Neural Networks and Genetic Algorithms in Materials Science and Engineering, 2006
January 11–13, 2006, Tata McGraw–Hill Publishing Company Ltd., India
Bengal Engineering and Science University, Shibpur, Howrah, India

Neural Networks in Materials Science: The Importance of Uncertainty

H. K. D. H. Bhadeshia

University of Cambridge, Materials Science and Metallurgy
Pembroke Street, Cambridge CB2 3QZ, U. K.
www.msm.cam.ac.uk/phase–trans

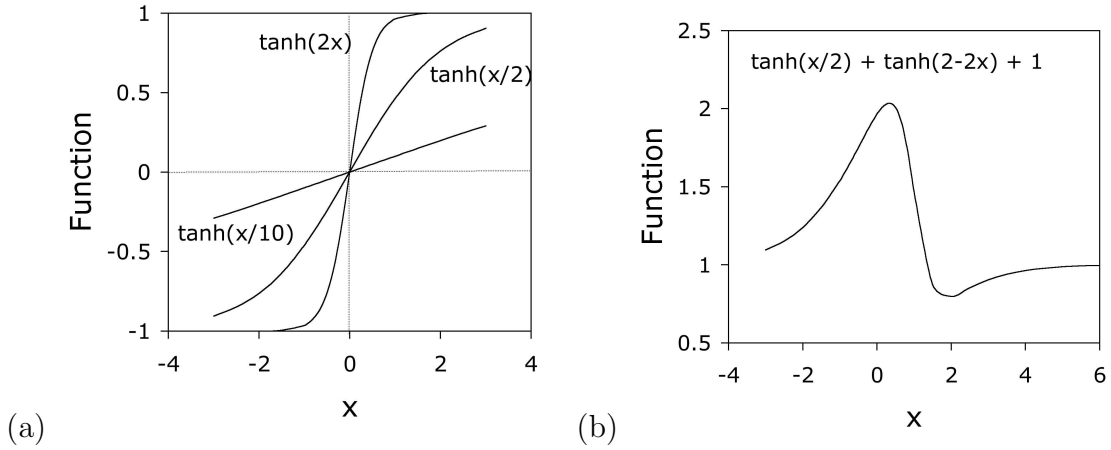
Abstract

It is immensely satisfying to be able to discover fundamental relationships and structure within vast arrays of ill–understood data. The neural network method is a powerful tool in this respect, a mathematically transparent technique which is able to capture complex relationships without the need to fix the mathematical form at the outset. These capabilities are enhanced by a proper consideration of errors and uncertainties. I introduce here the method and then go on to show why large uncertainties in the data need not be depressing. Uncertainty helps define novel experiments and stimulates questions about the fundamental relationships in nature.

INTRODUCTION

The usual approach when dealing with difficult problems is to correlate the results against chosen variables using linear regression analysis; a more powerful method of empirical analysis involves the use of neural networks. Since the method has been described elsewhere [1, 2, 3, 4, 5], what follows is an emphasis of the essential features. There is also a comprehensive world wide web resource on www.msm.cam.ac.uk/phase-trans/abstracts/neural.review.html

In conventional regression analysis the 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 $y = \sum_j w_j x_j + \theta$. Equations like these are used widely in industry, for example, in the formulation of the *carbon equivalents* which assign



(a) (b)

Figure 1: (a) Three different hyperbolic tangent functions; the "strength" of each depends on the weights. The diagram shows how flexible a hyperbolic tangent is. (b) A combination of two hyperbolic tangents to produce a more complex model. Such combinations can be continued indefinitely to produce functions of ever greater complexity.

the effect of individual solutes in steel on its overall behaviour:

$$\mathrm{CE} = \mathrm{C} + \frac{\mathrm{Mn} + \mathrm{Si}}{6} + \frac{\mathrm{Ni} + \mathrm{Cu}}{15} + \frac{\mathrm{Cr} + \mathrm{Mo} + \mathrm{V}}{5} \qquad \text{wt.\%} \tag{1}$$

Another example is the expression of mechanical properties of certain weld metals as a function of the chemical composition:

$$\text{yield strength / MPa} = 484 + 57 \times w_{Cu} \tag{2}$$

where w_{Cu} is the weight percent of copper. It is well understood that there is risk in using the relationships beyond the range of fitted data, but the risk is not quantified.

With neural networks, the input data x_j are again multiplied by weights, but the sum of all these products forms the argument of a flexible mathematical function, often a hyperbolic tangent. The output y is therefore a non–linear function of x_j. The exact shape of the hyperbolic tangent can be varied by altering the weights (Fig. 1a). Further degrees of non–linearity can be introduced by combining several of these hyperbolic tangents (Fig. 1b), so that the neural network method is able to capture almost arbitrarily non–linear relationships.

Fig. 2 illustrates the complexity of the surface that can be produced when representing the output (vertical axis) as a function of two inputs using just four hyperbolic tangents. A potential difficulty with the ability to produce complex, non–linear functions is the possibility of overfitting of data. It is possible, for example, to produce a neural network model for a random set of data.

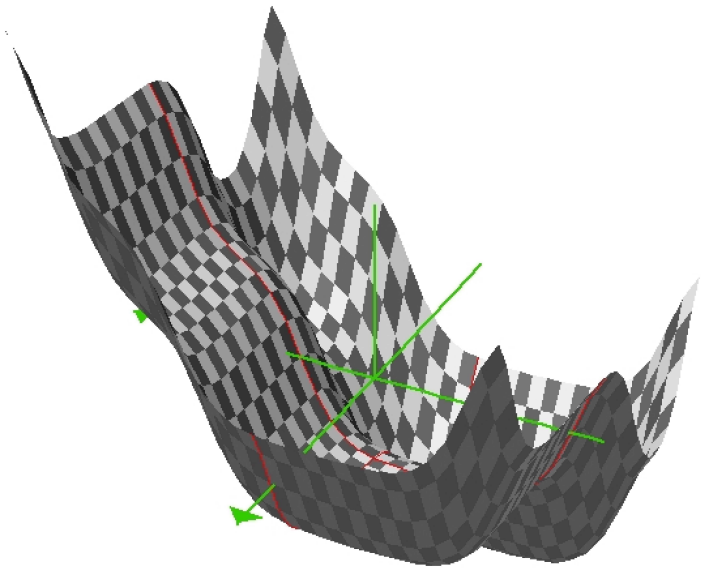


Figure 2: Variation in the output (vertical axis) as a function of two input variables (horizontal axes), the whole surface being generated using just four hyperbolic tangent functions.

OVERFITTING

A potential difficulty with the use of flexible non–linear regression methods is the possibility of overfitting data. In Fig. 3, the points represent all of the available (noisy) experimental data. There are two fitted functions, one a straight line and the other a non–linear polynomial function. It is not possible, without any guiding physical principles relating x to y, to assess which of these functions is the more reliable in extrapolation.

To avoid this difficulty, the experimental data can be divided into two sets, a *training* dataset and a *test* dataset. The model is produced using only the training data. The test data are then used to check that the model behaves itself when presented with previously unseen data. This is illustrated in Fig. 4 which shows three attempts at modelling noisy data for a case where y should vary with x^3. A linear model (Fig. 4a) is too simple and does not capture the real complexity in the data. An overcomplex function such as that illustrated in Fig. 4c accurately models the training data but generalises badly. The optimum model is illustrated in Fig. 4b. The training and test errors are shown schematically in Fig. 4d; not surprisingly, the training error tends to decrease continuously as the model complexity increases. It is the minimum in the test error which enables that model to be chosen which generalises best to unseen data.

This discussion of overfitting is rather brief because the problem does not

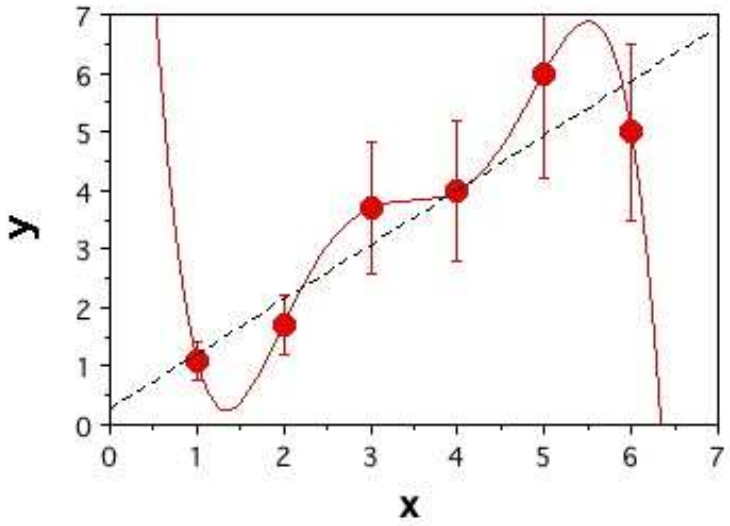


Figure 3: The points represent a set of noisy experimental data. There are two fitted functions, one a straight line and the other a non–linear polynomial.

simply involve the minimisation of test error. There are other parameters which control the complexity, which are adjusted automatically to try to achieve the right complexity of model [2, 3].

It is common practice in regression analysis to best fit a function to the data, *i.e.,* to use the most probable values of the weights for a given model. This results in an overall error obtained by comparing the predictions against experimental values, with no indication of the uncertainty as a function of position in the input space. MacKay has developed a treatment of neural networks in a Bayesian framework [2, 3, 4], which allows the calculation of error bars representing the uncertainty in the fitting parameters. The method recognises that there are many functions which can be fitted or extrapolated into uncertain regions of the input space, without unduly compromising the fit in adjacent regions which are rich in accurate data. Instead of calculating a unique set of weights, a probability distribution of sets of weights is used to define the fitting uncertainty. The error bars therefore become large when data are sparse or locally noisy. The Bayesian framework is also used to avoid overfitting and relevance determination [2, 3]. We now dwell into the uncertainties in greater depth.

NOISE AND UNCERTAINTIES

There are two kinds of errors to consider: when conducting experiments, the *noise* results in a different output for the same set of inputs when the experiment is repeated. This is because there are variables which are not controlled so their influence is not included in the analysis. The second kind deals with the *uncertainty of*

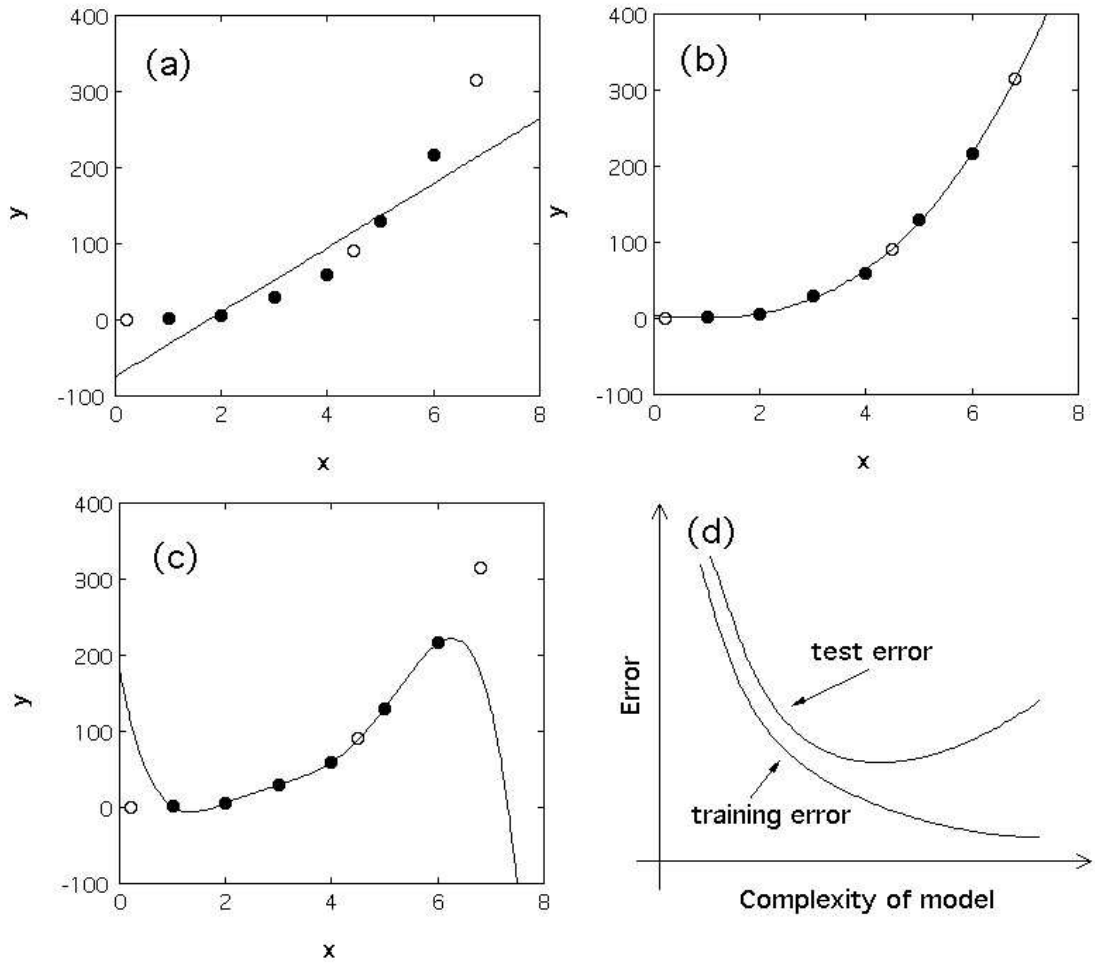


Figure 4: Variations in the test and training errors as a function of model complexity, for noisy data in a case where y should vary with x^3. The filled points were used to create the models (*i.e.,* they represent training data), and the circles constitute the test data. (a) A linear function which is too simple. (b) A cubic polynomial with optimum representation of both the training and test data. (c) A fifth order polynomial which generalises poorly. (d) Schematic illustration of the variation in the test and training errors as a function of the model complexity.

modelling; there many exist many mathematical functions which adequately represent the same set of empirical data but which behave differently in extrapolation.

The noise in the output can be assessed by comparing the predicted values (y_j) of the output against those measured (t_j), for example,

$$E_D \propto \sum_j (t_j - y_j)^2. \tag{3}$$

E_D is expected to increase if important input variables have been excluded from the

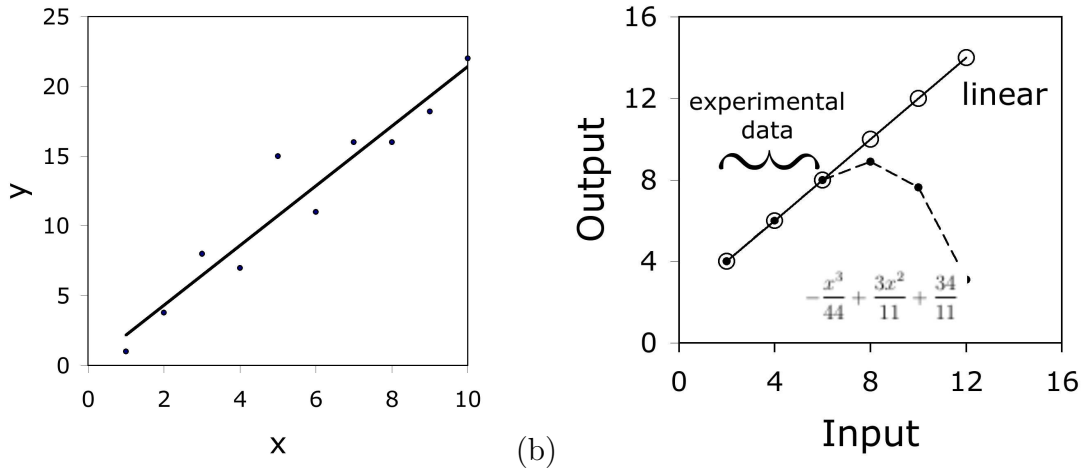


Figure 5: (a) Noise. (b) Uncertainty.

analysis. An equivalent description is the standard error σ_{SE} given by

$$\sigma_{SE} = \sqrt{\sum_{j=1}^{n} \frac{(t_j - y_j)^2}{n^2}}. \tag{4}$$

Fig. 5a shows this kind of noise where data have been fitted to a straight line with a standard error of ± 2 in the estimation of the output y.

Whereas E_D and σ_{SE} give an overall perceived level of noise in the output parameter, they are not, on their own, a satisfying description of the uncertainties of prediction. By contrast, the uncertainty of modelling is illustrated in Fig. 5b, where a set of precise data (2, 4, 6) are fitted to two different functions, one linear the other non–linear [6]:

$$y = -\frac{x^3}{44} + \frac{3x^2}{11} + \frac{34}{11} \tag{5}$$

Both of the functions illustrated precisely reproduce these data but behave quite differently when extrapolated (or indeed, interpolated, for $x = 3$, $y = 4.931$, not $y = 6$ according to the linear function). The difference in the predictions of the two functions in domains where data do not exist, is a measure of the uncertainty of modelling, since both functions correctly represent the data $x = 2, 4, 6$ used in creating the models. Notice also that unlike the noise, the magnitude of the modelling uncertainty is not constant, but varies as a function of the position in the input space. The uncertainty becomes larger in domains of input space where knowledge is sparse or non–existent, Fig. 5b.

One way of representing the uncertainty, is to create a large variety of models, all of which reasonably represent the experimental data. The predictions made by these models will not be identical; the standard deviation in the predicted values then is a quantitative measure of the modelling uncertainty.

Fig. 6 illustrates the problem; the practice of using the best–fit function (*i.e.,* the most probable values of the weights) does not adequately describe the uncertainties in regions of the input space where data are sparse (B), or where the data are particularly noisy (A).

In MacKay's method [2, 3], the modelling uncertainty is expressed by not having a unique set of weights, but rather a probability distribution of sets of weights. This recognises the existence of many functions which can be fitted or extrapolated into uncertain regions of the input space, without unduly compromising the fit in adjacent regions which are rich in accurate data. The error bars depicting the modelling uncertainty then become large when data are sparse or locally noisy, as illustrated in Fig. 6.

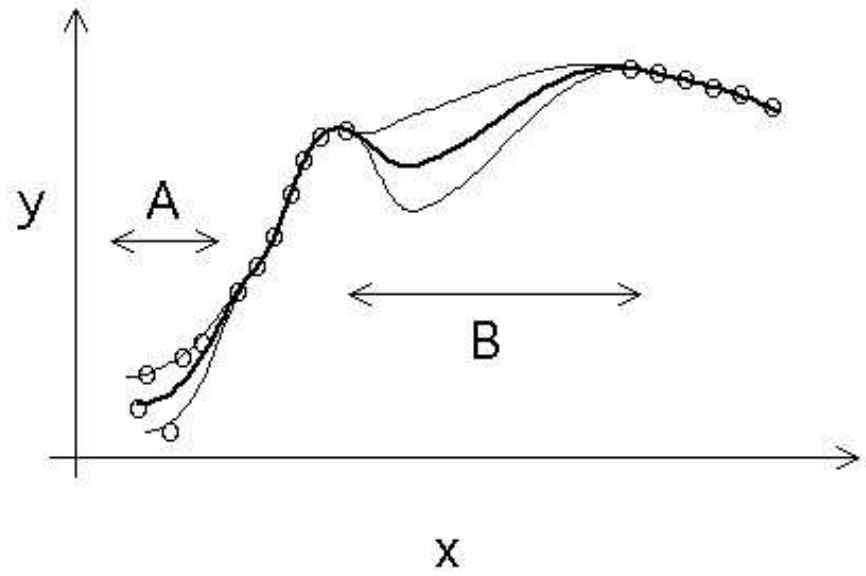


Figure 6: Schematic illustration of the uncertainty in defining a fitting function in regions where data are sparse (B) or where they are noisy (A). The thinner lines represent error bounds due to uncertainties in determining the weights.

This methodology has proved to be extremely useful in materials science [1] where properties need to be estimated as a function of a vast array of inputs. It is then most unlikely that the inputs are uniformly distributed in the input space.

SHADES OF GREY

Modelling techniques are sometimes classified as white, grey or black boxes, with the implication that the white–box model is completely transparent and comprehensible in its construction, whereas the black box is simply a magical instrument

which relates a set of inputs to the output without giving insight into how the conversion is achieved.

A great deal of what seems to me unnecessary terminology is used in describing a neural network; inappropriate comparisons are made with biological systems and synapses. This fanciful treatment of the networks gives them a status which is not justified and one which many treat with apprehension. Thus, neural networks are often referred to as black boxes. Unfavourable comparisons are even made against linear regression equations because the networks are supposed to lack transparency (this is outrageous given that a linear regression equation is a subset of a neural network).

The neural network is in fact a simple combination of transfer functions (in our case hyperbolic tangents) and weights. The number of hyperbolic tangents used is said to be the number of *hidden units*. A complete mathematical description of the network is straightforward. The function for a network with i hidden units, connecting the inputs x_j to the output y is given by

$$y = \sum_i w_i^{(2)} h_i + \theta^{(2)} \tag{6}$$

where

$$h_i = \tanh\Big(\sum_j w_{ij}^{(1)} x_j + \theta_i^{(1)}\Big) \tag{7}$$

where w represents weights and θ the constants as described in the context of linear regression.

The influence of the inputs on the output variable is, together with the transfer functions, implicit in the values of the weights. However, the weights may not always be easy to interpret given that there may be high–order interactions between the variables. For example, there may exist more than just pairwise interactions, in which case the problem becomes difficult to visualise from an examination of the weights. This visualisation problem is a feature of all non–linear methods, but is not a limitation because it is simple to use the trained network to make predictions, plot them, and to see how these depend on various combinations of inputs.

CHARPY TOUGHNESS: CASE STUDY

We can now see how these concepts can be applied to one of the most common tests used to characterise the energy absorbed during fracture, the Charpy test. A square section notched bar is fractured under specified conditions and the energy absorbed during fracture is taken as a measure of toughness. The Charpy test is empirical in that the data cannot be used directly in engineering design. It does not provide the most searching mechanical conditions. The sample has a notch, but this is less than the atomically sharp brittle crack. Although the test involves impact loading, there is a requirement to start a brittle crack from rest at the tip of the

notch, suggesting that the test is optimistic in its comparison against a propagating brittle crack. Most materials can be assumed to contain sub–critical defects so that the initiation of a crack seems seldom to be an issue.

The Charpy test is nevertheless a vital quality control measure which is specified widely in international standards, and in the ranking of samples in research and development exercises. It is the most common first assessment of toughness and in this sense has a proven record of reliability. The test is usually carried out at a variety of temperatures in order to characterise the ductile–brittle transition intrinsic to body–centred cubic metals with their large Peierls barriers to dislocation motion.

The toughness of a steel depends on many variables; that of a weld is dependent on many more. It is not possible to predict the Charpy toughness of a weld with any reliability.

Some of the variables that influence the impact toughness of a ferritic steel weld are listed in Table 1.

Table 1: Some variables controlling Charpy toughness.

Yield strength	Ultimate tensile strength
carbon	silicon
manganese	chromium
nickel	tungsten
sulphur	phosphorus
titanium	aluminium
nitrogen, hydrogen	oxygen *etc.*
primary microstructure	secondary microstructure
α, α_W, α_b, α_a *etc.*	welding parameters
degassing	heat treatment

A common belief is that the toughness of ferritic steel welds and of ferritic steels in general can be improved by adding nickel as an alloying element. The mechanism for this in not understood but it is speculated that those solutes which enable the cross–slip of screw dislocations, by reducing their dissociation, make them less effective as crack nuclei [7].

However, attempts at improving the Charpy toughness of commercial welding alloys of the type used in the fabrication of submarines, by increasing the nickel concentration, have consistently failed. These alloys have a base chemical composition which is summarised as follows:

Fe–0.03C–0.5Si–2Mn–0.5Cr–0.5Mo–3Ni wt%

A part of the reason for the failure to improve toughness using nickel is that we tend to generalise observations made on a limited range of alloy chemistry. A

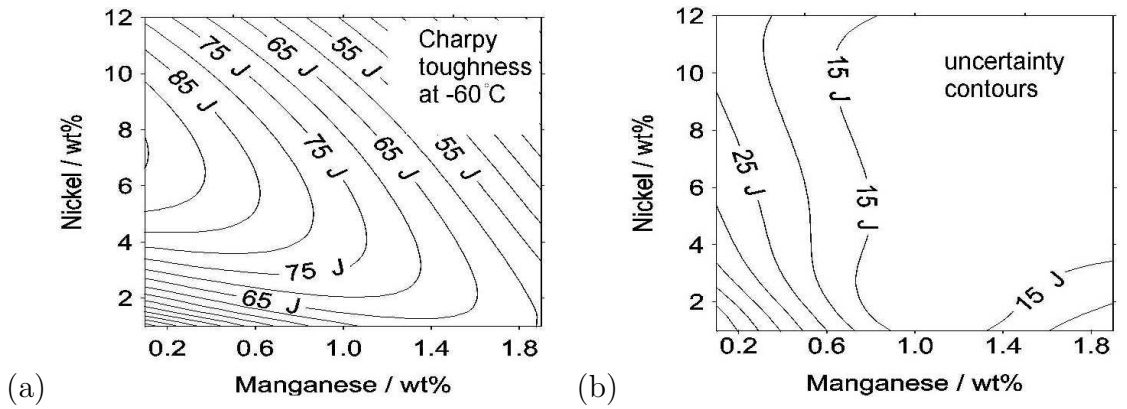


Figure 7: The combined effect of manganese and nickel on the calculated toughness for -60°C, of weld metal produced using arc welding with a heat input of 0.8 kJ mm^{-1}, with a base composition (wt%) 0.025 C, 0.37 Si, 0.006 S, 0.013 P, 0.21 Cr, 0.4 Mo, 0.011 V, 0.03 Cu, 0.039 O, 0.008 Ti, 0.019 N, and an interpass temperature of 250°C. (a) Contours showing the Charpy toughness. (b) Contours showing the $\pm 1\sigma$ uncertainty in the calculations.

neural network model was therefore created to cover a very large range of welding alloys; its purpose was to estimate the Charpy toughness as a function of many of the variables indicated in Table 1. Calculations using this model showed that in the class of welds described above, nickel is only effective in improving the Charpy toughness when the concentration of manganese is small (Fig. 7) [8]. No experiments were done in creating the model, which relied on published data. The only experiments were those conducted after the modelling, involving the manufacture of a new welding alloy which verified the results shown in Fig. 7.

Naturally, the plot (Fig. 7) does not in itself explain the mechanism by which the toughness is enhanced by nickel at low manganese concentrations, and *vice versa*. But it does suggest some clear paths for investigation, work which led to the discovery of a novel phase in weld metals, one which occurs when the nickel and manganese concentrations are both high [9]. This is illustrated in Fig. 8, which shows a scanning electron micrograph of extremely coarse coalesced bainite which is detrimental to toughness because of its scale.

The story presented here about the Charpy impact properties of welds represents an interesting example where a neural network was used to discover a pattern in huge quantities of complicated experimental data. The pattern was then used in conjunction with the uncertainties to identify a domain where a high–strength and high–toughness weld metal could be obtained (a domain never before accessed for manual metal arc welds) using an optimum combination of nickel and manganese in the steel. The weld metals were then manufactured and confirmed the calculations. A systematic high–resolution study of the metal then revealed the physical reason why the domain results in high toughness – the formation of coarse, coalesced bainite

as illustrated in Fig. 8.

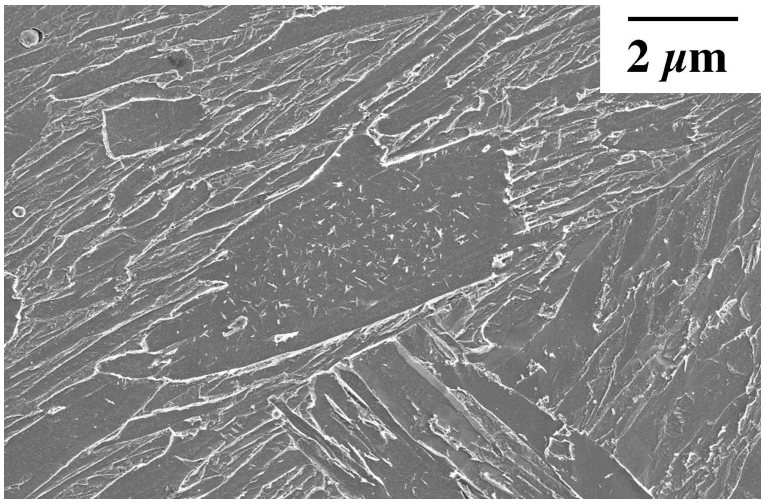


Figure 8: Coalesced bainite (centre) in a high nickel, high manganese weld metal. Photograph courtesy of Dr Enda Keehan, [9].

COPING WITH UNCERTAINTY

A large value of modelling uncertainty implies a lack of knowledge (sparse or noisy data). There are several advantages to making predictions with large modelling uncertainties. Firstly, a large uncertainty indicates a domain where it is useful to conduct new research. A small uncertainty by contrast could be interpreted to mean that new experiments are not needed. The estimate of uncertainty can therefore be used to design experimental programmes.

A large uncertainty does not necessarily mean that the prediction is likely to be wrong. If the network has been designed using physically meaningful inputs, and based on robust and large database, it is likely to make good predictions. This is illustrated in Fig. 9 where the uncertainties are large and yet the network has made excellent predictions [10].

SUMMARY

Neural networks now represent an indispensable tool in all aspects of material science. They become elegant when the predictions are associated with error bars consisting of contributions from noise and from modelling uncertainties. It is misleading to regard the networks as black boxes – they are mathematically transparent. Furthermore, their ability to cope with complexity and very large numbers of variables enables them to discover physical relationships and patterns which have not previously been possible. There are many examples where the networks have been exploited to design new materials and processes.

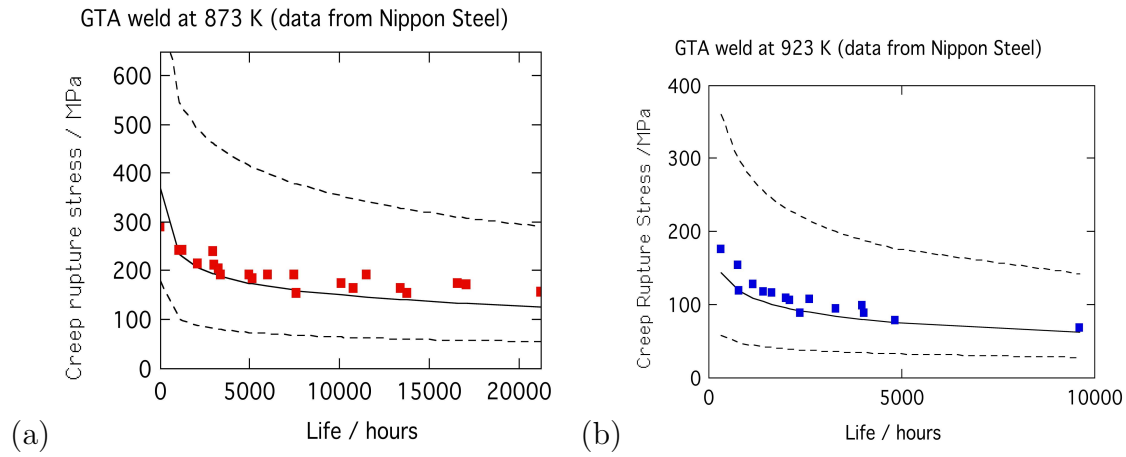


Figure 9: Comparison of the predictions of a neural network model (central curves) with experimental data which were not included in the creation of the model. The modelling uncertainty is indicated by the upper and lower error bounds [10].

We have emphasised the two kinds of uncertainties that should be considered in any analysis. The first is the noise that results from uncontrolled variables, generally reflected as a constant error bar in the output. The second is much more exciting, the uncertainty of modelling, which permits an assessment of the value of extrapolation or interpolation. This must be regarded as an essential part of any non–linear analysis since it is not otherwise possible to routinely determine whether one is extrapolating or interpolating in a complex field.

There are major advantages in taking the modelling uncertainty seriously. The first is that a large uncertainty defines a region where experiments would be useful and novel, that is, the domain in which *research* should be done. The second is that a large uncertainty leaves open the possibilities and hence stimulates ideas about whether the predicted relationship between the inputs and outputs has physical meaning.

> *"Uncertainty and mystery are energies of life. Don't let them scare you unduly, for they keep boredom at bay and spark creativity."*
>
> R. I. Fitzhenry

ACKNOWLEDGMENTS

I am grateful to Professors N. R. Bandyopadhyay, P. P. Chattopadhyay and S. Datta of the Bengal Engineering and Science University for organising this conference, and to the Royal Society (U.K.) for funding the journey to India.

REFERENCES

[1] H. K. D. H. Bhadeshia, "Neural Networks in Materials Science," *ISIJ International,* Vol. 39, 1999, 966–979.

[2] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation,* Vol. 4, 1992, 415–447.

[3] D. J. C. MacKay, "Practical bayesian framework of backpropagation networks," *Neural Computation,* Vol. 4, 1992, 448–472.

[4] T. Sourmail and H. K. D. H. Bhadeshia, "Neural Networks," *Introduction to Materials Modelling,* ed. Z. Barber, Institute of Materials, London, 2005.

[5] D. J. C. MacKay, "Probability theory and Occam's razor", *Mathematical Modelling of Weld Phenomena III,* eds H. Cerjak and H. K. D. H. Bhadeshia, The Institute of Materials, London, 1997, 359–389.

[6] D. J. C. MacKay, "Probability theory and Occam's razor," *Darwin College Magazine,* March 1993, 81–85.

[7] W. C. Leslie *Physical Metallurgy of Steels,* McGraw–Hill International Book Company, London, 1982, 124.

[8] M. Murugananth, H. K. D. H. Bhadeshia, E. Keehan, H. O. Andren, L. Karlsson. "Strong and Tough Steel Welds," *Mathematical Modelling of Weld Phenomena VI*, eds H. Cerjak and H. Bhadeshia, Institute of Materials, London, 2002, 205–230.

[9] E. Keehan, *Ph.D. Thesis,* Chalmers University, Sweden, 2005.

[10] D. Cole, C. Martin-Moran, A. Sheard, H. K. D. H. Bhadeshia and D. J. C. MacKay, "Modelling the Creep Rupture Strength of Ferritic Steel Welds," *Science and Technology of Welding and Joining,* Vol. 5, 2000, 81–89.