CHAPTER FOUR

Thermodynamic Analysis of Mechanical Alloying

4.1 Introduction

It has always been a question as to when a mixture of powders becomes a solution during the mechanical alloying process. Normal thermodynamic theory for solutions begins with the mixing of component atoms. Many solutions are, however, prepared by mixing together lumps of the components, each of which might contain millions of identical atoms. We examine here the way in which a solution evolves from these large clusters of components, from a purely thermodynamic point of view. There are some interesting results including the prediction that solution formation by the mechanical alloying of solid components cannot occur unless there is a gain in coherency as the particles become small. The nature of the barrier to mechanical alloying is discovered. There is also the possibility of a metastable state prior to the achievement of full solution, when the component atoms prefer like–neighbours.

4.2 Thermodynamic Analysis

Consider the pure components A and B with molar free energies μ_A^o and μ_B^o respectively. If the components are initially in the form of powders then the average free energy of mixture of such powders is given by

$$G\{\text{mixture}\} = (1-x)\mu_A^o + x\mu_B^o \tag{4.1}$$

where x is the mole fraction of B. It is assumed that the powder particles are so large that the A and B atoms do not "feel" each other's presence via interatomic forces between unlike atoms. It is also assumed that the number of ways in which the mixture of powder particles can be arranged is not

sufficiently different from unity to give a significant contribution to a configurational entropy of mixing. Thus, a blend of powders which obeys equation 4.1 is called a *mechanical mixture*. It has a free energy that is simply a weighted mean of the components, as illustrated by the point u in Figure 4.1 for a mean composition x.

A solution, on the other hand, describes a mixture of atoms or molecules, *i.e.* the smallest particle in the present context since mixing cannot be achieved on an even finer scale. There will in general be an enthalpy change associated with the change in near neighbour bonds. Because the total number of ways in which the "particles" can arrange is now very large, there will always be a significant contribution from the entropy of mixing. The free energy of the solution is therefore different from that of the mechanical mixture, as illustrated by the point v on Figure 4.1. The difference in the free energy between these two states of the components is the free energy of mixing, the essential term in all thermodynamic models for solutions.

In practice, many solutions do not form instantaneously from a mixture of large particles to an intimate mixture of atoms. Instead, the system must go through a series of transition states involving ever decreasing particle sizes and increasing number densities as the particles are subdivided. One example is the process of mechanical alloying [Benjamin, 1970] in which a mixture of large particles is attrited until the subdivision of particles ultimately leads to the formation of a solution. The violent mixing of two initially immiscible liquids is another example.

4.2.1 Configurational Entropy

We consider a binary system consisting of pure components A and B. In contrast to the conventional approach for solution theory, the equation for the free energy of mixing must contain particle sizes which can be much greater than an atom.

The change in configurational entropy as a consequence of mixing can be obtained using the Boltzmann equation $S = k \ln\{w\}$ where w is the number of configurations and k has its usual meaning.

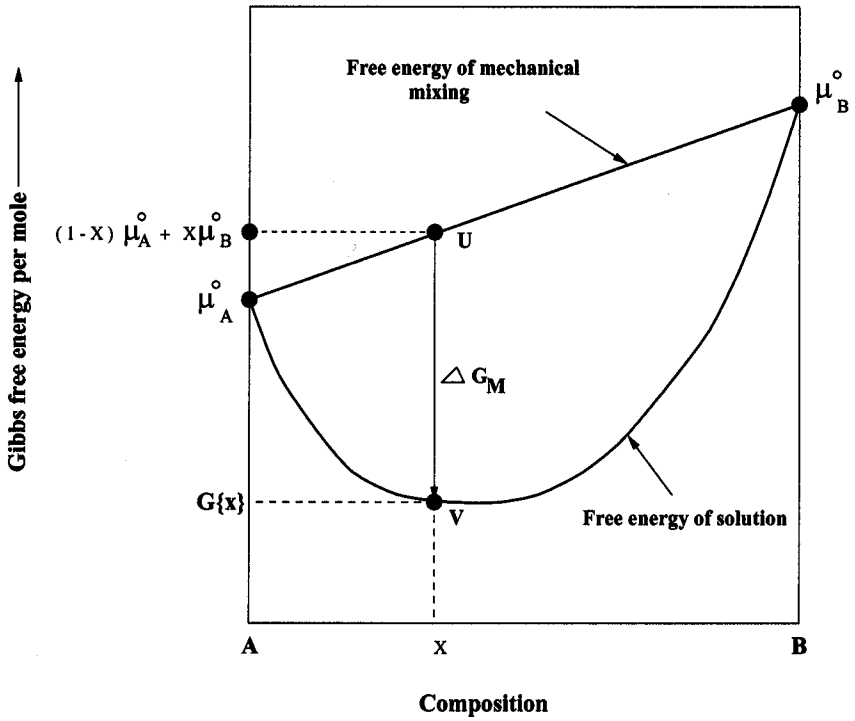

Figure 4.1 : Plot of the molar Gibbs free energy versus composition (mole fraction), both for mechanical mixtures and for solutions. $G\{x\}$ is the free energy of the solution of composition x whereas $(1-x)\mu_A^\circ + x\mu_B^\circ$ is the free energy of the corresponding mechanical mixture of large particles of A and B.

Suppose that there are m_A atoms per powder particle of A, and m_B atoms per particle of B; the powders are then mixed in a proportion which gives an average concentration of B which is the mole fraction x.

There is only one configuration when the heaps of pure powders are separate. When the powders are randomly mixed, the number of possible configurations for a mole of atoms becomes:

$$\frac{(N_a([1-x]/m_B + x/m_A))!}{(N_a[1-x]/m_A)!\ (N_a x/m_B)!} \tag{4.2}$$

where N_a is Avogadro's number. The numerator in equation 4.2 is the total number of particles and the denominator the product of the factorials of the A and B particles respectively. Using Stirling's approximation we obtain the molar entropy of mixing as

$$\begin{aligned}\frac{\Delta S_M}{kN_a} =& \frac{(1-x)m_B + xm_A}{m_A m_B} \ln\left\{N_a \frac{(1-x)m_B + xm_A}{m_A m_B}\right\} \\ &- \frac{1-x}{m_A} \ln\left\{\frac{N_a(1-x)}{m_A}\right\} \\ &- \frac{x}{m_B} \ln\left\{\frac{N_a x}{m_B}\right\}\end{aligned} \tag{4.3}$$

subject to the condition that the number of particles remains integral and non–zero. As a check, it is easy to show that this equation reduces to the familiar

$$\Delta S_M = -kN_a[(1-x)\ln\{1-x\} + x\ln\{x\}]$$

when $m_A = m_B = 1$.

Naturally, the largest reduction in free energy occurs when the particle sizes are atomic. This is illustrated in Figure 4.2 which shows the molar free energy of mixing for a case where the average composition is equiatomic assuming that only configurational entropy contributes to the free energy of mixing. An equiatomic composition maximises configurational entropy. When it is considered that phase changes often occur at appreciable rates when the accompanying reduction in free energy is just $10\,\mathrm{J\,mol^{-1}}$, Figure 4.2 shows that the entropy of mixing cannot be ignored when the particle size is less than a few hundreds of atoms. In commercial practice, powder metallurgically produced particles are typically 100 μ m in size, in which case the entropy of mixing can be neglected entirely, though solution formation must be considered to be advanced when the processing reduces particle dimensions to some 10^2 atoms. These comments must be qualified by the fact that we have neglected any enthalpy changes, which are treated in the next section.

4.2.2 Enthalpy

A major component of the excess enthalpy of mixing comes from the fact that there will in most cases be a change in the energy when new kinds of bonds are created during the formation of a solution.

In the regular solution model, the enthalpy of mixing is obtained by counting the different kinds of near neighbour bonds when the atoms are mixed at random. This information, together with the binding energies, gives the required change in enthalpy on mixing. The binding energy may be defined by considering the change in energy as the distance between a pair of atoms is decreased from infinity to an equilibrium separation. The change in energy during this process is the binding energy, which for a pair of A atoms is written $-2\epsilon_{AA}$. It follows that when $\epsilon_{AA} + \epsilon_{BB} < 2\epsilon_{AB}$, the solution will have a larger

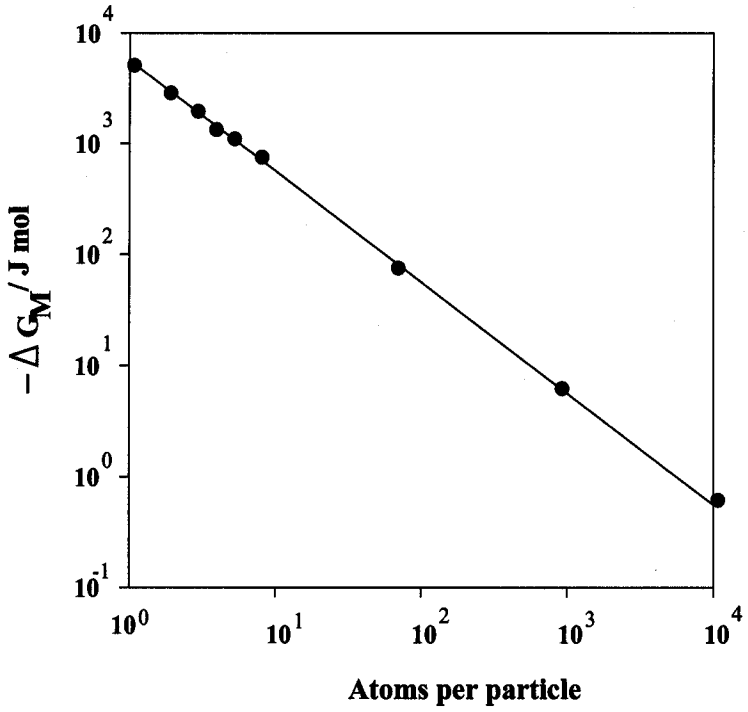

Figure 4.2 : The molar Gibbs free energy of mixing, $\Delta G_M = -T\Delta S_M$, for a binary alloy, as a function of the particle size when all the particles are of uniform size in a mixture whose average composition is equiatomic. $T = 1000\ K$.

than random probability of bonds between unlike atoms. The converse is true when $\epsilon_{AA} + \epsilon_{BB} > 2\epsilon_{AB}$ since atoms then prefer to be neighbours to their own kind.

With the approximation that atoms in a solution are randomly dispersed, the number of A–A bonds in a mole of solution is $zN_a(1-x)^2$, B–B bonds $zN_a x^2$ and A–B bonds $2zN_a(1-x)x$ where z is the co–ordination number. It follows that the molar enthalpy of mixing is given by:

$$\Delta H_M \simeq N_a z(1-x)x\omega \tag{4.4}$$

where

$$\omega = \epsilon_{AA} + \epsilon_{BB} - 2\epsilon_{AB} \tag{4.5}$$

The product $zN_a\omega$ is often called the regular solution parameter which we shall label Ω in subsequent discussion.

This treatment of the enthalpy of mixing has to be adapted for particles which are not monoatomic. For example, only those atoms at the interface between the A and B particles will feel the influence of the

unlike atoms. Assuming cubic particles, each of volume V_i and surface area $S = 6(V_i)^{2/3}$, we have

$$V_i = m_i\Phi_i \qquad \text{and} \qquad V_T = \sum_i \frac{N_a x_i}{m_i} V_i$$

where the subscript i represents the component, Φ_i is the volume per atom, x_i is the mole fraction, V_T is the total volume and N_a is Avogadro's number.

The total *surface area* of n_i isolated particles is $S_T = \sum_i n_i S$ but the total *grain boundary area* when the particles are compacted is half this value. It follows that the grain boundary area per unit volume, S_V is given by

$$S_V = \left(\frac{1}{2}\sum_i n_i 6(m_i\Phi_i)^{\frac{2}{3}}\right) \Big/ \left(\sum_i \frac{N_a x_i}{m_i} V_i\right) \tag{4.6}$$

If it is assumed that the boundary has a thickness 2δ (two monolayers) then the volume fraction of material within the boundary is $S_V \times 2\delta$. The enthalpy of mixing can only be generated within this region where the unlike atoms meet. It follows that

$$\Delta H_M = \Omega \ 2\delta S_V \ x(1-x) \tag{4.7}$$

4.2.3 Interfacial Energy

The role of the interface as discussed in the previous section is simply to identify the number of different atoms that are physically close enough to interact. However, there is a further term which must be taken into account, which does not occur in conventional solution theory. This comes from the inevitable disorder present at the interface, giving a structural component of the interfacial energy as σ per unit area. The chemical component of interfacial energy is already included in the enthalpy of mixing term. The net cost due to the structural component is

$$\Delta H_I = V_m S_V \sigma \tag{4.8}$$

where V_m is the molar volume.

ΔH_I is the *change* in enthalpy content due to interfaces. It is assumed that $S_V \simeq 0$ for very large particles. The process is envisaged as one in which the very large particles are reduced to smaller ones on the route towards the formation of a solution. This is a reasonable description of the mechanical alloying process. It is also assumed in this analysis that σ is identical for interfaces between $A - A$ particles, $B - B$ particles and $A - B$ particles.

4.3 Results & Discussions

The modelling of atomic solutions is well–established but Figure 4.3 nevertheless presents the results for particles which are one–atom in size. This is to illustrate the magnitude of the free energy changes involved for comparison against later results, and to highlight the fact that the energy of mixing is zero for the pure components. There are three cases illustrated corresponding to solutions in which like–atoms tend to cluster ($\Omega > 0$), those in which they tend to order ($\Omega < 0$) and lastly, the ideal solution ($\Omega = 0$). The calculations are for 1000 K, the temperature dependence appearing only via the $(-T\Delta S_M)$ term. Finally, interfacial energy does not feature in these plots because the solution is atomic and therefore fully coherent.

Difficulties arise when considering the evolution of a solution from a mixture of large particles to mixtures with ever decreasing sizes. This scenario precisely describes mechanical alloying, where the component powders are repeatedly fractured and deformed until an atomic solid solution is formed. The refinement of particle size leads to an increase in the amount of interface per unit volume (S_V). If the interface energy is constant then the cost indicated by equation 4.8 must eventually overwhelm any advantage from the entropy or enthalpy of mixing. The inescapable conclusion is that mechanical alloying cannot occur unless there is a gain in coherency, *i.e.* a reduction in the interfacial energy as atomic dimensions are approached. This is not surprising in hindsight, since the process envisaged is the opposite of the normal precipitation and growth event in which a small particle begins coherently and loses coherency as it grows.

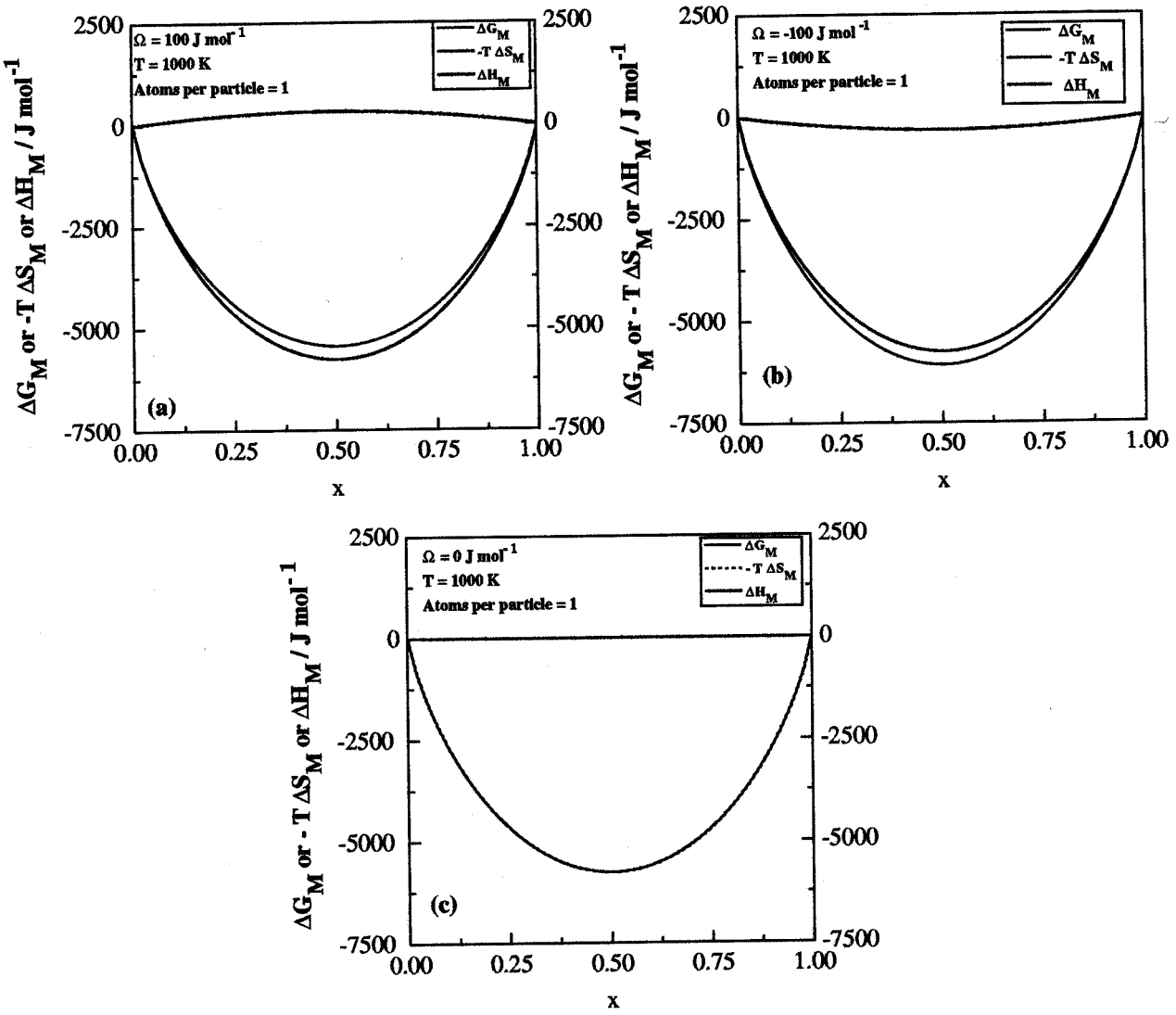

Figure 4.3 : Plots of the free energy, entropy and enthalpy of mixing in a binary system at 1000 K. (a) $\Omega \doteq 100\,\mathrm{J\,mol^{-1}}$, (b) $\Omega = -100\ \mathrm{J\,mol^{-1}}$ and (c) $\Omega = 0\,\mathrm{J\,mol^{-1}}$.

Since the change in the interfacial energy with particle size is not known, we have assumed that σ remains constant until a size of 10^7 atoms is reached. Beyond that the energy decreases linearly to zero when the particle size is one atom:

$$\begin{aligned} \sigma &= 0.5\ \mathrm{J\,m^{-2}} && m > 10^7 \text{atoms per particle} \\ \sigma &= 0.5\left(1 - \frac{10^7 - m}{10^7}\right)\ \mathrm{J\,m^{-2}} && m \leq 10^7 \text{atoms per particle} \end{aligned} \tag{4.9}$$

The results as a function of particle size are illustrated in Figure 4.4. A typical powder particle at the

beginning of the mechanical alloying process might be some 10^{17} atoms in size but the calculations begin at 10^9 in order to see significant effects. The free energy change as the particle size is reduced to 10^9 atoms is at first mainly a contribution from the increase in the structural component of interfacial energy. The net free energy change remains positive until contributions from the enthalpy and entropy of mixing begin to become significant and when the interfacial energy term begins to decrease below 10^7 atoms due to a gain in coherency.

The energy changes are all very much smaller than associated with the direct formation of an atomic solution (cf. Fig. 4.3). This is because the contribution from configurational entropy is small until the particle size reaches about 100 atoms, and that from enthalpy is small because only those atoms at unlike–particle interfaces can interact. The details also depend on the nature of the solution, the sign and size of Ω. For the case where like atoms tend to cluster ($\Omega > 0$), a pair of minima develop, in the A and B rich regions, in the ΔG_M function as the particle size is reduced. Solution formation is not favoured when the concentrations of the two components are about equal. The opposite is true when $\Omega \geq 0$ and mixing is favoured at all compositions once the particle size becomes small enough for coherency to set in.

Figure 4.5 shows how there is a barrier to the formation of a solid solution during the mechanical alloying process. The barrier occurs for all the examples illustrated because of the incorporation of interfacial energy in the analysis. This dominates in the early stages as the particle size is reduced, until a size is achieved below which coherency begins to set in. It is important to note that when $\Omega > 0$, there is an energy barrier even in the absence of interfacial energy since the alloying forces unlike atoms to mix leading to an increase in enthalpy. Naturally, this latter effect is minimised at low or high concentrations which are A–rich or B–rich respectively. Figure 4.5a reveals the possibility that there is an energy well in the curve of energy versus particle size, suggesting a possible metastable state making the alloying process particularly difficult. However, the general result is that there is a certain particle size below which mechanical alloying proceeds rapidly with a reduction in free energy.

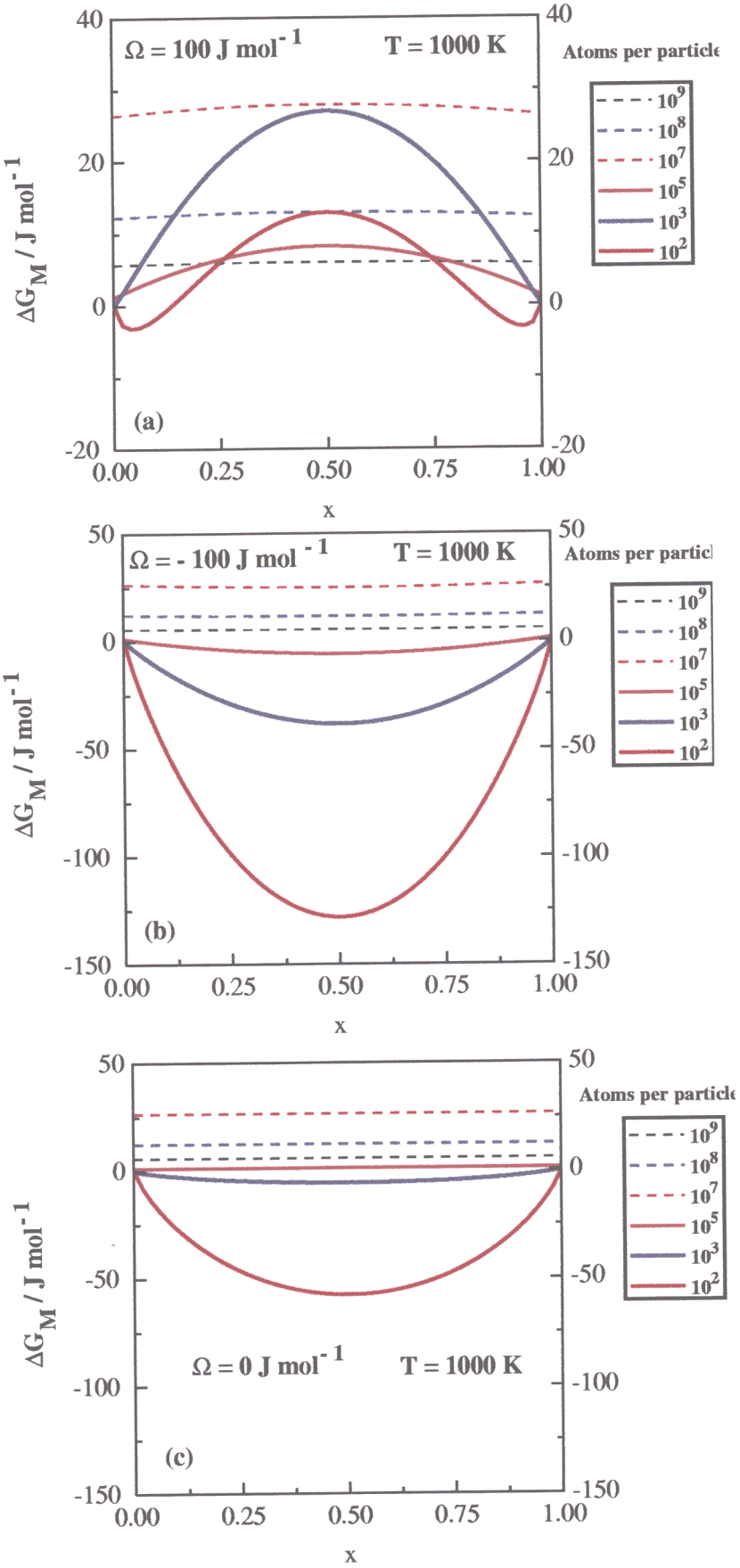

Figure 4.4 : Plots of the free energy, entropy and enthalpy of mixing in a binary system at 1000 K, as a function of the number of atoms in the particles. (a) $\Omega = 100\,\mathrm{J\,mol^{-1}}$, (b) $\Omega = -100\,\mathrm{J\,mol^{-1}}$ and (c) $\Omega = 0\,\mathrm{J\,mol^{-1}}$.

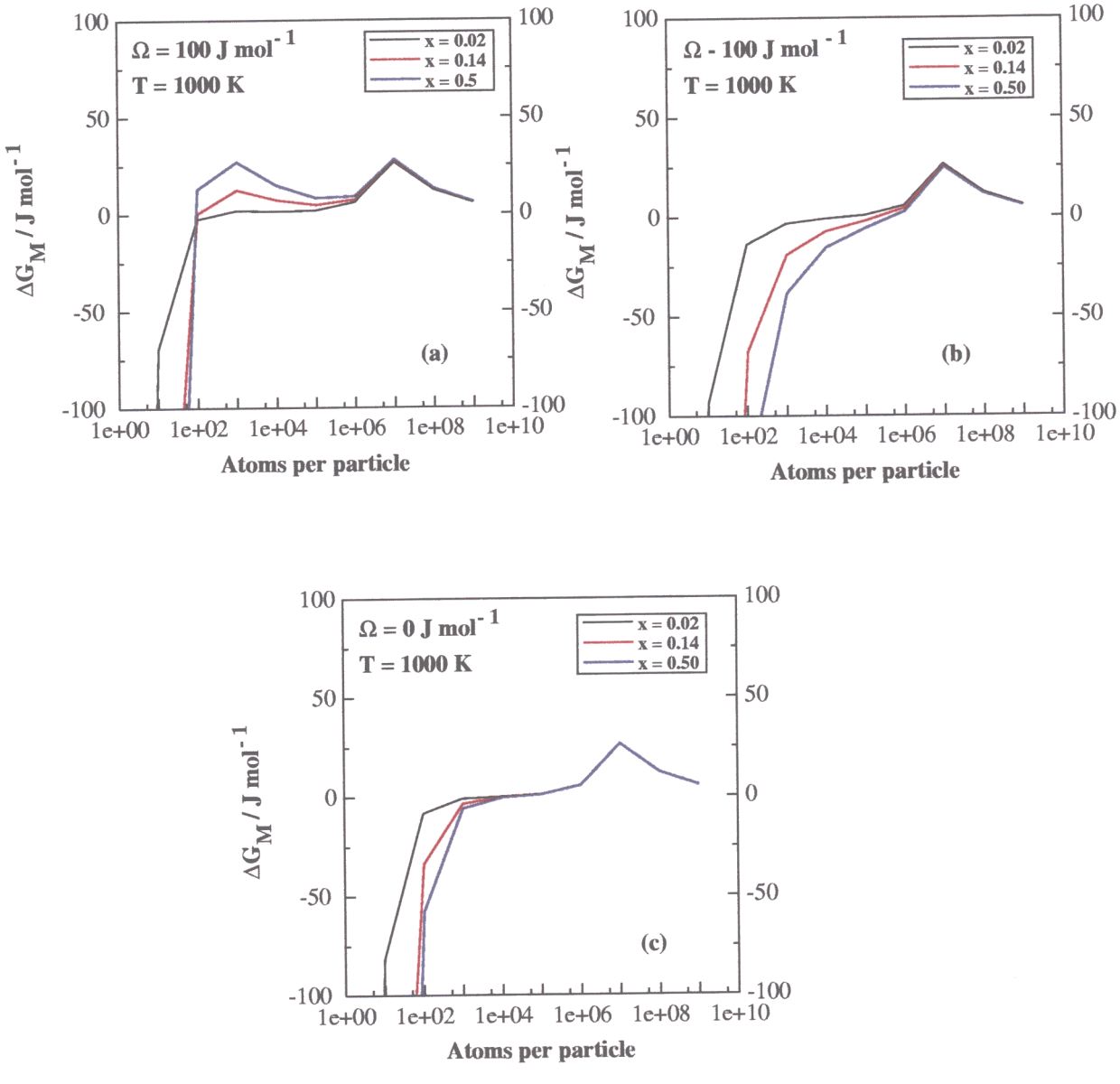

Figure 4.5 : Free energy versus the number of atoms per particle as a function of the concentration, at $T = 1000K$. (a) $\Omega = 100\,\mathrm{J\,mol^{-1}}$, (b) $\Omega = -100\,\mathrm{J\,mol^{-1}}$ and (c) $\Omega = 0\,\mathrm{J\,mol^{-1}}$. The jagged appearance is because the calculated points have been connected with straight segments. The true variation is expected to be smooth.

Finally, the entropy term $-T\Delta S$ ensures that the effect of increasing the temperature is, for all finite x, to make the alloying process easier (Figure 4.6).

4.4 Conclusions

A model has been developed to deal with a situation in which a solution is created by continuously refining a mixture of powder particles of the pure components. This process of solution formation is a good representation of the mechanical alloying process.

It is predicted that mechanical alloying is not possible unless initially incoherent interfaces approach coherency and eventually disappear as true solution is approached. The inclusion of interfacial energy also predicts the existence of a barrier to the evolution of the solution. For cases where like atoms tend to cluster, it is possible in principle to obtain a metastable state before solution formation is completed.

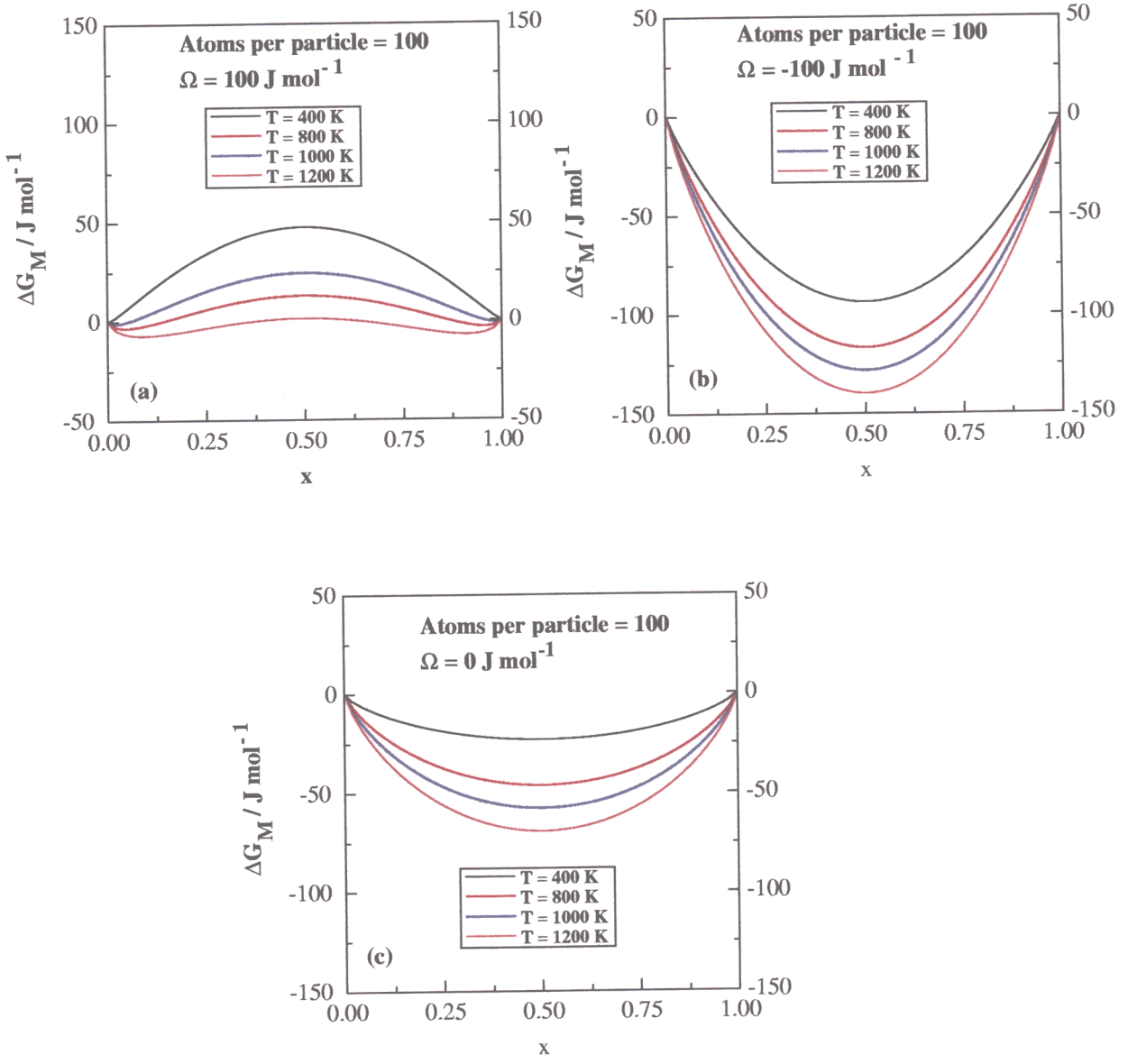

Figure 4.6 : Effect of temperature on the free energy of mixing for a particle size containing 100 atoms per particle (a) $\Omega = 100\,\mathrm{J\,mol^{-1}}$, (b) $\Omega = -100\,\mathrm{J\,mol^{-1}}$ and (c) $\Omega = 0\,\mathrm{J\,mol^{-1}}$.

CHAPTER FIVE

Neural Network Models for the Tensile Properties Of Mechanically Alloyed ODS Iron–Alloys

5.1 Introduction

Very limited information is available in the published literature on the mechanical properties of mechanically alloyed materials. The experimental data which are available do not seem to have been adequately interpreted in terms of the numerous variables known to affect the mechanical properties. The technique of mechanical alloying is unconventional and this is coupled with the complex microstructures. The objective of this work was to investigate whether an artificial neural network [Rumelhart *et al.*, 1986] can be trained to predict the yield strength, ultimate tensile strength and elongation of the mecanically alloyed ferritic stainless steels as a non-linear function of the important processing and service variables.

This chapter is presented first with an introduction to the neural network technique, followed by a description of how the optimum model was obtained in each case. The final section deals with the use of the models to ensure as far as is possible that the percieved relationships are metallurgically significant.

5.2 The Neural Network

Most people 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 $y = \sum_j w_j x_j + \theta$.

It is well understood 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 chosen being the hyperbolic tangent because of its flexibility. The exact shape of the hyperbolic tangent can be varied by altering the weights (Figure 5.1a). Further degrees of non-linearity can be introduced by combining several of these hyperbolic tangents (Figure 5.1b), so that the neural network method is able to capture almost arbitrarily non-linear relationships. It is well known that the effect of chromium on the microstructure of steels is quite different at large concentrations than in dilute alloys. Ordinary regression analysis cannot cope with such changes in the form of relationships.

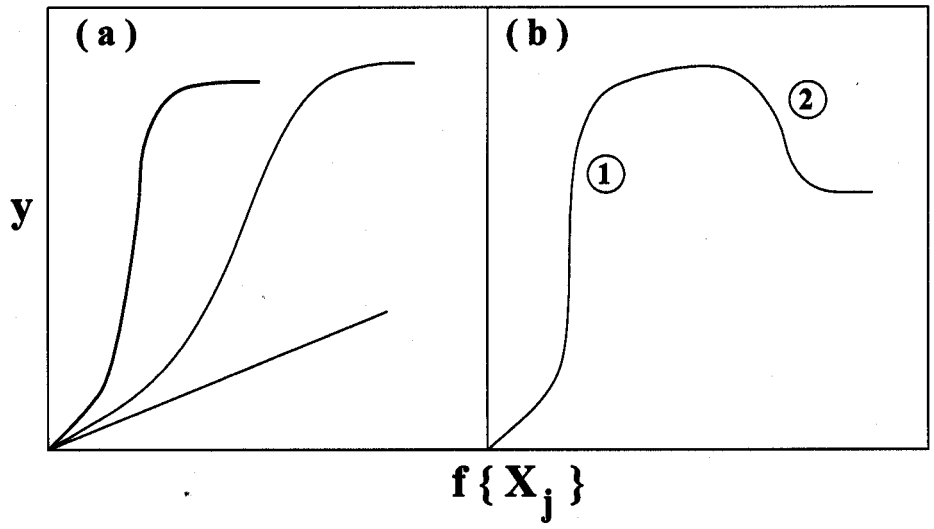

Figure 5.1 : (a) Three different hyperbolic tangent functions - the 'strength' of each depends on the weight (b) A combination of two hyperbolic tangents to produce a more complex model.

A neural network is 'trained' on 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 with blind data modelling is that of 'overfitting', in which spurious details and noise in the training data are overfitted by the model (Figure 5.2). This gives rise to solutions that generalise poorly. Mackay (1992, 1994, 1995, 1997) has developed a Bayesian framework for neural networks in which the appropriate model complexity is inferred from the data. The Bayesian theory is a subject in its own right and beyond the scope of the present work.

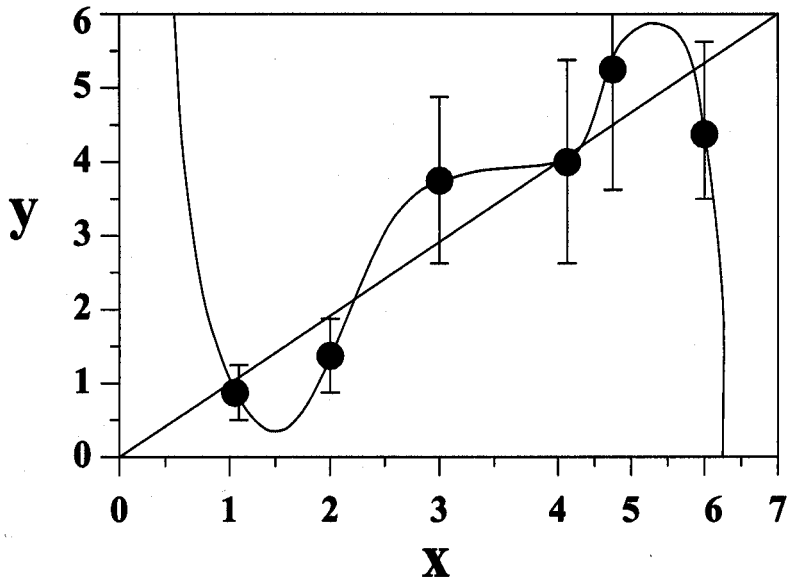

Figure 5.2 : A complicated model may fit the data, but in this case, a linear relationship may be all that is justified by the noise in the data.

However, the Bayesian framework for neural networks has two important advantages. First, the significance of the input variables with respect to explaining the variation in the output is automatically quantified. Consequently the significance perceived by the model of each input variable can be compared against metallurgical theory. Second, the network's predictions are accompanied by error bars which depend on the specific position in input space. These quantify the model's certainty about its predictions. The error bars are therefore not fixed as in conventional regression analysis.

The neural network method has recently been applied to many materials problems and examples include: the impact toughness of C–Mn steel arc welds by Bhadeshia *et al.* (1995); an analysis of the strength of Ni-base superalloys by Jones *et al.* (1995); austenite formation in steels by Garvard *et al.* (1996); yield and ultimate tensile strength of steel welds by Cool *et al.* (1997); fatigue crack growth rate in nickel base superalloys by Fujii *et al.* (1996); mechanical properties in the HAZ of power plant steels by Cool *et al.* (1997); prediction of martensite-start temperature by Vermeulen *et al.* (1996); prediction of the continuous cooling transformation diagram of some selected steels by Vermeulen *et al.* (1997), and prediction of the measured temperature after the last finishing stand in hot rolling, by Vermeulen *et al.* (1997).

5.3 The Analysis

Both the input and output variables were first normalized with the range $+0.5$ to -0.5. The normalisation is obtained through a procedure which is expressed quantitatively as :

$$x_N = \frac{x - x_{min}}{x_{max} - x_{min}} - 0.5 \tag{5.1}$$

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

Figure 5.3 shows a typical network. Each network consisted 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:

$$h_i = \tanh\left(\sum_j w_{ij}^{(1)} x_j + \theta_i^{(1)}\right) \tag{5.2}$$

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:

$$y = \sum_i w_{ij}^{(2)} h_i + \theta^{(2)} \tag{5.3}$$

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.

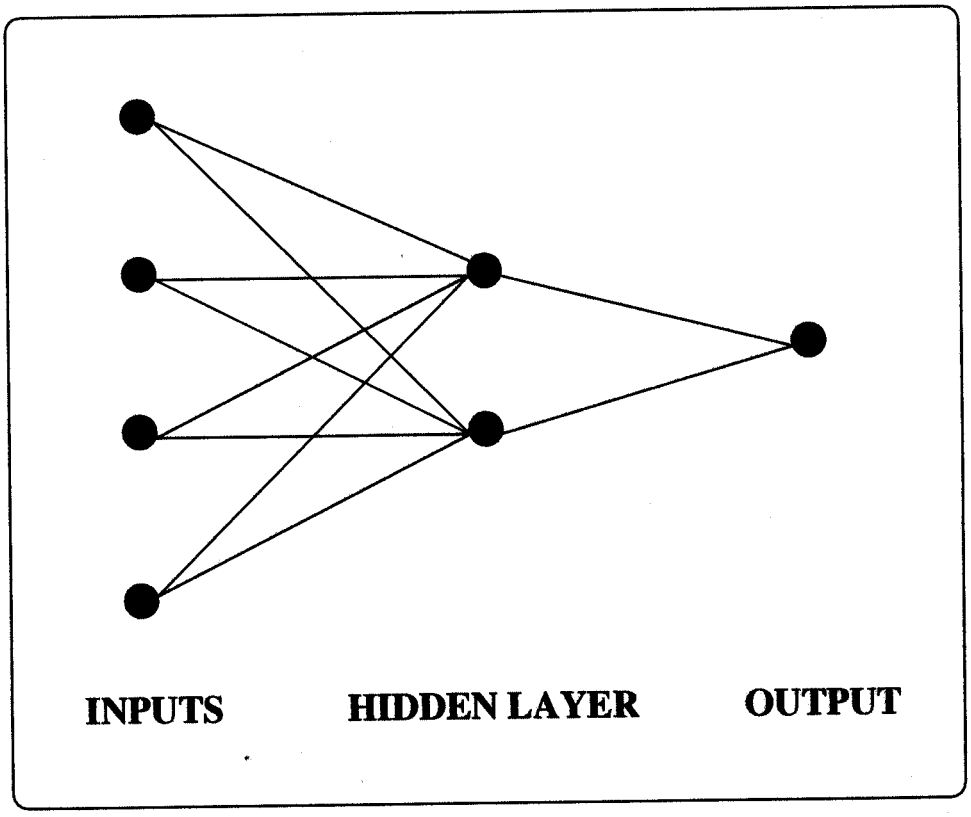

Figure 5.3 : Typical network used in the analysis.

5.4 Training and Optimisation

The 'training' is the time consuming stage in the production of a functioning neural network. A neural network as defined in equations (3.2) and (3.3) is trained using a data set $D = \{x^m, t^m\}$ by adjusting **w** so as to minimize an error function, e.g.,

$$E_D(w) = \frac{1}{2}\sum_m (y(x^m; w) - t^m)^2 \tag{5.4}$$

This objective function is a sum of terms, one for each input-target pair $\{x, t\}$, measuring how close the output is to the target **t** (Mackay, 1995).

The minimization is based on the repeated evaluation of the gradient of E_D using 'backpropagation'[Rumelhart *et al.*, 1986]. The backpropagation algorithm computes for each input-output pair m the gradient of $\frac{1}{2}\left(y(x^m; w) - t^m\right)^2$ by following the 'forward pass' of equations (3.2–3.3) by a 'backward pass', in which information about the errors $\left(y(x^m; w) - t^m\right)$ propagates back through the network by the chain rule. Often, regularization is included, modifying the objective function to

$$M(w) = \beta E_D + \alpha E_W \tag{5.5}$$

where, for example, $E_W = \frac{1}{2}\sum w_i^2$. This additional term favours small values of **w** and thus encourages the model to find simpler solutions with less tendency to 'overfit' noise in the data. The control parameters, α and β define the assumed Gaussian noise level $\sigma_\nu^2 = \frac{1}{\beta}$ and the assumed weight variances $\sigma_w^2 = \frac{1}{\alpha}$. σ_ν is the noise level inferred by the model. The parameter α has the effect of encouraging the weights to decay. Therefore, a high value of σ_w implies that the input parameter concerned explains a relatively large amount of the variation in the output. Thus, σ_w is regarded as a good expression of the significance of each input though not of the sensitivity of the output to that input. The values of the hyperparameters are inferred from the data using the Bayesian method by [Mackay, 1992].

The complexity of the model is controlled by the number of hidden units and the values of the regularization 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, σ_ν, 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 (1992, 1993, 1994, 1997) has made a detailed study of this problem and defined a quantity (the 'evidence') which comments on the probability of a model. In circumstances where two models give similar results for the known dataset, 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 the noise level . The number of hidden units is set by examining performance on test data. A combination of Bayesian and pragmatic statistical techniques are therefore used to control the complexity of the model. A further procedure used to avoid the overfitting problem was to randomly divide the experimental data into two equal sets, the *training* and *test* datasets. The models are developed using just the training data. The unseen test data are then used to assess how the model generalizes. A good model would produce similar levels of errors 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 chosen using this procedure, it can be retrained using all the data with a small but significant reduction in the error.

The test error (sum squared error, E_D) is a reflection of the ability of the model to predict the target values in the test data. It is popular to use the test error as the default performance measure whereby the model with the lowest test error is reckoned to be the best [Mackay, 1994]. In many applications there will be an opportunity not to simply make a scalar prediction, but rather to make a prediction with error bars, or maybe an even more complicated predictive performance. 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 [Mackay, 1997]. Assuming that for each example m the model gives a prediction with error, $(y^{(m)}, \sigma^{(m)2})$, the log predictive error (LPE) is [Mackay, 1997]:

$$LPE = \sum_m \left[\frac{1}{2} \left(t^{(m)} - y^{(m)} \right)^2 / \sigma_y^{(m)^2} + \log(\sqrt{2\pi}\sigma_y^{(m)}) \right] \tag{5.6}$$

5.5 Committee Model

It is common practice in the application of neural networks to train many different candidate networks from the same data by varying either the number of hidden units or starting value of σ_w, and then to select the best on the basis of the test error or log predictive error, and to keep only this network and discard the rest. There are two disadvantages with such an approach. First, all of the effort involved in training the remaining networks is wasted. Second, the generalization performance on the validation set has a random component due to the noise on the data, and so the network which had best performance on the validation set might not be the one with the best performance on new test data.

These drawbacks can be reduced by combining the networks together to form a *committee* [Perrone and Cooper, 1993; Perrone, 1994; Mackay, 1994]. The importance of such an approach is that it can lead to significant improvements in the predictions on new data, while involving little additional computational effort. In fact the performance of a commitee can be better than the performance of the best single network used in isolation, although there is an optimum size to the committee.

Suppose we have a set of L trained network models y_i where $i = 1, ..., L$. The commitee output is taking to be the average of the outputs of the L networks which comprise the committee. Thus, the commitee prediction, $\overline{y}$ is written as:

$$\overline{y} = \frac{1}{L}\sum_{i}^{L} y_i \tag{5.7}$$

The optimum size of the committee is determined from the validation error of the committee's predictions using the test dataset. The test error of the predictions made by a committee is calculated by replacing the y in equation (3.5) with $\overline{y}$.

5.6 The Database

The database used for the analysis was compiled from published literature. There are 232 data points in the database. The dataset is restricted because at the moment, published information on the mechanical properties of mechanically alloyed materials is very limited. Moreover, discrepances in the experimental conditions and inadequate reporting in some cases necessitated the exclusion of some published data.

Table 5.1 shows the chemical compositions of some commercial MA-ODS steels which are represented in the data used for the analysis [Whittenberger, 1981; Singer and Gessinger, 1994; Zakine *et al.*, 1993; Dubiel *et al.*, 1994; Hendrix and Vandermeulen, 1982; Alamo *et al.*, 1990, 1992; Regle, 1994 etc.].

Table 5.1 : Chemical composition (wt.%) of some commercial MA-ODS steels.

Steel	Cr	Al	Mo	Ti	Y_2O_3	Fe
MA956	20.0	4.5	-	0.5	0.5	Balance
MA957	14.0	-	0.3	1.0	0.27	Balance
DY (DT2203Y05)	13.0	-	1.5	2.2	0.5	Balance
DT (DT2906)	13.0	-	1.5	2.9	-	Balance

5.7 The yield strength model

The technique was applied to the variables listed in Table 5.2 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 (Cr, Al, Ti, Mo) 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 ageing treatment [Hendrix and Vandermeulen, 1982]. 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 one which is coarse and columnar. Cold work is naturally expected to increase the yield strength; the dataset included a variety of methods of cold deformation including rolling and swaging. The yield strength of body-centred cubic metals is particularly sensitive to temperatrure because of the large Peierls barriers to dislocation motion. A further temperature dependence comes from the possibility of the climb of dislocations over dispersoids. There may, in unrecystallised alloys be an additional effect due to the onset of dynamic recrystallisation.

The plot of σ_{ν}, as a function of the complexity of the models is shown in Figure 5.4 . Note that a number of values are presented for each hidden unit because the training process was started off using different randomly chosen seeds which determine the starting values of the weights. The test error and log predictable error versus hidden units are shown in Fig. 3.5.

The numerical data for the top ten models ranked by their test error are shown in Table 5.3. Table 4 shows 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 Figure 5.6 plotted to the same scale to show the usual reduction in test error when an appropriate committee is formed. The committee consisting of five top ranking

Table 5.2 : The variables used in the analysis of the 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
Recrystallization time, s	0-120	28.44	33.38
Ageing temperature, °C	20-800	163.3	303
Ageing 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}	0.00000003-0.03	0.000989	0.0025
Yield strength, MPa	63-1600	497	388

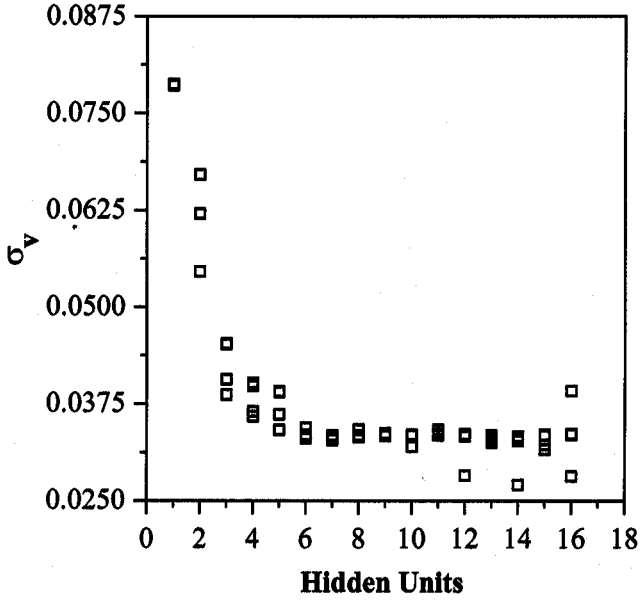

Figure 5.4 : Variation in σ_v as a function of hidden units.

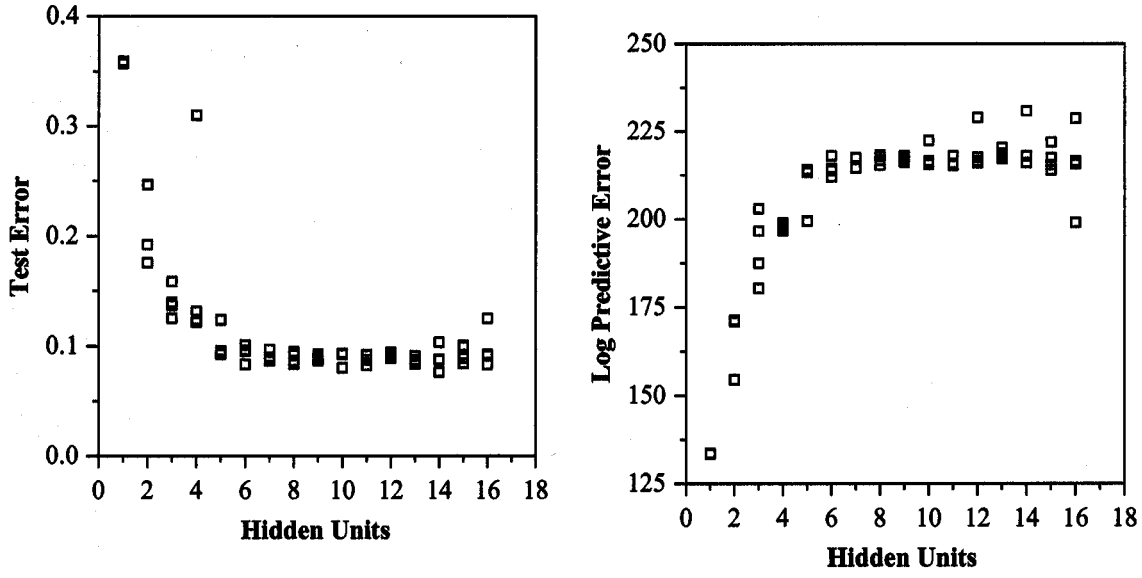

Figure 5.5 : Variation in test error and log predictive error as functions of the number of hidden units. Note that a larger log predictive error stands for a better model.

models has the least test error and was used for the study of the yield strength presented later.

The plots of the predicted values versus experimental values of the training and test dataset for the single best model and the commitee are shown in Figures 3.7 and 3.8 respectively. Figure 3.9 shows the plot for the committee after retraining using the whole dataset.

The perceived significance of each of the input variables by the various models contained in the committee is shown by σ_w in Figure 5.10. 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 the each of the inputs. As expected, the yield strength correlates strongly with temperature.

Table 5.3 : Ranking by test error of the ten best models of the 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 5.4 : Test errors of the committees. Notice that the test error of the best committee is less than that of the 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

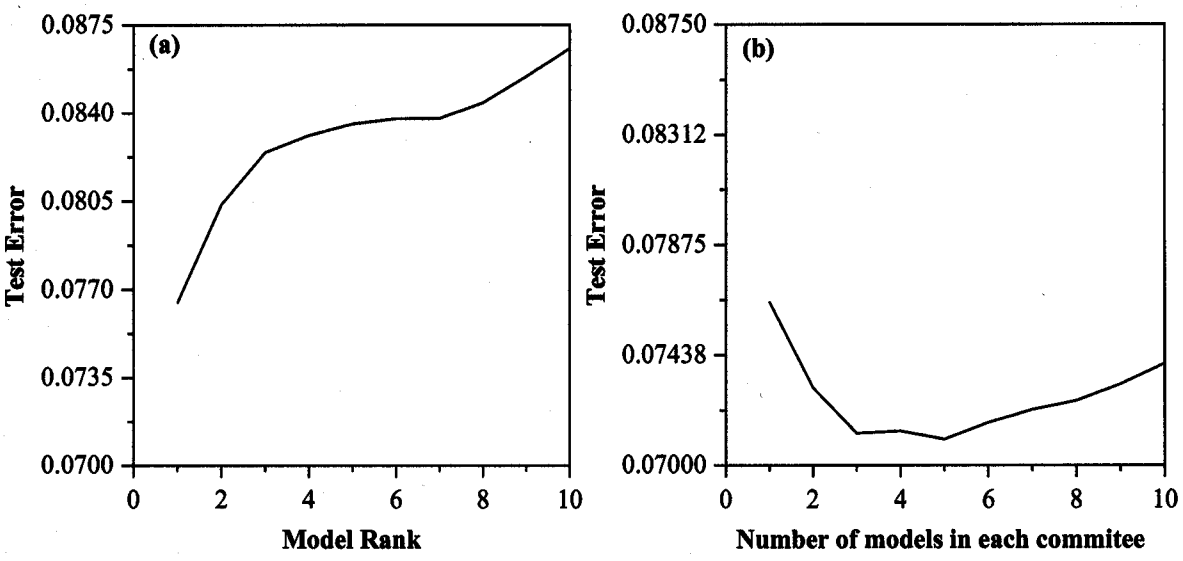

Figure 5.6 : (a) Test errors of the top ten of yield strength models; and (b) corresponding test errors for the commitee models.

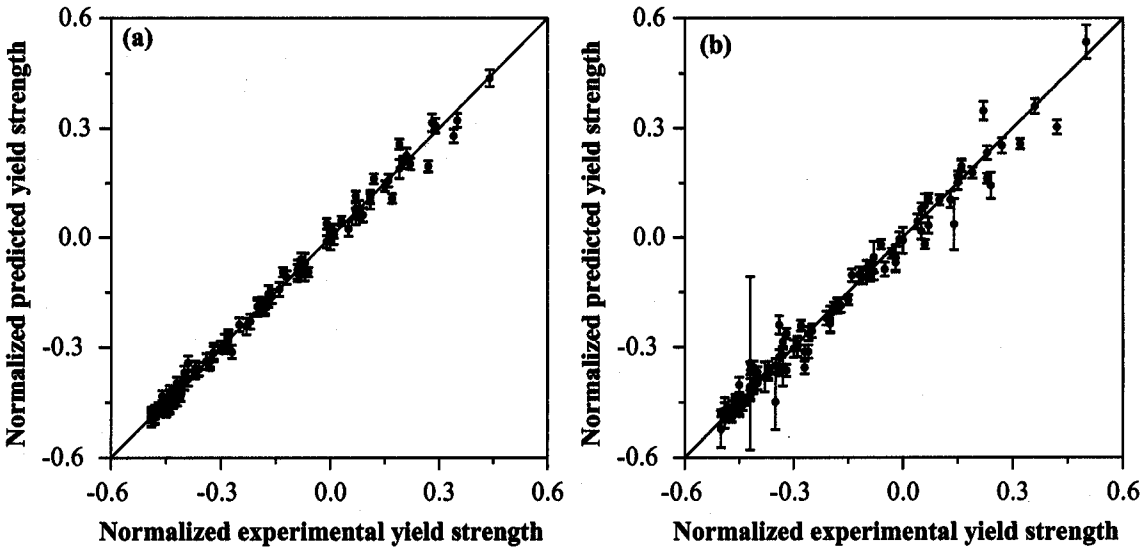

Figure 5.7 : Predicted versus experimental results using single best model. (a) training dataset, (b) test dataset.

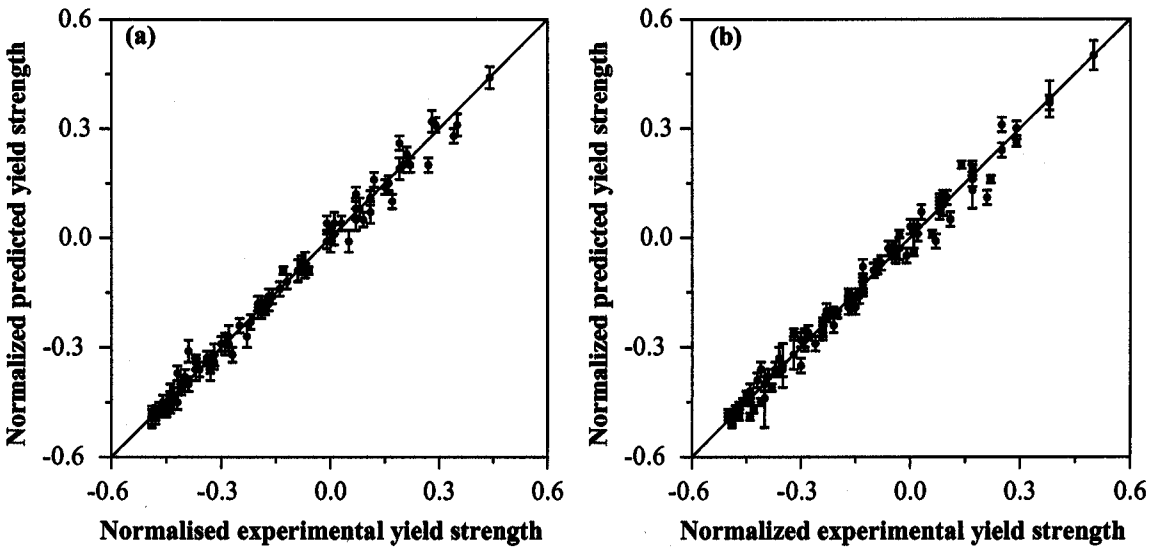

Figure 5.8 : Predicted versus experimental results using the optimum committee. (a) training dataset, (b) test dataset.

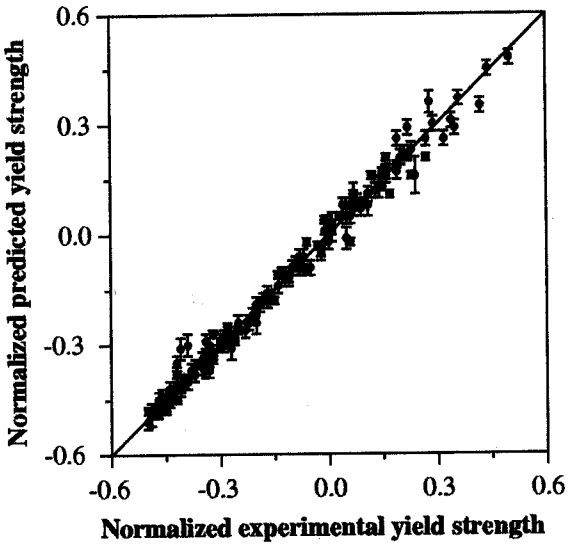

Figure 5.9 : Predicted versus experimental results for the whole dataset after retraining, using the optimum committee.

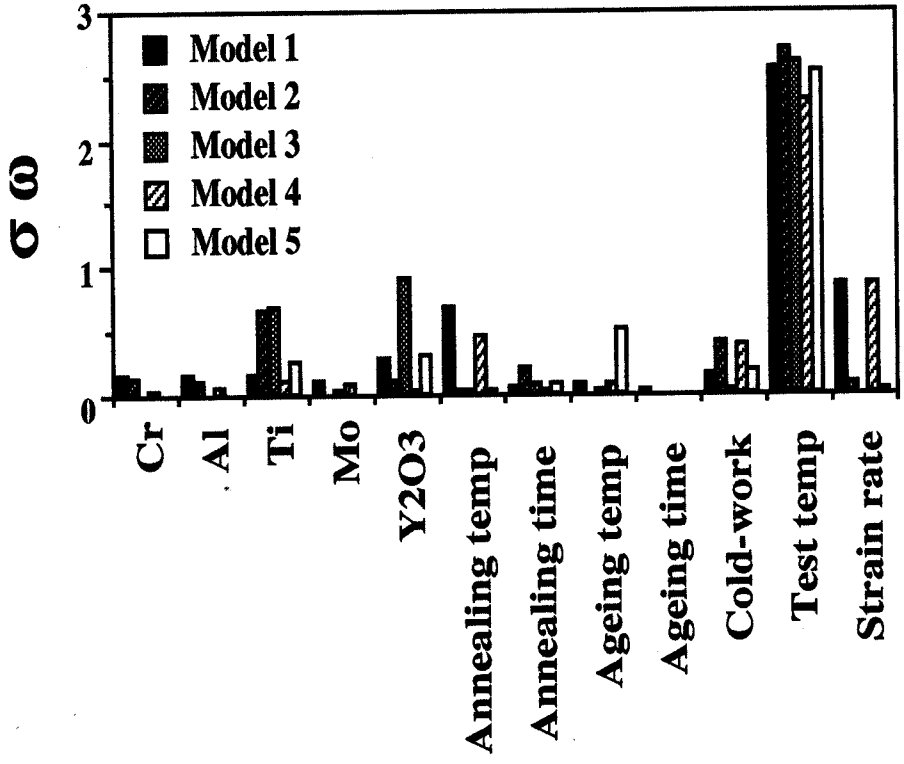

Figure 5.10 : The model-percieved significance of the input parameters for the committee model trained on all data. The σ_w value for all the members of the committee are presented for each of the variables.

5.8 The ultimate tensile strength model

The neural network technique was applied to the ultimate strength data shown in Table 5.5. The variables are identical to those used for the yield strength analysis. It would ideally be interesting to include the strain hardening coefficient since this determines the plastic instability which defines the ultimate strength. However, such data were not found at all in the published literature. There were 12 input variables and one output variable, the ultimate tensile strength. A total of 232 data were used. The data were divided equally and randomly into a 'test' dataset and a 'training' dataset. The 'training' dataset was used to train the model and the ability of the model to generalise was examined by checking its performance on the unseen 'test' data.

Table 5.5 : The variables used in the analysis of the ultimate tensile strength.

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
Recrystallization temperature, °C	20-1330	684	594
Recrystallization time, s	0-120	27.56	33.24
Ageing temperature, °C	20-800	174.7	311.7
Ageing 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}	0.00000003-0.03	0.0011	0.0025
UTS, MPa	70.7-1680	575.3	407.3

Figure 5.11 shows the plot of σ_ν versus hidden units. As expected, the inferred noise level decreases monotonically as the number of hidden units increases.

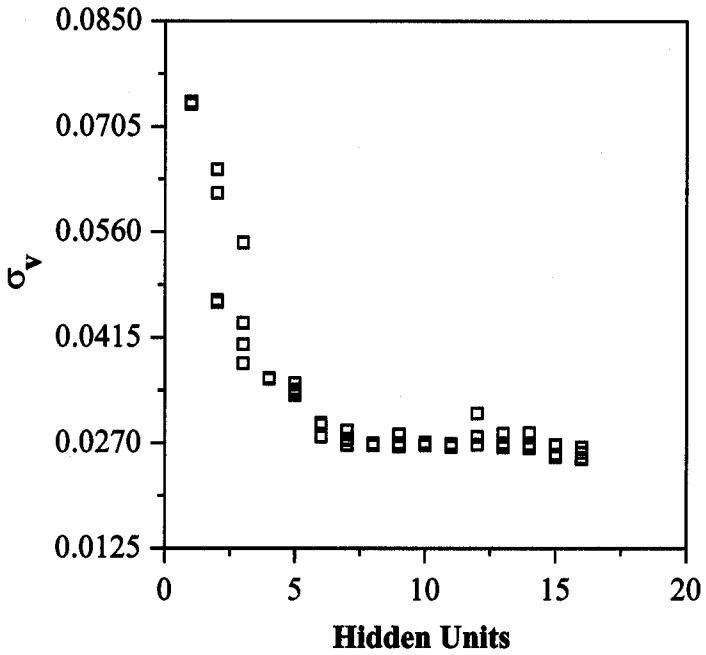

Figure 5.11 : Variation in σ_ν as a function of hidden units. Several values are presented for each set of hidden units because the training for each network was started with a variety of random seeds.

Figure 5.12 show the variation of test error and log predictive error as functions of the number of hidden units. The calculated test error goes through a minimum at 16 hidden units and the log predictive error also goes through a maximum at the same number of hidden units. This would have been the optimum model if a single model were to be 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 best. The models were ranked using their test error values as shown in Table 5.6.

The optimum number of models in the committee was determined from the calculated validation errors of the different possible committees. Figure 5.13 shows the variation of the test error of the best models

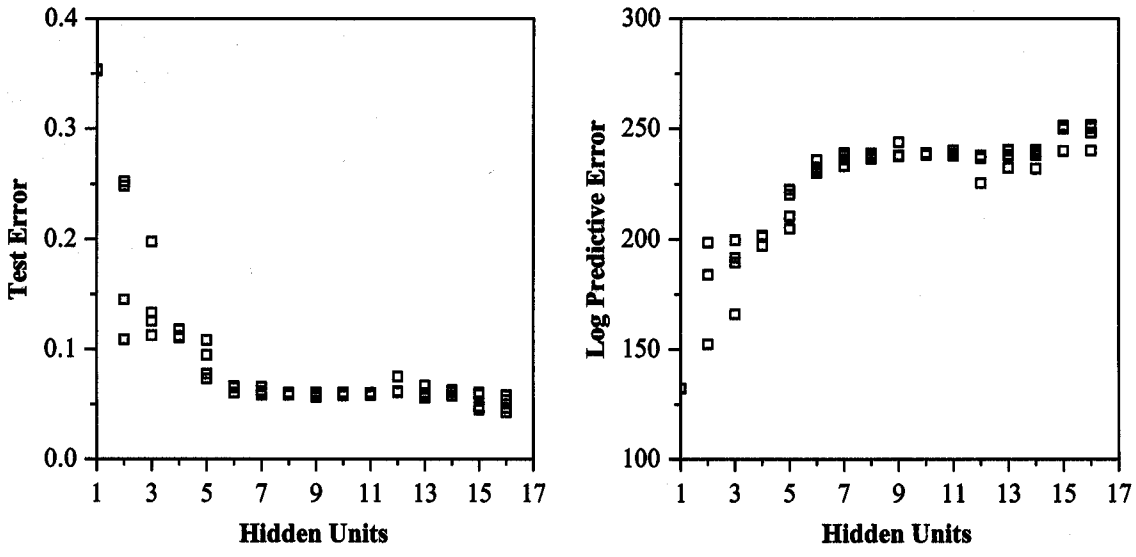

Figure 5.12 : Variation in test error and in log predictive error as functions of the number of hidden units.

Table 5.6 : Ranking by test error for the four best models of ultimate strength .

Ranking	Hidden units	Seed	Test energy	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

as a function of their position on the ranking table and the test error of the commitees as a function of the number of models. It is evident that forming a commitee reduces the test error, and hence improves predictions.

As shown in Fig. 11 the commitee that is made up of the top three models shows the least test error and was used for the analysis. The agreement between the predicted and experimental values for the training and test datasets are shown in Figures 3.14 and 3.15 for the single best model and the committee respectively.

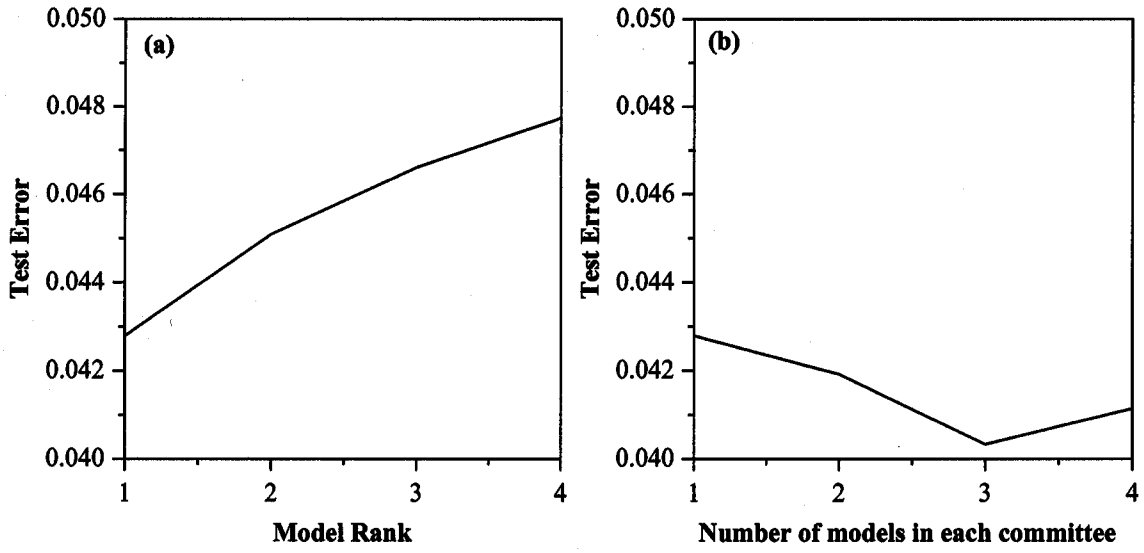

Figure 5.13 : (a) Test error of the four best ultimate tensile strength models and, (b) corresponding test errors for commitees of models.

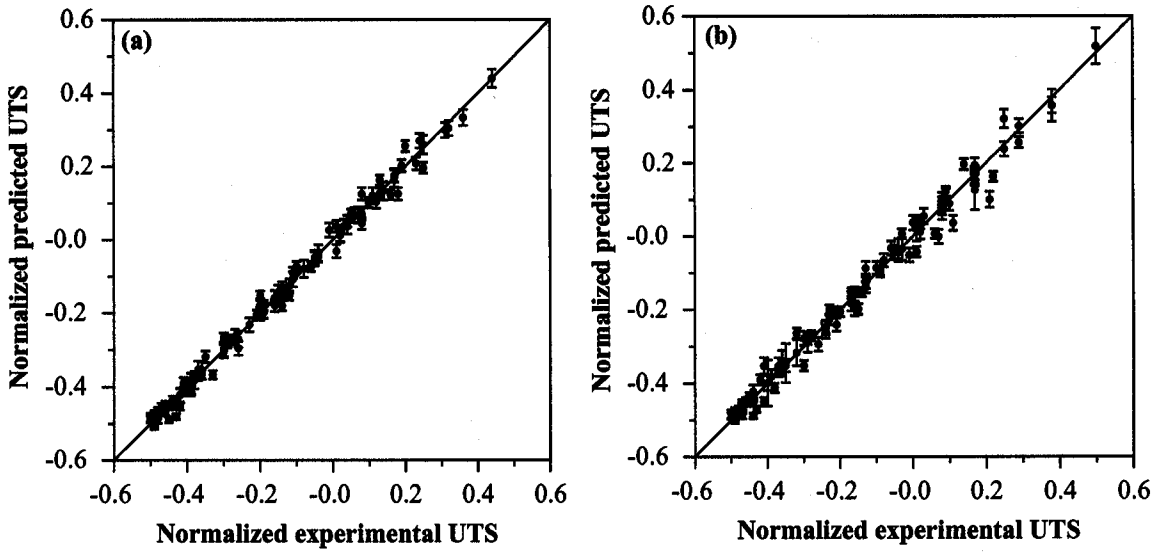

Figure 5.14 : Predicted versus experimental results using the single best model. (a) training dataset (b) test dataset.

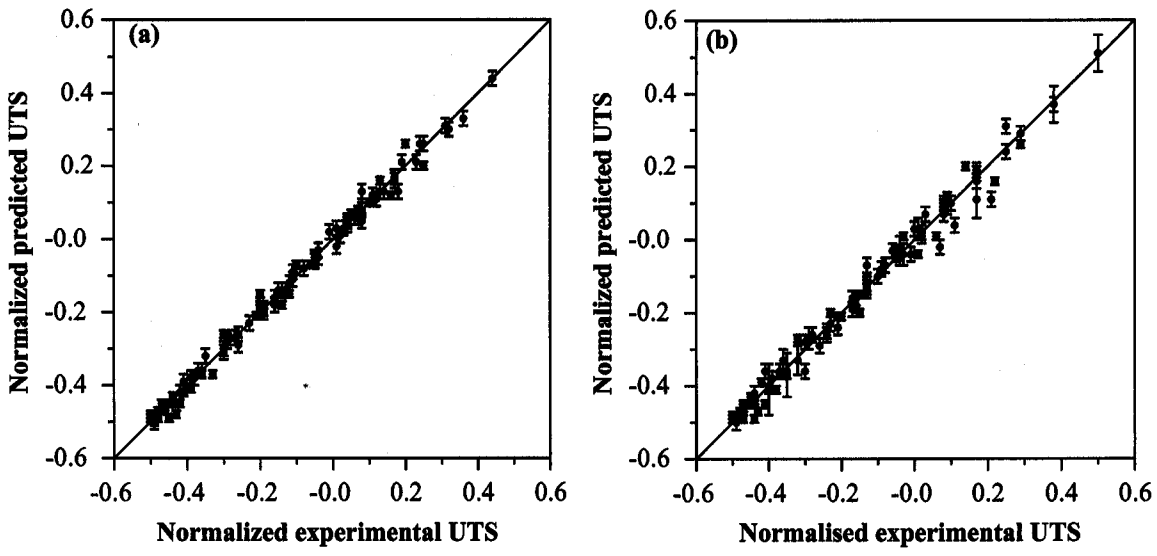

Figure 5.15 : Predicted versus experimental results using the optimum committee. (a) training dataset (b) test dataset.

The committee models were then retrained on the whole dataset beginning with the weights determined from the previous training exercise. Figure 5.16 shows the plot of the predicted values versus experimental values of the whole dataset 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.

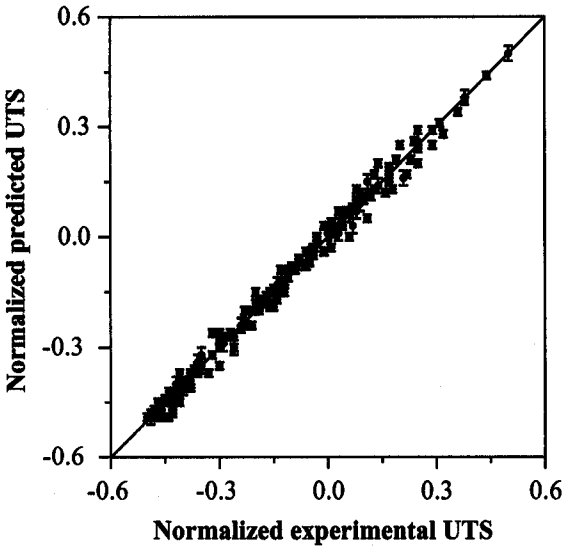

Figure 5.16: Predicted versus experimental results for the whole dataset after the retraining using the committee.

The perceived significance of each of the input variables by the various models contained in the committee is shown by σ_w in Figure 5.17 . The test temperature is shown to have the largest σ_w for all the three models in the committee. This shows the models have recognised a correct pattern because temperature is mostly varied than any other input in the database. Moreover, temperature is known to affect strength very significantly.

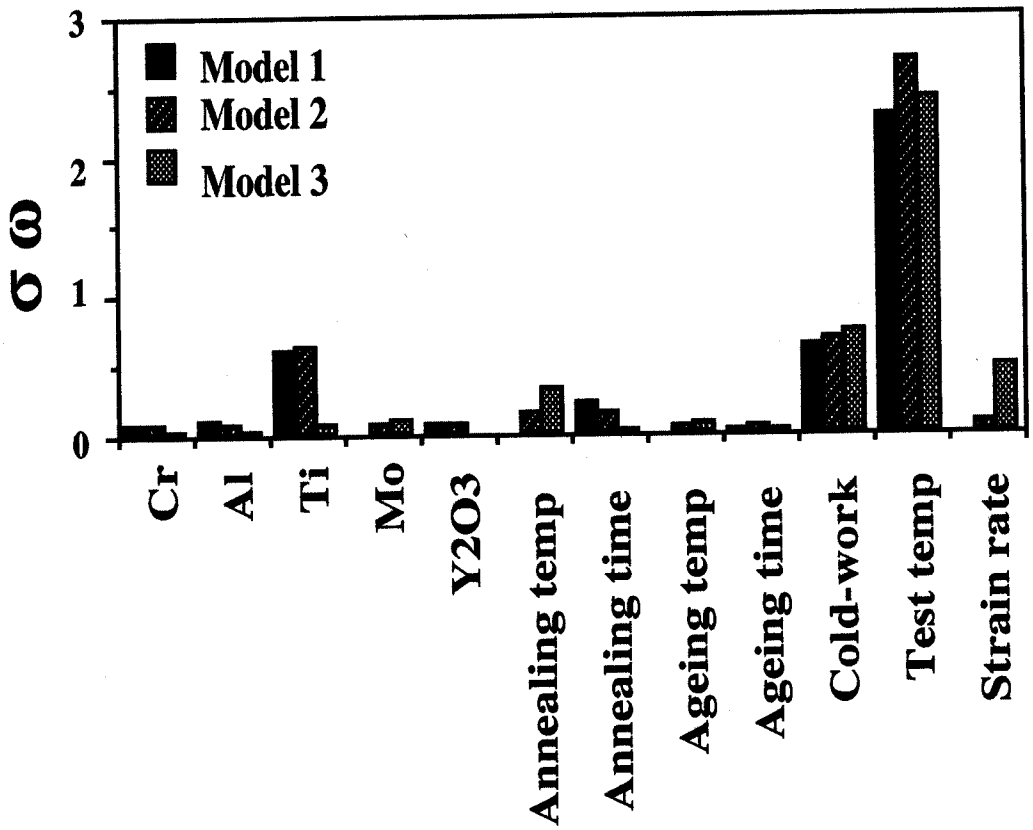

Figure 5.17: The model-percieved significance of the input parameters for the committee model trained on all data. The σ_w value for all the members of the committee are presented for each of the variables.

5.9 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 5.7). The inclusion of the strength parameters became necessary after the initial attempts to train the network without the parameters failed to produce good result. The reason for this behaviour is obvious, 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 Figure 5.18. Test error and log predictive error as functions of hidden units are shown in Figure 3.19.

Table 5.8 gives the numerical data of the ten best models ranked according to their log predictive errors. Though the ranking of the models was by their log predictable error the optimum committee was determined using the validation error. The test errors of the ten committees are shown in Table 3.9

Table 5.7 : The variables used in the 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
Recrystallization temperature, °C	20-1330	729	585
Recrystallization time, s	0-120	28.92	33.72
Ageing temperature, °C	20-800	171	309
Ageing 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}	0.00000003-0.03	0.0013	0.0027
UTS, MPa	70.7-1680	545	390
Yield strength, MPa	63-1600	468	367
Elongation, MPa	0.8-49.29	12.13	8.18

and Figure 5.20 shows the test error versus ranking of the best models and test errors of the committees.

As shown in Fig. 3.20(b) the commitee consisting of the top six models shows the least energy error and was therefore chosen as the optimum model. Figures 3.21 and 3.22 give the plots of the predicted values versus experimental values for the single best model and the committee respectively. The plot for the committee after retraining the committee models with the whole dataset is given in Figure 3.23.

Figure 5.24 shows the plot of σ_ν, the model percieved significance of the variables for the six models contained in the committee. Test temperature is shown to have a large effect on elongation. This is expected and it is exciting to note the consistency with the patterns shown in the yield and ultimate strength models. The percieved influence of yield strength on elongation is slightly higher than for ultimate tensile strength on elongation. This is expected metallurgically since it is the difference between the yield strength and UTS that relates to the uniform component of elongation.

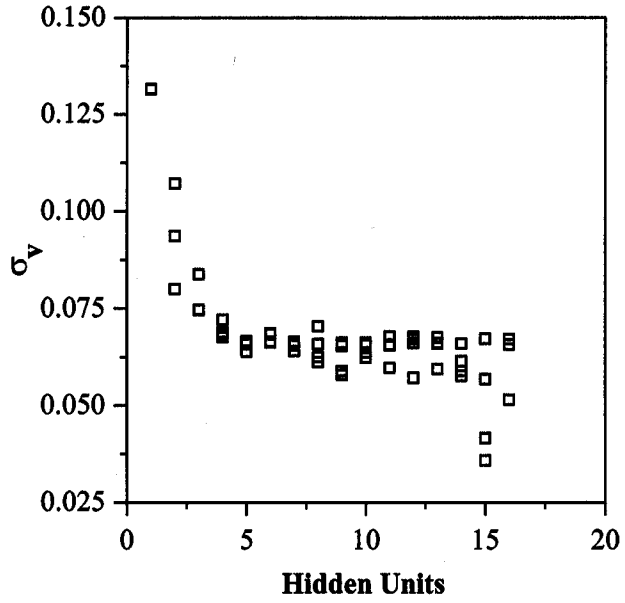

Figure 5.18: Variation in σ_ν as a function of hidden units.

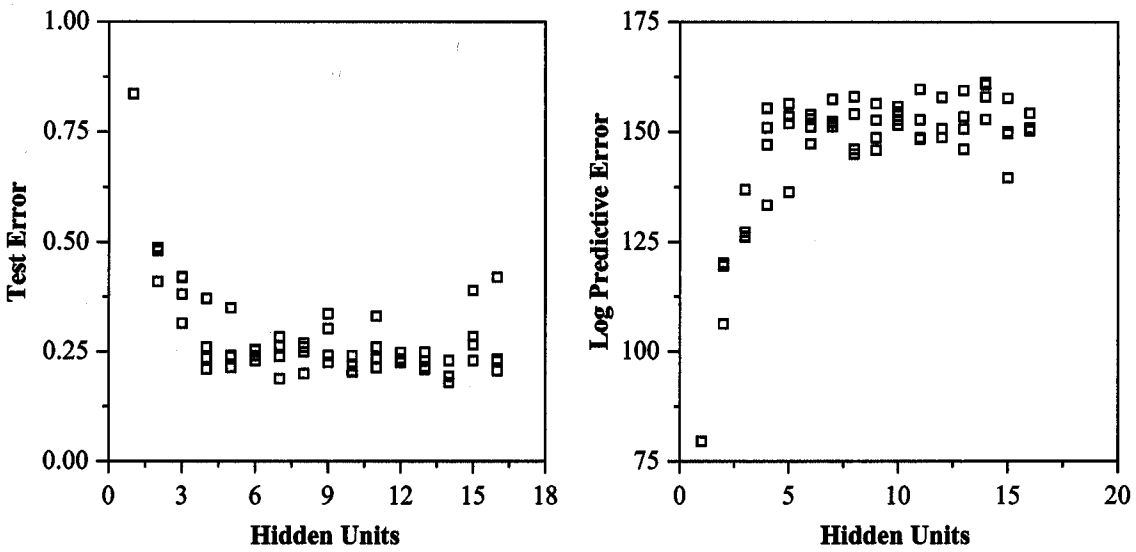

Figure 5.19: Variation in test error and log predicive error as functions of hidden units.

Table 5.8 : Ranking by log predictive error for the ten best models of the elongation

Ranking	Hidden units	Seed	Test error	Log predictive error
1	14	50	0.17988	161.23
2	14	30	0.19329	160.86
3	11	30	0.21316	159.76
4	13	50	0.20960	159.47
5	8	100	0.19966	158.05
6	7	50	0.18807	157.47
7	5	100	0.21390	156.42
8	4	30	0.21008	155.44
9	10	100	0.20382	155.81
10	10	10	0.20522	154.51

Table 5.9 : Test errors of the committees.

Models in committee	Test error
1	0.17987
2	0.17642
3	0.17930
4	0.17921
5	0.18035
6	0.17368
7	0.17504
8	0.17702
9	0.17639
10	0.17726

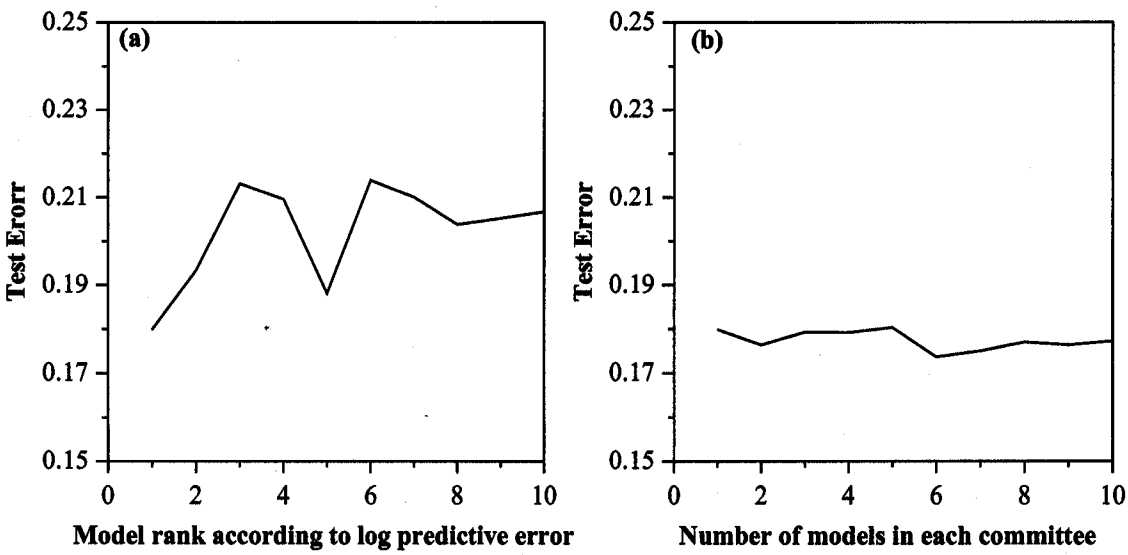

Figure 5.20: (a) Test errors of the top ten of the elongation models and, (b) corresponding test errors for commitees of models.

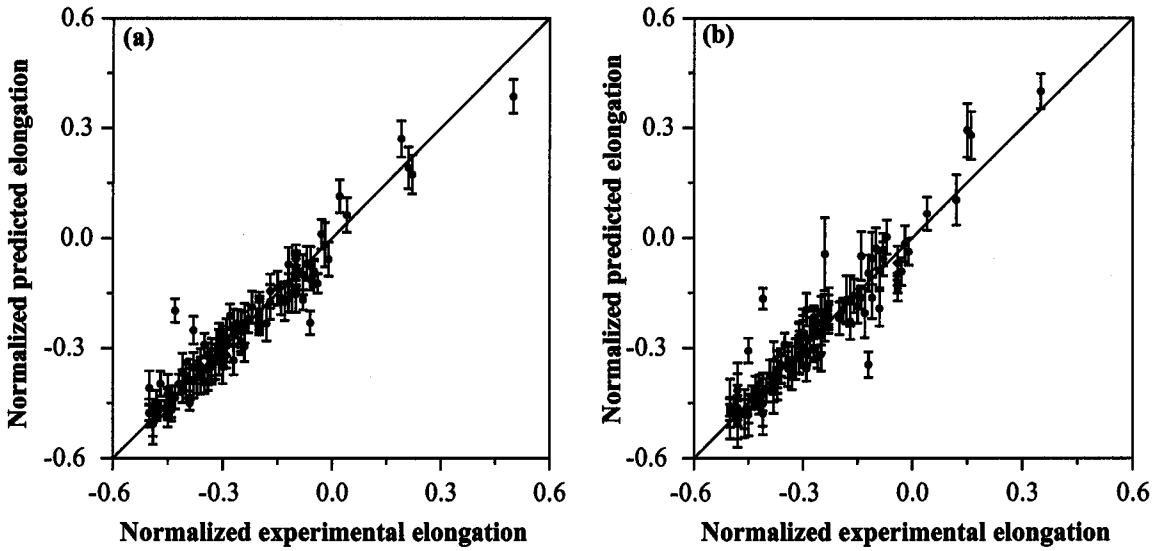

Figure 5.21: Predicted versus experimental results using the single best model. (a) training dataset (b) test dataset.

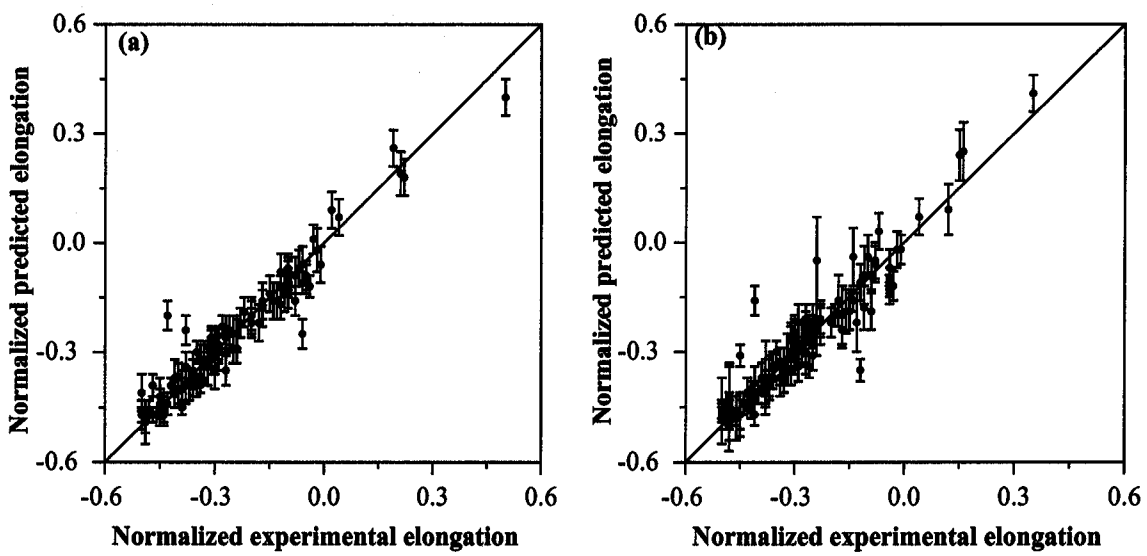

Figure 5.22 : Predicted versus experimental results using the optimum committee. (a) training dataset (b) test dataset.

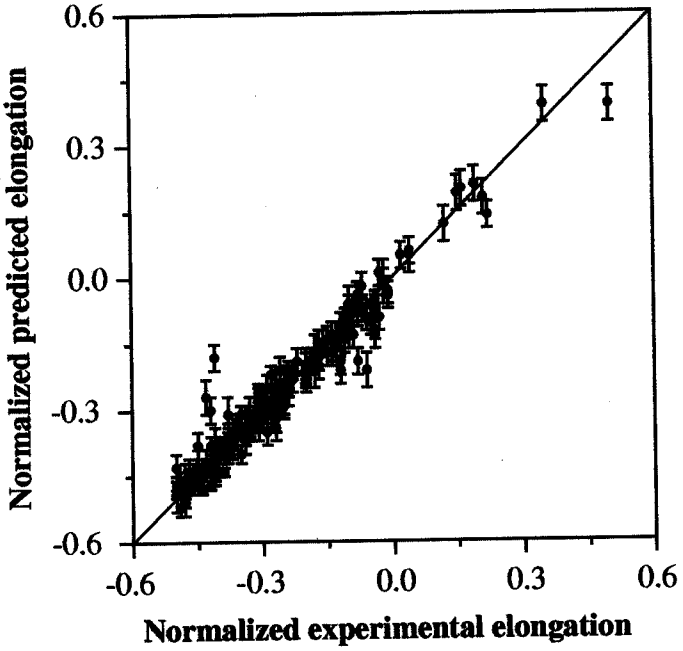

Figure 5.23 : Predicted versus experimental results for the whole dataset after the retraining using the committee.

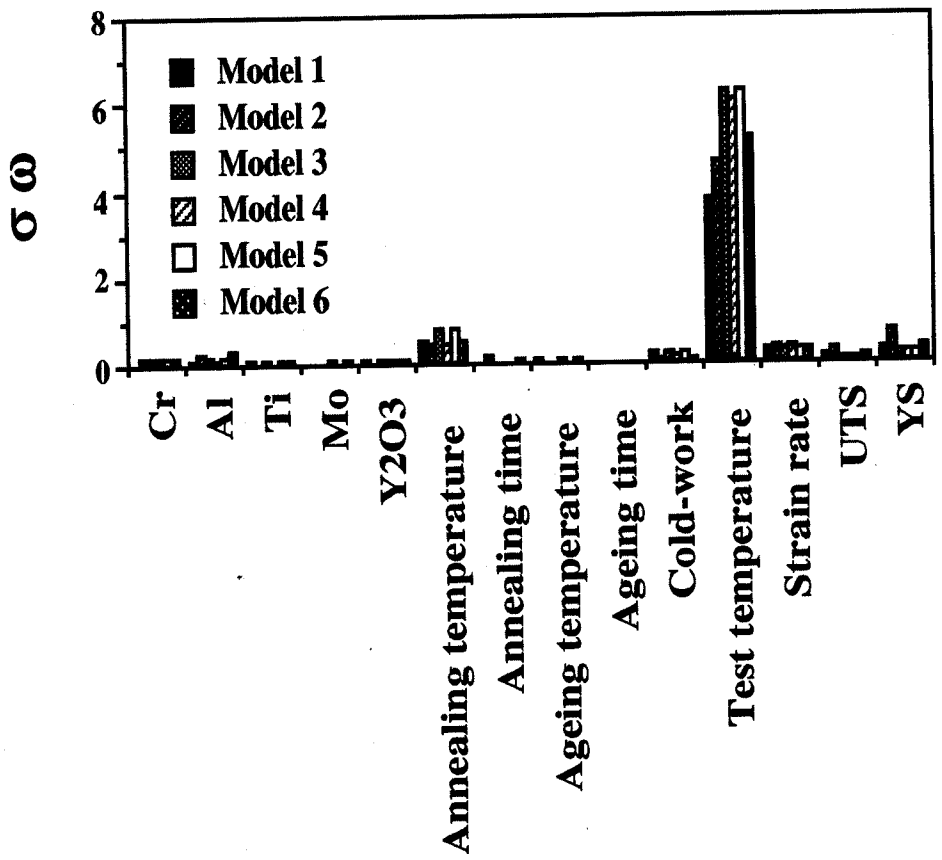

Figure 5.24 : The model-percieved significance of the input parameters for the committee model trained on all data. The σ_w value for all the members of the committee are presented for each of the variables.

5.10 Application of the models

Attempts were made to use the commitee models to predict the influence of the variables on the ultimate tensile strength, yield strength and percent elongation of the iron-base MA-ODS alloys and to see if the perceived relationships are reasonable from the point of view of the established metallurgical facts. All the results are presented along with the ± 1 standard deviation predicted error bars.

5.10.1 Effect of temperature

Figure 5.25 shows the predicted effect of test temperature on the ultimate strength, yield strength and elongation of MA956 for both recrystallized and unrecrystallised conditions.

The predicted patterns are quite reasonable. There is no significant fall in strength until around 500°C. Similarly, there isn't any noticeable change in elongation until precisely the same region of fall in strength when the elongation increases. This is consistent with the established facts that increasing temperature leads to decrease in strength and an increase in elongation. However, the sharp drop in strength and corresponding sharp rise in elongation is peculiar though well known [Alamo *et al.*, 1990, 1992; Regle, 1994]. An explanation that the sharp changes occur at a temperature where the dislocation density effectively drops would have been appropriate if the pattern had not been the same for both the recrystallized and unrecrystallized conditions. It may be the case that dislocations climb over the fine yttria particles becomes prominent in the regime where the sharp drop is observed. The alloy is seen to show higher percent elongation in the unrecrystallised condition than the recrystallised. This seems unusual as the alloys in the unrecrystallised condition are harder and would be expected to be less ductile. However, this result is consistent with the extensive experimental work by Alamo (1992) which indicates that it is the coarse columnar grain structure which leads to poor ductility in the recrystallised alloys.

The strengths and percent elongation of the various commercial MA-ODS ferritic steels are compared in Figure 5.26.

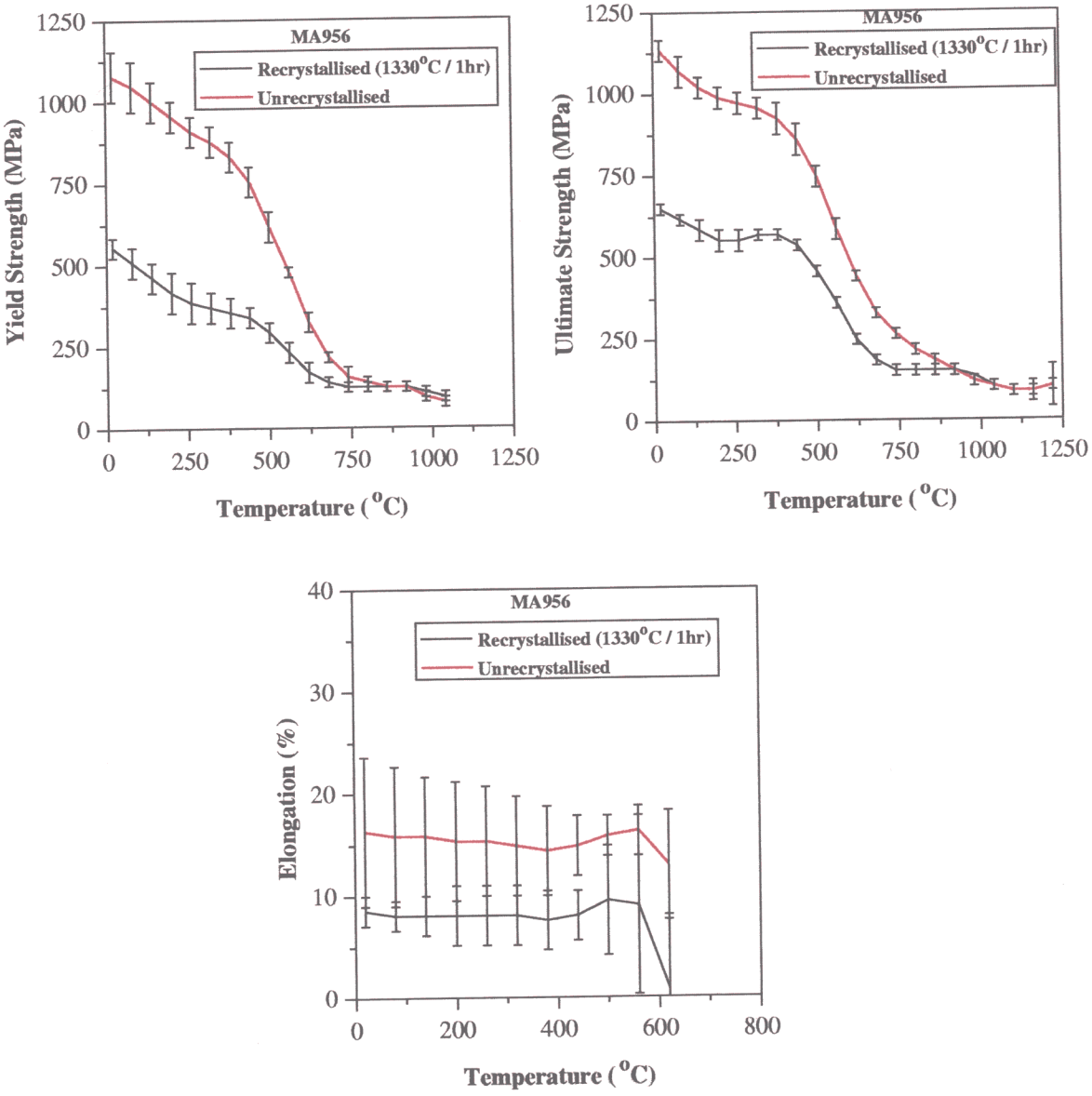

Figure 5.25: Effect of temperature on the yield strength, ultimate strength and percent elongation of MA956.

The heat-treatment conditions of the alloys are as for commercial applications. DT and DY are given ageing treatment to precipitate χ-phase for higher strength. This effect is correctly predicted with higher ultimate strength for DT and DY than MA956 which contains no χ-phase. DY contains yittria particles whereas DT doesn't and this explains the higher strength of DY than DT alloy. As shown the yield strength of MA956 is higher than DT despite the fact that DT contains χ-phase. This is because of the yittria particles in MA956 and clearly demonstrates that the effect of dispersoids strengthening

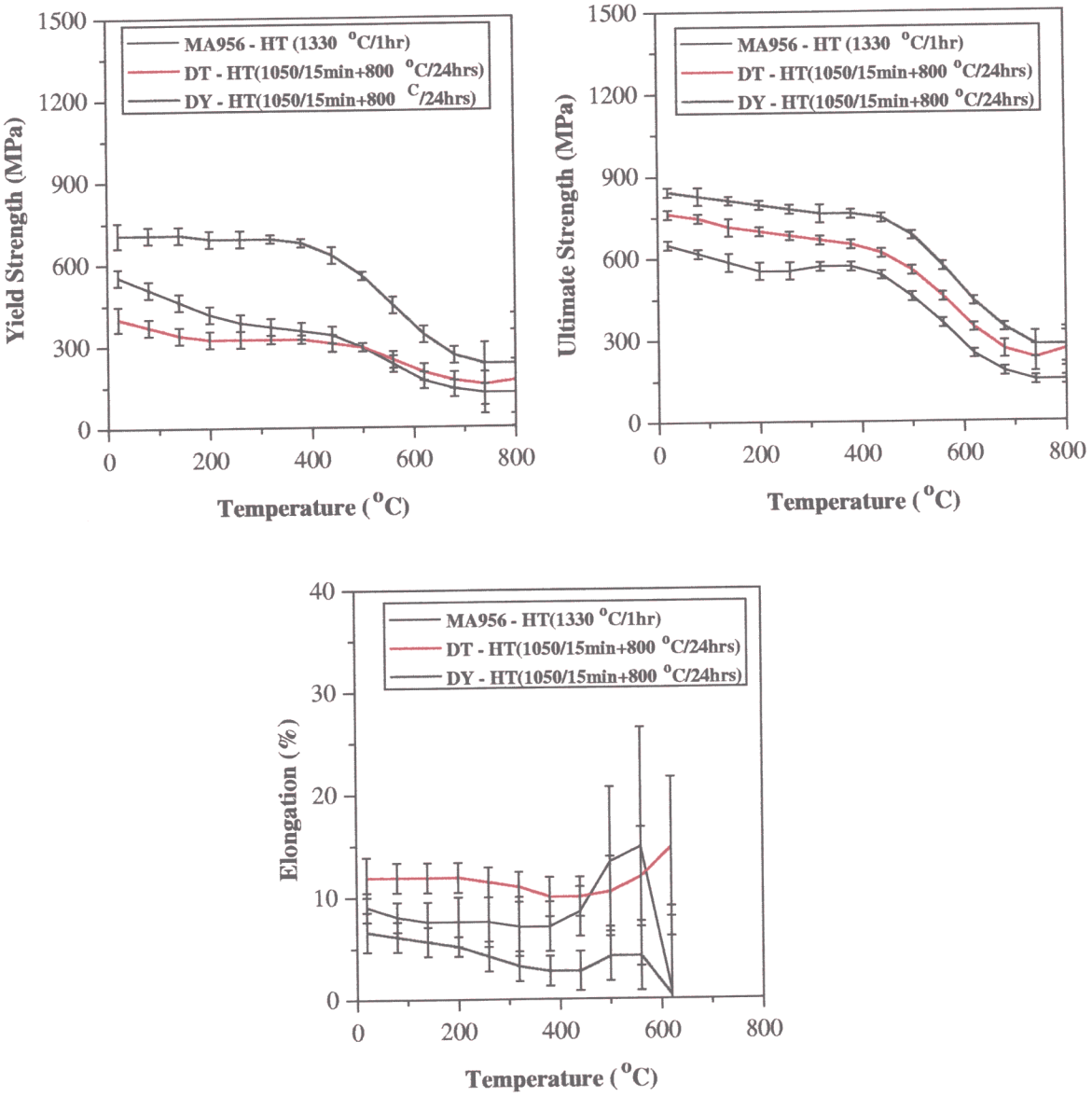

Figure 5.26: Comparison of the predicted properties of some MA-ODS ferritic steels.

is primarily on the yield strength rather than on the ultimate strength. The plots of percent elongation further show the effects of dispersoids with DT showing a better ductility because yittria particles are not present. It can then be summarized that yittria particles increases yield strength and reduces ductility and that the lower ductility in DY in which both χ-phase and yittria particles are present is as a result of the yittria particles.

5.10.2 Effect of Titanium, Molybenum and Yttria content

Figure 5.27 shows the effects of titanium, yittria , and molybdenum on the ultimate strength, yield strength and elongation of MA956. The ultimate strength and yield strength increase with weight percent content of titanium, elongation, however, appears insensitive. The increase in strength with increasing titanium is supposed to be via χ-phase but this is not formed in MA956, in which the titaninum solid solution strengthens, thus affecting both the ultimate and yield strength almost equally. The effect of yittria content is slightly more pronunced on the yield than the ultimate tensile strength, though the error bars are large. The bar charts of the model percieved significance of variables for the ultimate strength and yield strength in Figures 3.10 and 3.17 respectively, agreeably show a higher yittria effect for the yield strength than the ultimate strength. The results of an experimental work by Kawasaki *et al.*(1996) on the effect of dispersoids on tensile deformation of Fe-20Cr ODS alloys has helped to establish the reliability of the predicted patterns. Addition of ytrria particles was reported [Kawasaki, *et al.*, 1996] to increase the 0.2% yield stress all over the experimental temperature range (300–1073 K) and that at higher temperatures than 673 K, the increament of work hardening due to the dispersoids is small. The ultimate tensile strength is a function of work hardening and this probably explains why the predicted effect of dispersoids on the ultimate tensile strength is generally smaller than for the yield strength. These predicted behaviours with respect to titanium and yittria concentrations are significant as they seem to corroborate the explanation given earlier for the predicted tensile properties of the different MA-ODS steels. Titanium through χ-phase is responsible for the higher ultimate strength of DT and DY than MA956. The higher yield strength of MA956 than DT and the higher ductility of DT than MA956 and DY is because of the yittria particles in MA956 and DT. Strength and elongation are shown to be insensitive to molybdenum content in MA956. Apart from being a solid solution strengthener, molybdenum is a constituent in χ-phase and as such increasing its concentration is expected to contribute positively to strength. However, the titanium level upon which formation of χ-phase depends is very low in MA956 compared to the levels in the alloys where χ-phase is formed so

increasing molybdenum concentration may not have any effect. The model seems to have recognised correct patterns.

5.10.3 Effect of Chromium and Aluminium

The predicted effects of chromium and aluminium on the tensile properties of MA956 are shown in Figure 5.28 . Changes in concentration have negligible effects on the strength and elongation of MA956. However, the error bars are so large that it is not possible to reach a satisfactory conclusion. Large error bars occur either due to noisy data or due to sparce data. It is suspected that both of these factors are responsible for the lack of a significant relationship for chromium or aluminium.

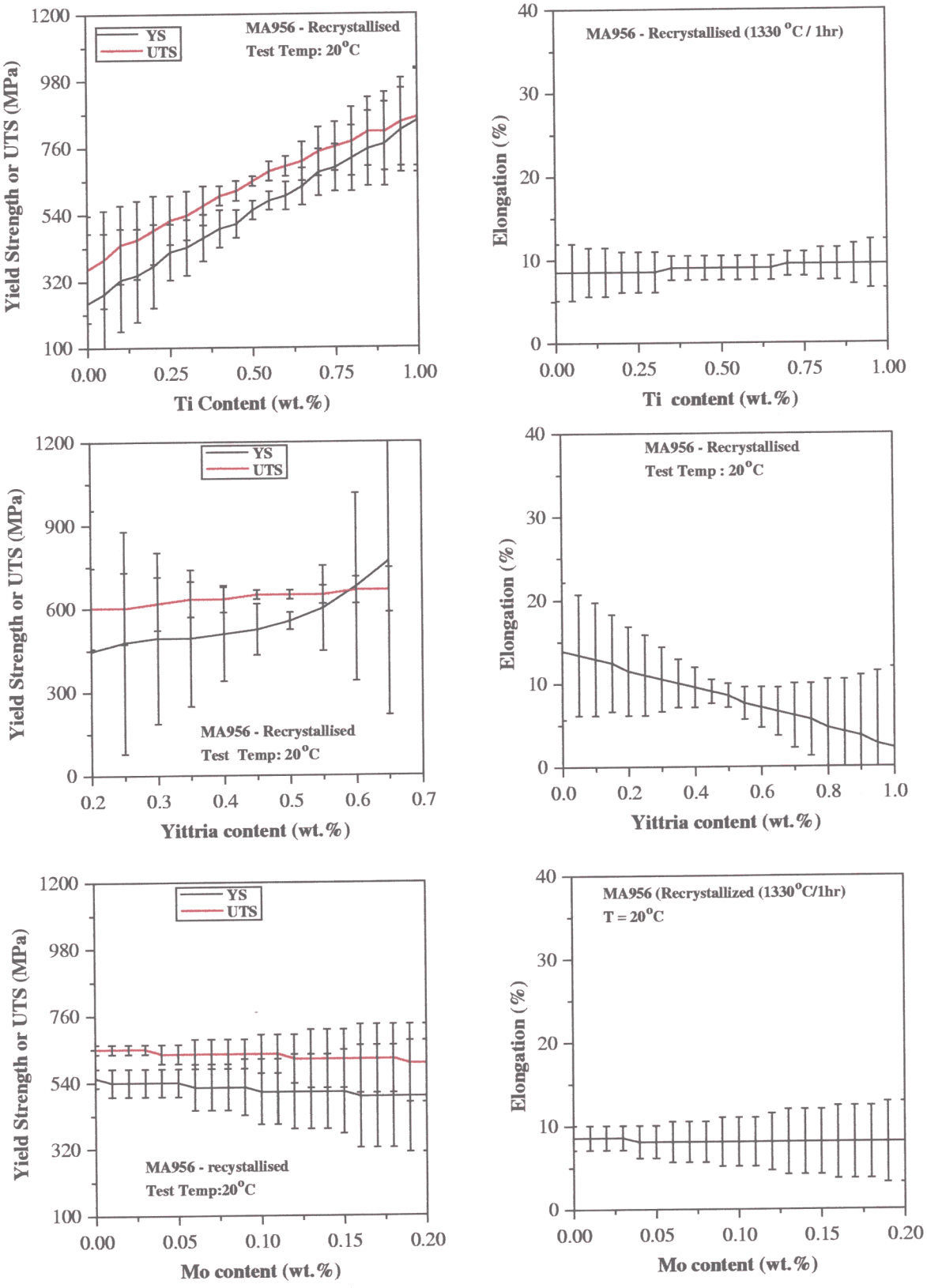

Figure 5.27 : Predicted effect of titanium, yttria and molybdenum on the yield strength, ultimate strength and percent elongation of MA956.

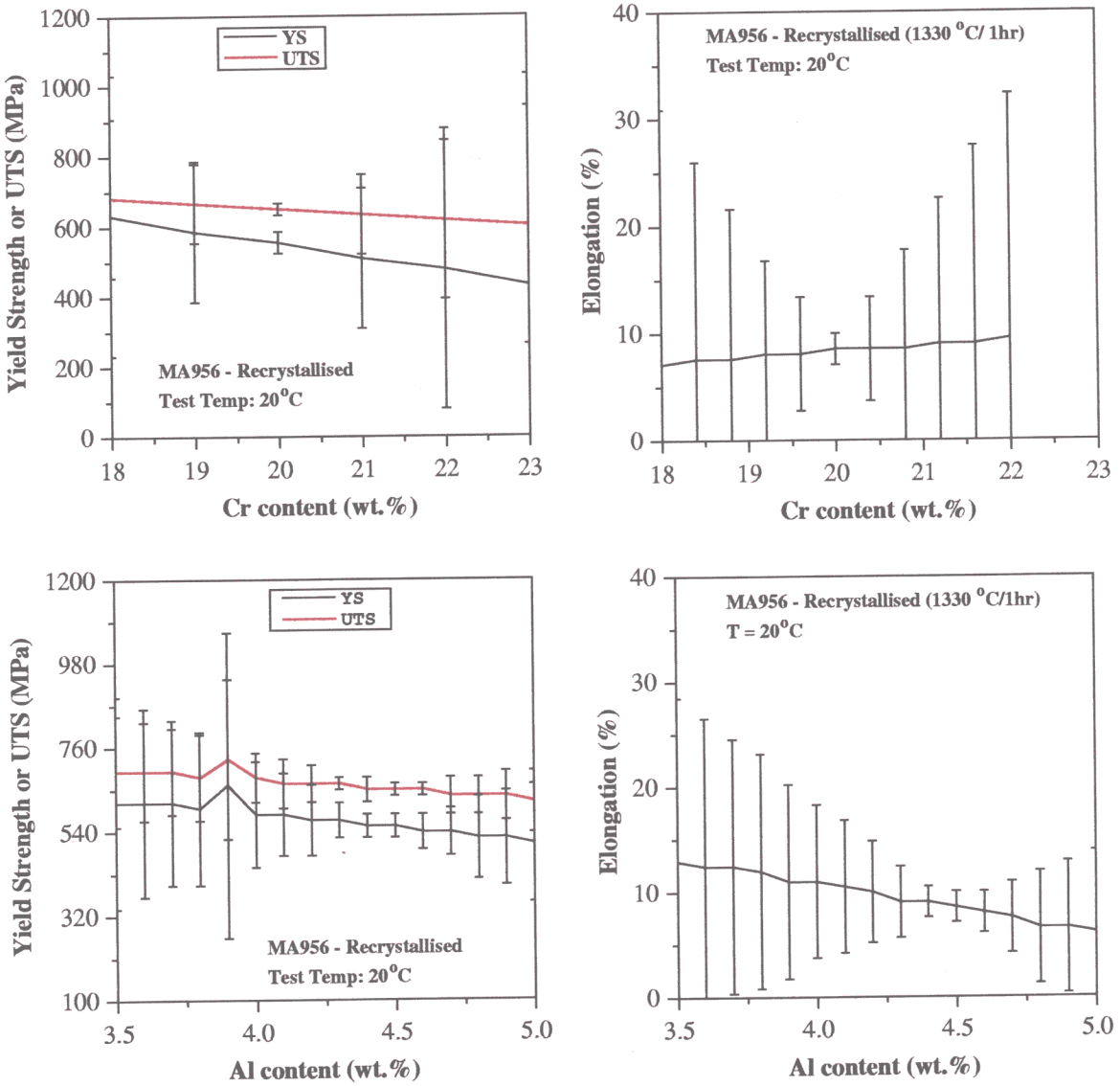

Figure 5.28 : Predicted effect of chromium and aluminium on the yield strength, ultimate strength and percent elongation of MA956.

5.10.4 Effects of recrystallisation temperature and time

The effects of the recrystallisation temperature and recrystallisation time on the ultimate strength, yield strength and elongation are shown in Figure 5.29 . As expected the ultimate strength and yield strength decrease with increasing recrystallisation temperature and time whilst the elongation increases.

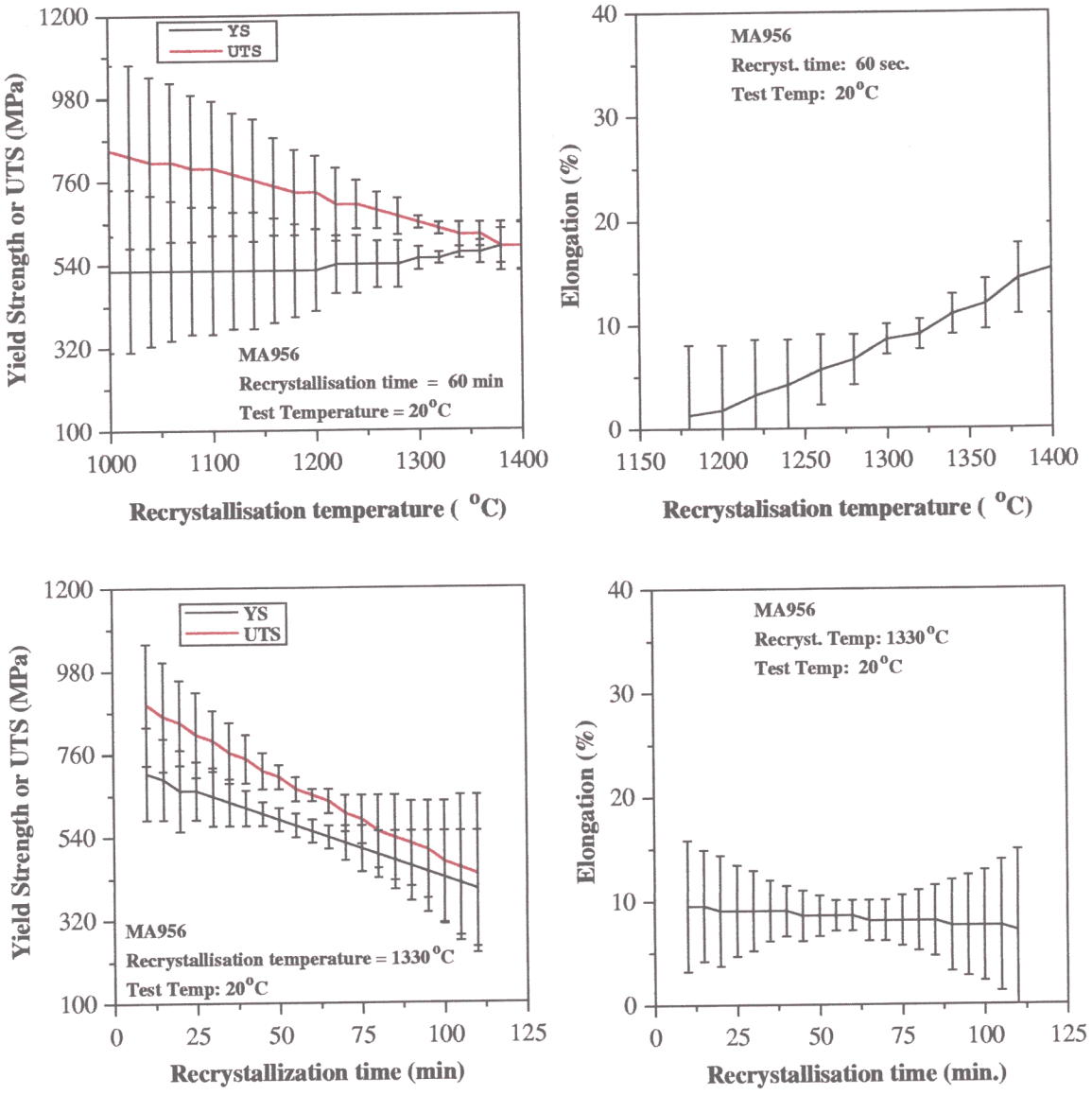

Figure 5.29 : Effects of recrystallisation temperature and recrystallisation time on the tensile properties of MA956.

5.10.5 Effects of cold-work and strain rate

Figure 5.30 shows the the predicted effects of cold-work and strain rate on the ultimate strength, yield strength and elongation of MA956. As expected, increasing cold-work increases the ultimate strength and the yield strength and saturation level is predicted above which further cold-work does not lead to increase in strength. Though the error bars are large for the elongation, the predicted pattern is what is expected, decreasing ductility with increasing cold-work. The large error bar is due to the reason that there is very limited examples where cold-work is varied in the database. The strength and elongation

are shown to be insensitive to low strain rate but at high strain rate, strengths increase and elongation decreases accordingly with increasing strain rate. These results are in excellent agreement with the report of an experimental work by [Whittenberger, 1979].

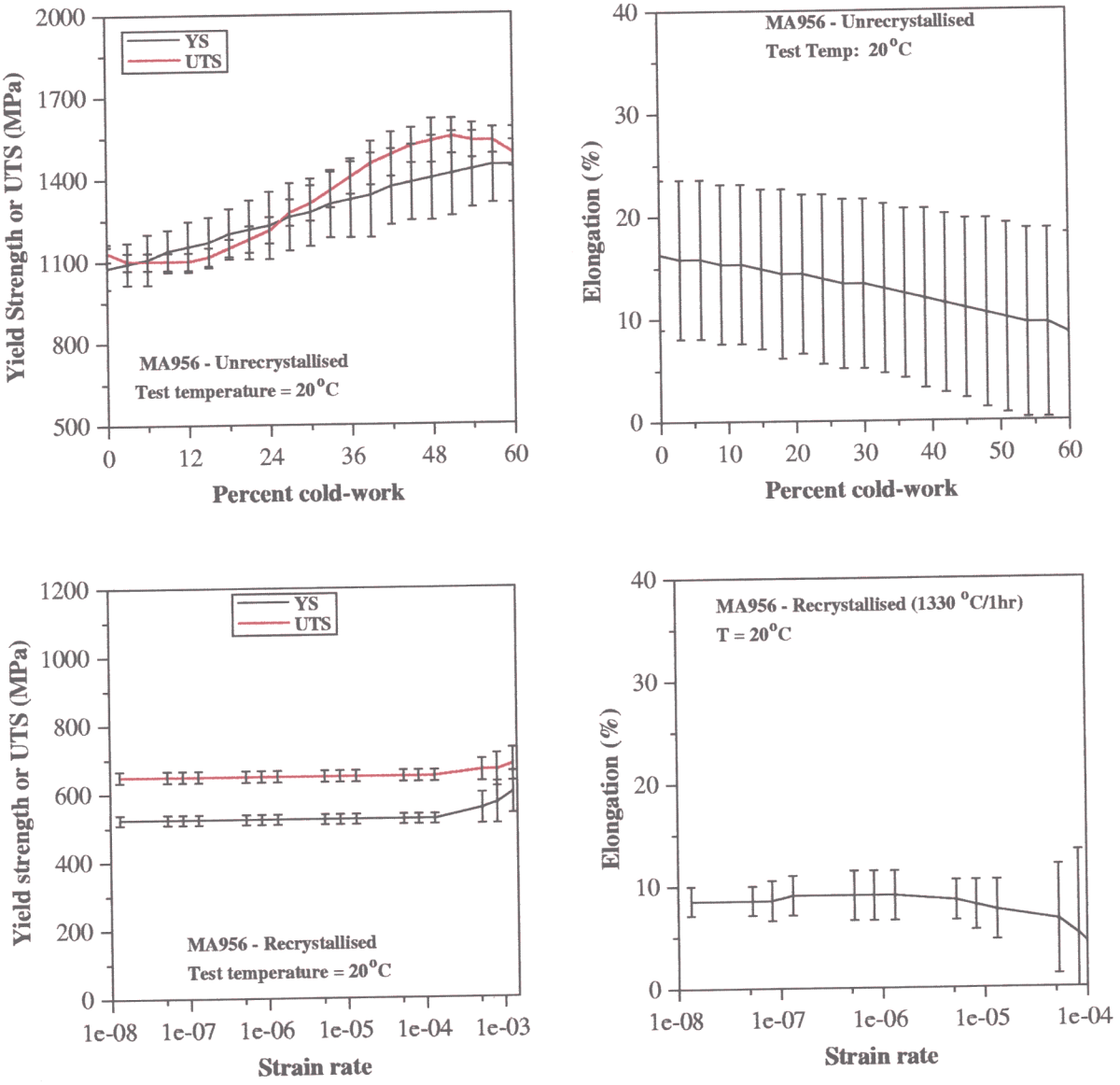

Figure 5.30 : Effects of cold-work and strain rate on the tensile properties of MA956.

5.11 Summary

Tensile properties data for mechanically alloyed, oxide dispersion strengthened ferritic stainless steels have been analysed using a neural network technique within a Bayesian framework. The analysis, though empirical, can after appropriate training and with the use of committee of models, produce results which are metallurgically reasonable.

Our experience of the neural network method suggests that it has considerable potential for useful applications in materials science. It is particularly useful in circumstances where there is extreme complexity, such that physical models are not available to construct within a reasonable time scale.

Neural networks are frequently used for regression problems in which continuous variables are modelled. They can also be applied to classification problems where the variables to be predicted adopt discrete values [Ichikawa, *et al.*, 1996].

The technique is extremely powerful; it can in principle produce a model for a random set of points. There are many models available, for example, on the world-wide web or from commercial sources. Care must however, be taken to select those which have a good strategy for avoiding the problem of overfitting the data. Methods in which the error bar depends on the position in the input space are particularly safe.

Finally, the neural network, like all regression methods, is a purely mathematical tool which cannot necessarily distinguish between cause and effect. The selection of proper inputs and outputs is important in deducing physically sound relationships.