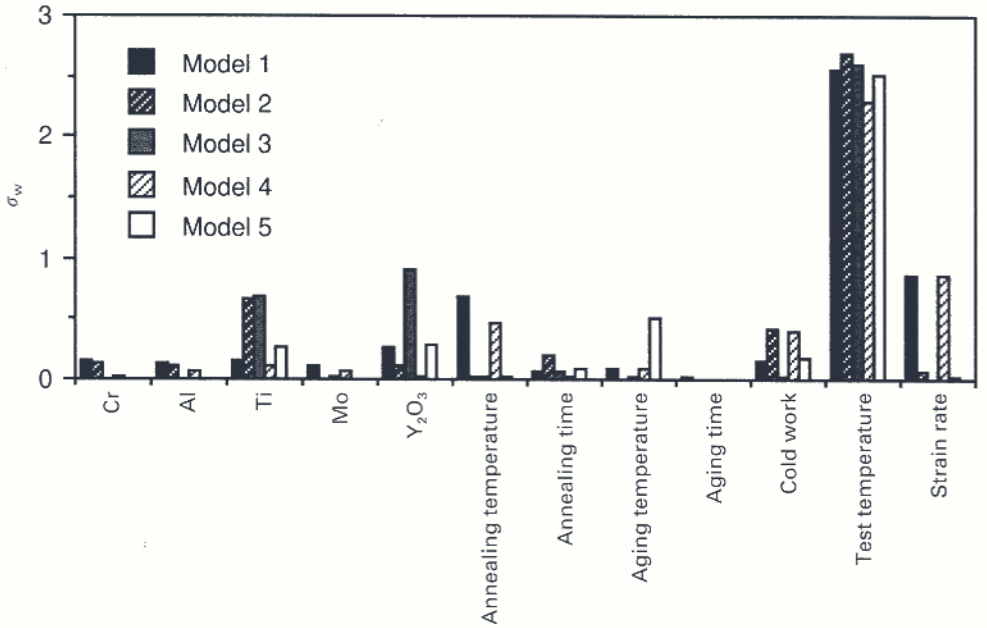
The training for each network is started with a variety of random seeds. The value of a term σ_v gives the framework estimate of the overall noise level of the data. The complexity of the model is controlled by the number of hidden units and the values of the regularisation constants (σ_w), one associated with each of the inputs, one for biases, and one for all weights connected to the output. The noise level decreases monotonically as the number of hidden units increases. However, the complexity of the model also increases with the number of hidden units. A high degree of complexity may not be justified if the model attempts to fit the noise in the experimental data. MacKay[12,13,15,16] has made a detailed study of this problem and defined a quantity (the 'evidence') which acts as an indicator of the probability of a model. In circumstances where two models give similar results for the known data set, the more probable model would be predicted to be that which is simpler; this simple model would have a higher value of evidence. The evidence framework is used to control the regularisation constants and σ_v. The number of hidden units is set by examining performance on test data. A combination of Bayesian and pragmatic statistical techniques is therefore used to control the complexity of the model. A further procedure used to avoid the overfitting problem was to divide the experimental data randomly into two equal sets, namely, the training and test data sets. The models are developed using the training data only. The unseen test data are then used to assess how well the model generalises. A good model would produce similar levels of error in both the test and training data whereas an overfitted model might accurately predict the training data but badly estimate the unseen test data. Once the correct complexity of the model has been determined using this procedure, it can be retrained using all the data with a small but significant reduction in the error.



8 Normalised predicted yield strength versus normalised experimental results for whole data set after retraining using optimum committee

Table 2 Variables used in analysis of yield strength

Variable	Range	Mean	Standard deviation
Chromium, wt-%	13–20	17·30	3·20
Aluminium, wt-%	0–4·5	2·62	2·23
Titanium, wt-%	0·5–3·50	1·03	0·86
Molybdenum, wt-%	0–1·5	0·35	0·56
Yttria, wt-%	0–0·5	0·41	0·15
Recrystallisation temperature, °C	20–1330	697	595
Recrystallisation time, s	0–120	28·44	33·38
Aging temperature, °C	20–800	163·3	303
Aging time, s	0–2888	327	739
Cold work, %	0–70	10·43	19·64
Test temperature, °C	0–1200	562·1	340·4
Strain rate, s^{-1}	3×10^{-8}–3×10^{-2}	$9{\cdot}89 \times 10^{-4}$	$2{\cdot}5 \times 10^{-3}$
Yield strength, MPa	63–1600	497	388



9 Model perceived significance of input parameters for committee model trained on all yield strength data; σ_w values for all members of committee are presented for each variable

The test error T_{en} is a reflection of the ability of the model to predict the target values in the test data

$$T_{en} = 0{\cdot}5 \sum_{n} (y_n - t_n)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

where y_n is the set of predictions made by the model, and t_n the corresponding target (experimental) values previously unseen by the model.

It is popular to use the test error (sum squared error) as the default performance measure whereby the model with the lowest test error is considered to be the best.[16] In many applications there will be an opportunity to make a prediction with error bars rather than a simple scalar prediction, or maybe to carry out an even more complex predictive procedure. It is then reasonable to compare models in terms of their predictive performance as measured by the log predictive probability of the test data. Under the log predictive error, as contrasted with the test error, the penalty for making a 'wild' prediction is much less if the wild prediction is accompanied by appropriately large error bars. Assuming that for each example m the model gives a prediction with error $(y^{(m)}, \sigma^{(m)^2})$, the log predictive error (LPE) is

$$\mathrm{LPE} = \sum_{m} \left\{ \frac{1}{2}(t^{(m)} - y^{(m)})^2 \Big/ \sigma_y^{(m)^2} + \log[(2\pi)^{1/2}\sigma_y^{(m)}] \right\} \quad \ldots\ldots\ldots (6)$$

When making predictions, MacKay[16] has recommended the use of multiple good models instead of just one best model. This is termed 'forming a committee'. The committee prediction $\bar{y}$ is obtained using the expression

$$\bar{y} = \frac{1}{L} \sum_{i} y_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

where L is the size of the committee and y_i is the estimate of a particular model i. The optimum size of the committee is determined from the validation error of the committee's predictions using the test data set. The test error of the predictions made by a committee is calculated by replacing the y_n in equation (5) with $\bar{y}$.

Table 3 Ranking by test error of 10 best models of yield strength

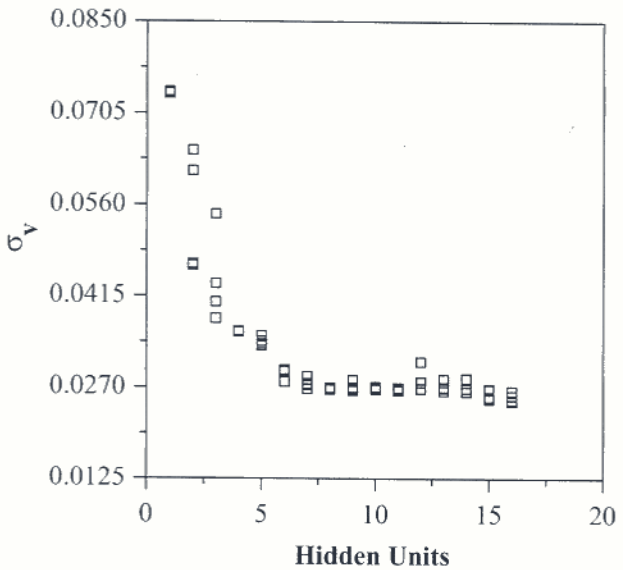
Ranking	Hidden units	Seed	Test error	Log predictive error
1	14	100	0·0764770	230·75
2	10	100	0·080370	222·37
3	11	100	0·082443	217·96
4	16	30	0·083116	228·67
5	6	30	0·083576	218·03
6	8	100	0·083781	217·81
7	13	10	0·083790	220·32
8	15	30	0·084411	221·84
9	13	100	0·085464	218·76
10	6	10	0·086566	218·26

Table 4 Test errors of committees organised for estimation of yield strength: note that test error of best committee is less than that of single best model

Number of models in committee	Test error
1	0·07648
2	0·07309
3	0·07128
4	0·07137
5	0·07104
6	0·07170
7	0·07221
8	0·07257
9	0·07322
10	0·07403

Yield strength model

The technique was applied to the variables given in Table 2 for the analysis of the yield strength. There were 232 data, 12 input variables, and one output which is the yield strength. The major alloying elements (chromium, aluminium, titanium, and molybdenum) are expected to influence the yield strength primarily via solid solution strengthening. In some alloys molybdenum, titanium, and chromium also precipitate as an intermetallic compound, χ phase (FeCrTiMo) after a low temperature aging treatment.[8] Yttrium oxide is present as a very fine dispersion and must enhance strength at all temperatures by impeding the glide of dislocations. The recrystallisation heat treatment has a very severe effect on the microstructure since it changes an ultrafine primary recrystallised grain structure to a structure that is coarse and columnar. Cold work is naturally



10 Ultimate tensile strength analysis: variation in model perceived noise level σ_v as function of hidden units (i.e. complexity of model) – several values are presented for each set of hidden units because training for each network was started with various random seeds

expected to increase the yield strength; the data set included a variety of methods of cold deformation including rolling and swaging. The yield strength of body centred cubic metals is particularly sensitive to temperature because of the large Peierls barriers to dislocation motion. A further temperature dependence originates from the possibility of the climb of dislocations over dispersoids. There may in unrecrystallised alloys be an additional effect due to the onset of dynamic recrystallisation.

The plot of σ_v as a function of the complexity of the models is shown in Fig. 4. Note that a number of values are presented for each hidden unit because the training process was started using different randomly selected seeds which determine the starting values of the weights. The test error and log predictive error versus number of hidden units are shown in Fig. 5.

The numerical data for the top ten models ranked by their test error are given in Table 3. Table 4 gives the test errors of the ten committees formed, starting with the best model and progressively increasing the number of models in the committee. The plots of the test errors of the top ten models and those of the committees are shown in Fig. 6 plotted on the same scale to show the usual reduction in test error when an appropriate committee is formed. The committee consisting of five top ranking models has the least test error and was used for the study of the yield strength presented below.

The plots of the predicted values versus experimental values using the training and test data sets for the single best model and the committee are shown in Fig. 7. Figure 8 shows the plot for the committee after retraining using the whole data set.

The significance of each of the input variables perceived by the various models contained in the committee is shown by σ_w in Fig. 9. The parameter σ_w is rather like a partial correlation coefficient in that it represents the amount of variation in the output that can be attributed to any particular input parameter and does not necessarily represent the sensitivity of the output to each of the inputs. As expected, the yield strength correlates strongly with temperature.

Ultimate tensile strength

The neural network technique was applied to the ultimate tensile strength (UTS) data given in Table 5. The variables are identical to those used for the yield strength analysis. It would ideally be of interest to include the strain hardening coefficient since this determines the plastic instability which defines the ultimate strength. However, no such data could be found in the published literature. There were 12 input variables and one output variable, the ultimate tensile strength. A total of 232 data were used. As before, the data were divided equally and randomly into test and training data sets, the training data set being used to train the model and the ability of the model to generalise being examined by checking its performance on the unseen test data.

Figure 10 shows the plot of σ_v versus hidden units. As expected, the inferred noise level decreases monotonically as the number of hidden units increases.

Figure 11 shows the variation of test error and log predictive error as functions of the number of hidden units. The calculated test error reaches a minimum at 16 hidden units and the log predictive error also exhibits a maximum at the same number of hidden units. This would have been the optimum model had a single model been used for the analysis.

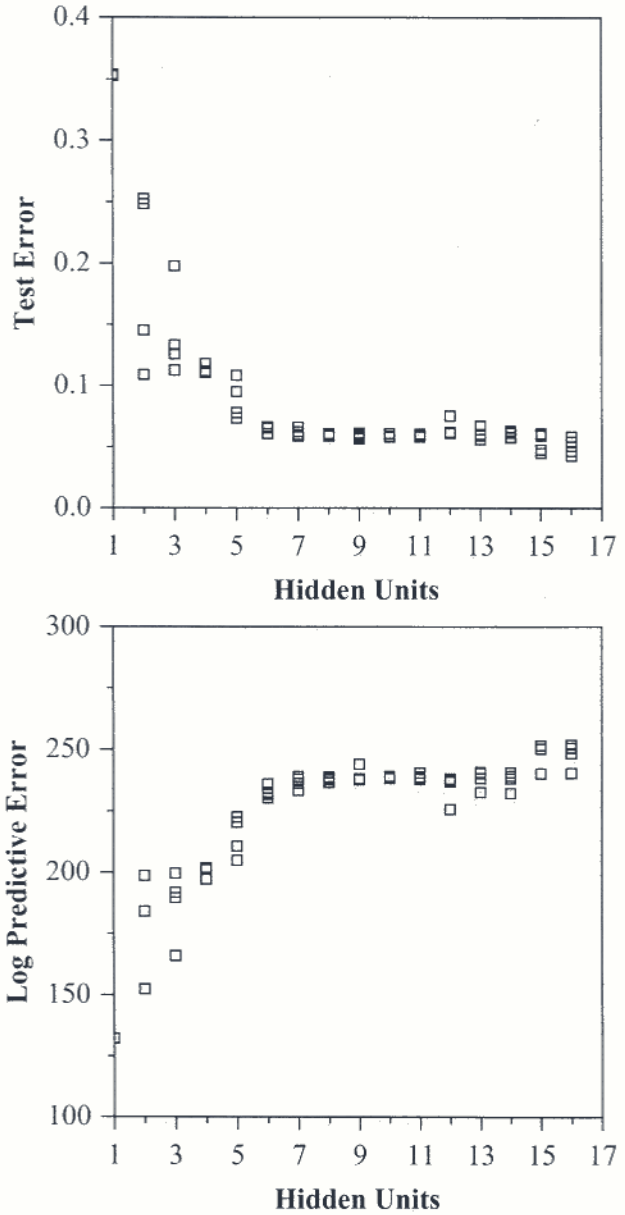
A committee model was used. Based on the values of the test error and log predictive error, four models were selected as the best. The models were ranked using their test error values as presented in Table 6.

The optimum number of models in the committee was determined from the calculated validation errors of the different possible committees. Figure 12 shows the variation of the test error of the best models as a function of their position on the ranking table and the test error of the committees as a function of the number of models. It is evident that forming a committee reduces the test error, and hence improves predictions.

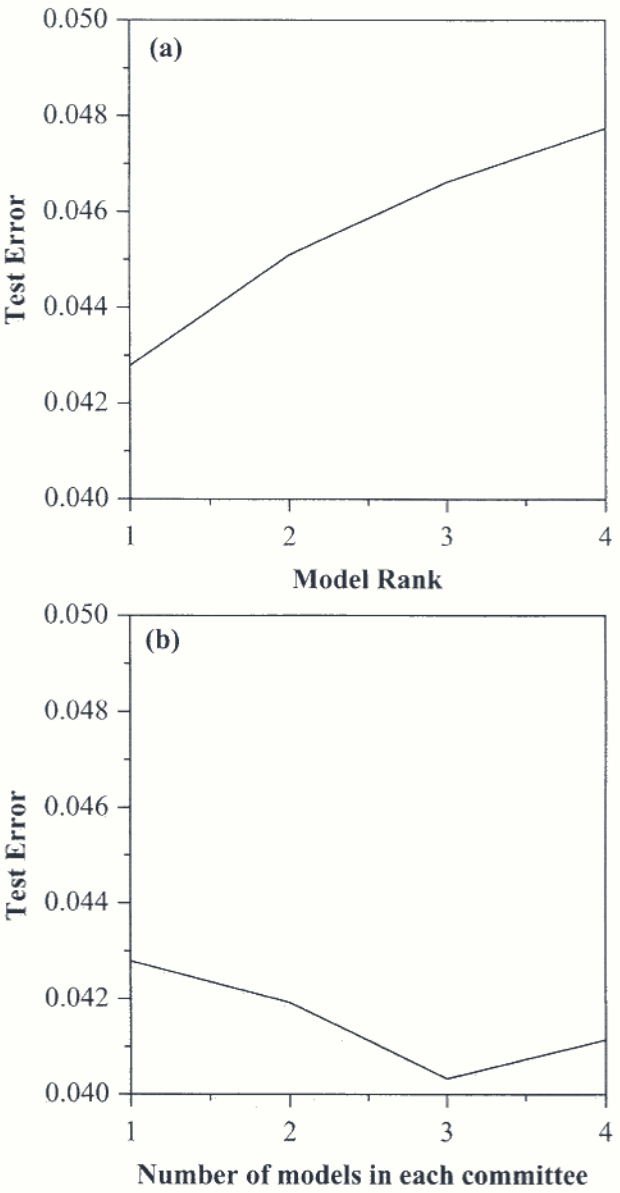
As shown in Fig. 12 the committee consisting of the top three models shows the least test error and was used

Table 5 Variables used in analysis of ultimate tensile strength (UTS)

Variable	Range	Mean	Standard deviation
Chromium, wt-%	13–20	17·19	3·22
Aluminium, wt-%	0–4·5	2·54	2·24
Titanium, wt-%	0·5–3·50	1·1	0·87
Molybdenum, wt-%	0–1·5	0·37	0·58
Yttria, wt-%	0–0·5	0·41	0·15
Recrystallisation temperature, °C	20–1330	684	594
Recrystallisation time, s	0–120	27·56	33·24
Aging temperature, °C	20–800	174·7	311·7
Aging time, s	0–2888	361	781
Cold work, %	0–70	10·47	19·71
Test temperature, °C	0–1200	561·1	347·6
Strain rate, s^{-1}	3×10^{-8}–3×10^{-2}	$1{\cdot}1 \times 10^{-3}$	$2{\cdot}5 \times 10^{-3}$
UTS, MPa	70·7–1680	575·3	407·3



11 Ultimate tensile strength analysis: variation in test error and in log predictive error as functions of number of hidden units – note that larger log predictive error indicates superior model



12 *a* test error of four best ultimate tensile strength models and *b* corresponding test errors for committees of models

for the analysis. The agreement between the predicted and experimental values for the training and test data sets is shown in Fig. 13 for the single best model and the committee.

The committee models were then retrained on the whole data set starting with the weights determined from the previous training exercise. Figure 14 shows the plot of the predicted values versus experimental values for the whole data set after the retraining. The retraining is shown to have significantly improved the model with the reduction in error bars and the apparent absence of outliers.

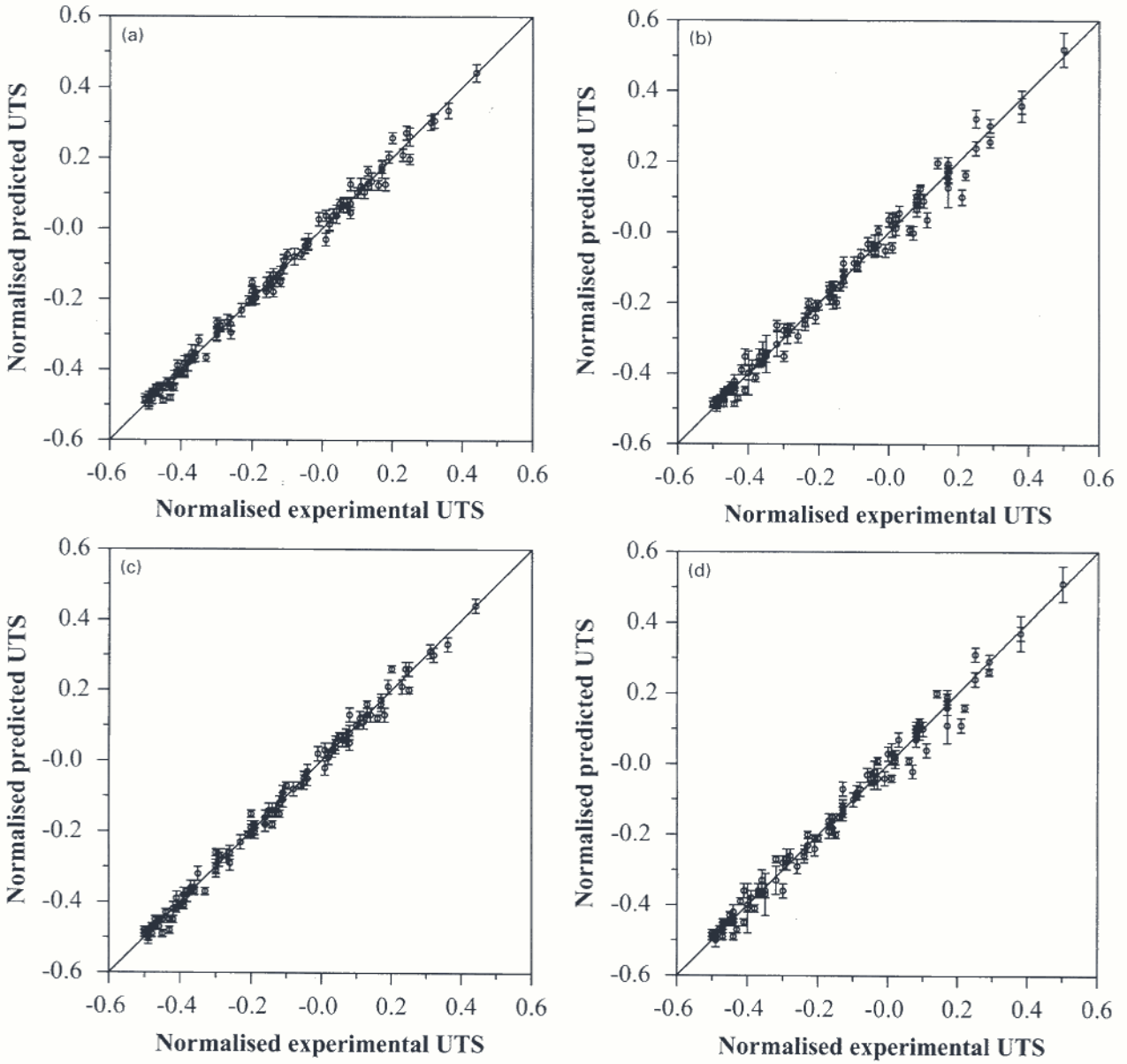
Table 6 Ranking according to test error for four best models of ultimate tensile strength

Ranking	Hidden units	Seed	Test error	Log predictive error
1	16	30	0·042786	250·73
2	15	30	0·045090	250·15
3	16	100	0·046607	250·49
4	15	100	0·047735	251·45

The significance of each of the input variables perceived by the various models contained in the committee is shown by σ_w in Fig. 15. The test temperature is shown to have the largest σ_w for all three models in the committee. This shows that the models have recognised a pattern correctly because temperature is more widely varied than any other input in the database. Moreover, temperature is known to affect strength very significantly.

Elongation model

The ultimate tensile strength and yield strength were included as input variables in addition to the 12 input variables used in the strength analysis (Table 7). The inclusion of the strength parameters became necessary after the initial attempts to train the network without these parameters failed to produce acceptable results. The reason for this behaviour is obvious, in that the ductility of a material is a function of strength. The database consists of 232 examples and the noise level in the data is as plotted in Fig. 16*a*.



a single model, training data set; *b* single model, test data set; *c* committee, training data set; *d* committee, test data set

13 Normalised predicted ultimate tensile strength versus normalised experimental results using *a*, *b* single best model and *c*, *d* optimum committee

Test error and log predictive error as functions of hidden units are shown in Fig. 16*b* and *c*.

The numerical data for the ten best models ranked according to their log predictive errors are given in Table 8. Although the ranking of the models was by their log predictive error the optimum committee was determined using the validation error. The test errors of the ten committees are given in Table 9 and Fig. 17 shows the test

Table 7 Variables used in analysis of elongation

Variable	Range	Mean	Standard deviation
Chromium, wt-%	13–20	17·45	3·19
Aluminium, wt-%	0–4·5	2·74	2·20
Titanium, wt-%	0·5–3·50	1·02	0·86
Molybdenum, wt-%	0–1·5	0·35	0·58
Yttria, wt-%	0–0·5	0·42	0·15
Recrystallisation temperature, °C	20–1330	729	585
Recrystallisation time, s	0–120	28·92	33·72
Aging temperature, °C	20–800	171	309
Aging time, s	0–2888	354·8	778
Cold work, %	0–70	11·4	22·59
Test temperature, °C	0–1250	561·4	340·2
Strain rate, s^{-1}	3×10^{-8}–3×10^{-2}	$1{\cdot}3 \times 10^{-3}$	$2{\cdot}7 \times 10^{-3}$
UTS, MPa	70·7–1680	545	390
Yield strength, MPa	63–1600	468	367
Elongation, %	0·8–49·29	12·13	8·18