

# REVIEW

## Neural Networks and Information in Materials Science

H. K. D. H. Bhadeshia\*

*Department of Materials Science and Metallurgy, University of Cambridge, Cambridge CB2 3QZ, U.K*

Received 11 February 2008; revised 25 July 2008; accepted 18 August 2008

DOI:10.1002/sam.10018

Published online 16 January 2009 in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** Neural networks have pervaded all aspects of materials science resulting in the discovery of new phenomena and have been used in quantitative design and control. At the same time, they have introduced a culture in which both noise and modeling uncertainties are considered in order to realize the value of empirical modeling. This review deals with all of these aspects using concrete examples to highlight the progress made, whilst at the same time emphasizing the limitations of the method. © 2009 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 1: 296–305, 2009

### 1. INTRODUCTION

Neural networks are wonderful tools, which permit the development of quantitative expressions without compromising the known complexity of the problem. This makes them ideal in circumstances where simplification of the problem, in order to make it mathematically tractable, would lead to an unacceptable loss of information. As pointed out by Ziman [1], there is a fine balance between over-idealizing the initial hypothesis in order to make it amenable to mathematical analysis, and abandoning reality.

Materials science involves many of the pure disciplines [2,3] and the content is usually of technological interest. There are two common strands in all of its subdivisions (metals, ceramics, polymers . . .), experimental characterization and mathematical modeling, both of which cover vast length and time scales in their applications. The experiments collectively generate huge quantities of data, which are used to infer the properties of matter, or to formulate and validate theories. The mathematical models include combinations of atomistic calculations, thermodynamics, kinetics, finite element analysis, etc. Their functions may be categorized as follows [4]:

- Those which lead to unexpected outcomes that can be verified,
- Those which are created or used in hindsight to explain diverse observations,

- Existing models which are adapted or grouped to design materials or processes,
- Models used to express data, reveal patterns, or for implementation in control algorithms.

The successful application of the models depends to varying degrees on the existence of experimental data, simple observations, insight and knowledge beyond the boundaries of the pure mathematical framework. The purpose of this paper is to show how neural networks when used in context, are capable of achieving the goals of materials science in all of the four functions outlined above.

### 2. ESSENCE OF THE METHOD

When dealing with difficult problems, it is useful to correlate the results against chosen variables using regression analysis in which 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  $\theta$  then give an estimate of the output  $y = \sum_j w_j x_j + \theta$ . Equations like these are used widely, for example in the specification of thermodynamic data during phase diagram calculation methods. The heat capacity at constant pressure ( $C_P$ ) is expressed empirically as a function of the absolute temperature  $T$  as follows:

$$C_P = w_1 + w_2 T + w_3 T^2 + \frac{w_4}{T^2} \quad (1)$$

*Correspondence to:* H. K. D. H. Bhadeshia (hkdb@cam.ac.uk)

where  $w_i$  are the fitting constants or *weights*. The use of the empirical equations in this case helps create generalized methods of phase diagram calculations [5–7]. It is well understood that there is risk in using the relationships beyond the range of fitted data.

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 nonlinear function of  $x_j$ . The exact shape of the hyperbolic tangent can be varied by altering the weights (Fig. 1(a)). Further degrees of nonlinearity can be introduced by combining several of these hyperbolic tangents (Fig. 1(b)), so that the neural network method is able to capture almost arbitrarily nonlinear relationships.

Figure 2 illustrates the complexity of the surface that can be produced when representing the output  $y$  (vertical axis) as a function of two inputs  $x_1, x_2$  using just four hyperbolic tangents:

$$y = 0.8[\tanh\{8x_1 - 2\} + \tanh\{x_1^2 - 8\} + \tanh\{8x_2 + 2\} + \tanh\{x_2^2 - n\} + 1] \quad (2)$$

To summarize, a neural network is an explicit 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*. The function for a network with  $i$  hidden units, connecting the inputs  $x_j$  to the output  $y$  is given by

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

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

where  $w$  represents weights and  $\theta$  the constants as described in the context of linear regression. Modeling techniques

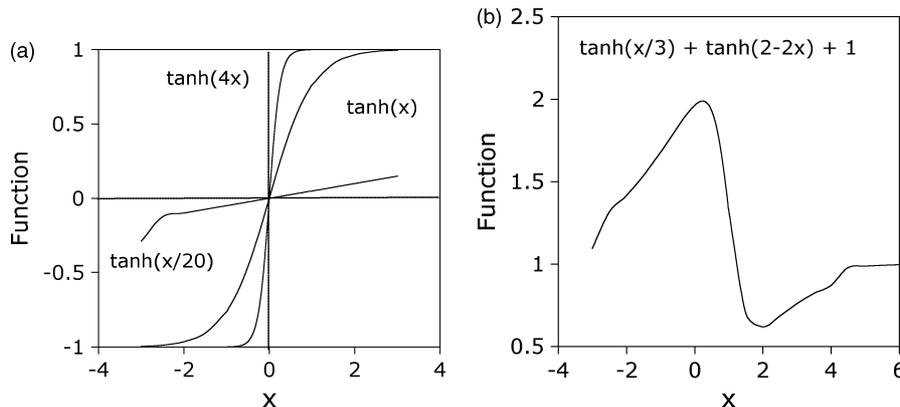


Fig. 1 (a) Three different hyperbolic functions; the ‘strength’ of each depends on the weights. The diagram illustrates the flexibility of a hyperbolic tangent. (b) A combination of two hyperbolic tangents to produce a more complex model. Such combinations can be continued indefinitely to produce functions of ever greater complexity.

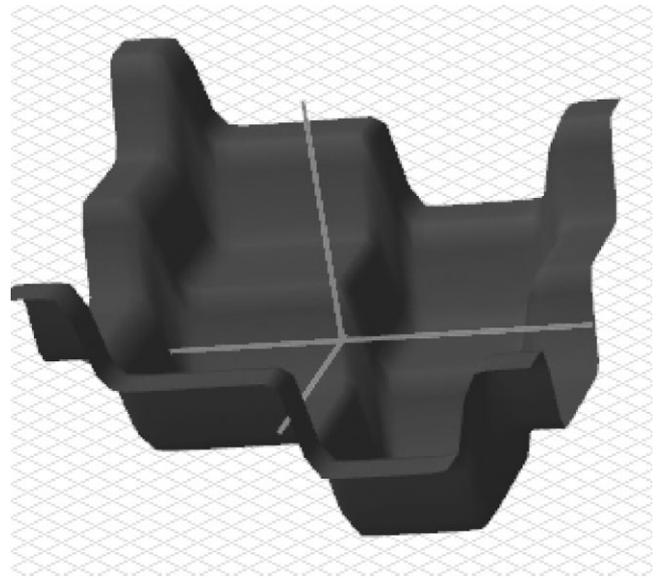


Fig. 2 Variation in the output (vertical axis) as a function of two input variables (horizontal axes), the whole surface being generated using just four hyperbolic tangent functions.

are sometimes classified as white, gray 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. Equation 3 is an explicit statement of the network and it is wrong to call such a network a black box [8] as is sometimes done [9].

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 visualize from an examination of the weights. This visualization 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.

A potential difficulty with the ability to produce complex, nonlinear functions is the possibility of overfitting of data. 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. 3 which shows three attempts at modeling noisy data for a case where  $y$  should vary with  $x^3$ . A linear model (Fig. 3(a)) is too simple and does not capture the real complexity in the data. An overcomplex function such as that illustrated in Fig. 3(c) accurately models the training data but generalizes badly. The optimum model is illustrated in Fig. 3(b). The training and test errors are shown schematically in Fig. 3(d); 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 generalizes best to unseen data. Additional methods for avoiding overfitting by penalizing overly complicated functions are described elsewhere [10–12].

### 3. UNCERTAINTIES

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 modeling*; there might 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,

$$E_D \propto \sum_j (t_j - y_j)^2. \tag{4}$$

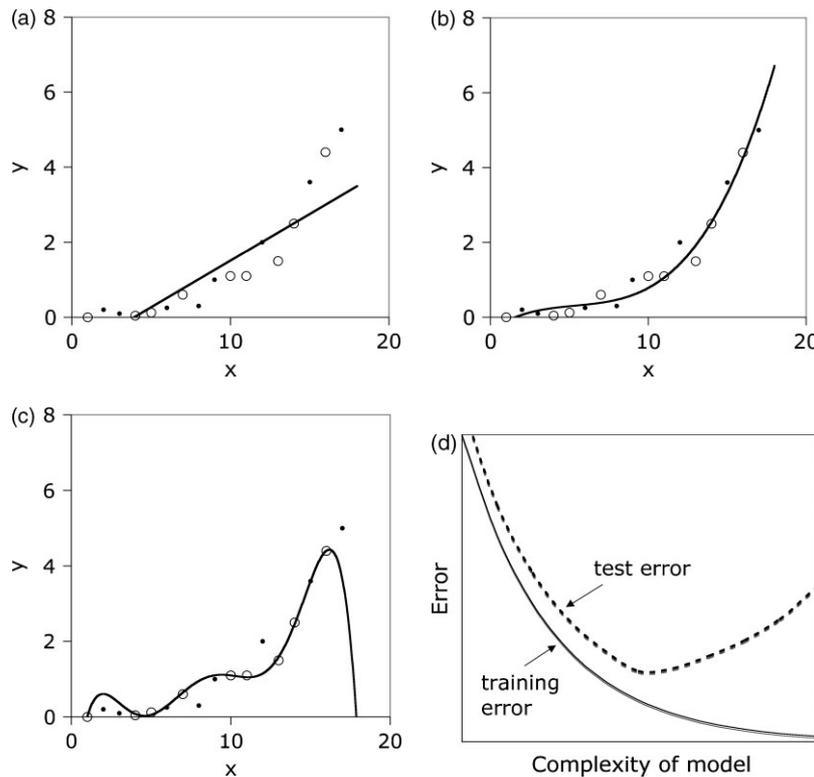


Fig. 3 Variations in the test and training errors as a function of model complexity, for noisy data in a case where  $y$  should vary with  $x^3$ . The large, open points were used to create the models (*i.e.* they represent training data), and the black dots constitute the test data. (a) A linear function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

$E_D$  is expected to increase if important input variables have been excluded from the analysis. An equivalent description is the standard error  $\sigma_{SE}$  given by

$$\sigma_{SE} = \sqrt{\sum_{j=1}^n \frac{(t_j - y_j)^2}{n^2}} \quad (5)$$

Figure 4(a) shows this kind of noise where data have been fitted to a straight line with a standard error of  $\pm 2$  in the estimation of the output  $y$ .

Whereas  $E_D$  and  $\sigma_{SE}$  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 modeling is illustrated in Fig. 4(b), where a set of precise data (2, 4, 6) are fitted to two different functions, one linear the other nonlinear [13]:

$$y = -\frac{x^3}{44} + \frac{3x^2}{11} + \frac{34}{11} \quad (6)$$

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 modeling, 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 modeling 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 nonexistent, Fig. 4(b).

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 modeling uncertainty.

Figure 5 illustrates the problem; the practice of using the best-fit function (*i.e.* the most probable values of the

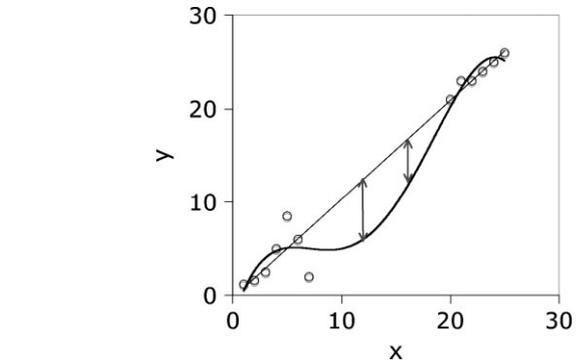
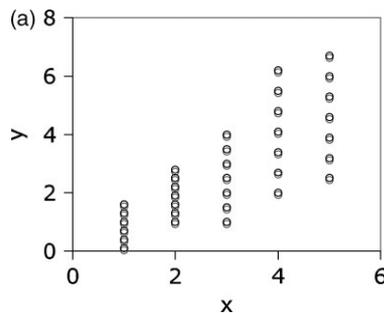


Fig. 5 Two different functions, each of which explains some 98% of the variation in the output  $y$ . And yet, they extrapolate differently in regions where the data are sparse or noisy. The difference between these functions is an indication of modeling uncertainty.

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 [10,11], the modeling uncertainty is expressed by not having a unique set of weights, but rather a probability distribution of sets of weights. This recognizes 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 modeling uncertainty then become large when data are sparse or locally noisy, as illustrated in Fig. 5.

This methodology has proved to be extremely useful in materials science [8] 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.

There has been controversy recently about whether published models are 'mathematically logical and determined', on the basis that inadequate quantities of data may have been used in creating sophisticated models [14–17].

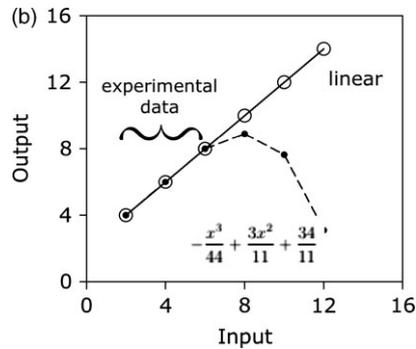


Fig. 4 (a) Noise. (b) Uncertainty.

It is important to realize that with neural network models that indicate modeling uncertainty, the question about how many data are needed to produce a satisfactory model becomes less relevant. Insufficient data are associated with large uncertainties, which the user must then assess to see whether the prediction is useful or not. Furthermore, in the Bayesian framework [10,11] there are other parameters which penalize excessive complexity, resulting automatically in models which are ‘as simple as possible, but not simpler’.

The argument that there must be more data than fitting points [14–17] needs qualification in the context of *empirical* analysis, *i.e.* where there are no principles which enable the actual relationship to be justified. It is suggested that with the two data, it would not be justified to fit anything other than a straight line (two fitting parameters). However, the relationship could equally well be a high-order polynomial with many more fitting parameters than the number of data. The fact that more than one relationship is possible is simply a reflection of the modeling uncertainty rather than a justification for using only two fitting parameters. Further experiments would be needed to confirm the right form but neither can be ruled out without access to additional knowledge such as the underlying physics.

One example where a network in a Bayesian framework has been used to model the transformation behavior of a stainless steel using just six data points (some of which could be regarded as being clustered) illustrates this discussion regarding the quantity of data [18]. Martensite tends to form when austenite in certain austenitic stainless steels is plastically deformed, but at large strains the transformation is suppressed by a phenomenon known as mechanical stabilization [19–29]. Plastic strain therefore first accelerates transformation and then retards it. Figure. 6 shows that with just six data, a neural network in a Bayesian framework with appropriate measures taken to avoid overfitting [18], has been able to capture correctly the relationship between martensite fraction and plastic strain. The form of the relationship is consistent with that expected from a dislocation-based theory of mechanical stabilization [18,26], in spite of the fact that the number of fitting parameters is very large indeed [18].

It is obvious from the discussion above that a failure to assess the modeling uncertainty and to test the trained network for generalization may lead to neural networks which overfit the data since the only measure available to fix the complexity of the network is the residue between the measured and calculated values. Mehta *et al.* [30] have attempted to avoid overfitting by adding noise to each existing data vector, thus artificially creating a larger dataset prior to training. The method requires an essentially subjective assessment of the level of noise that is appropriate.

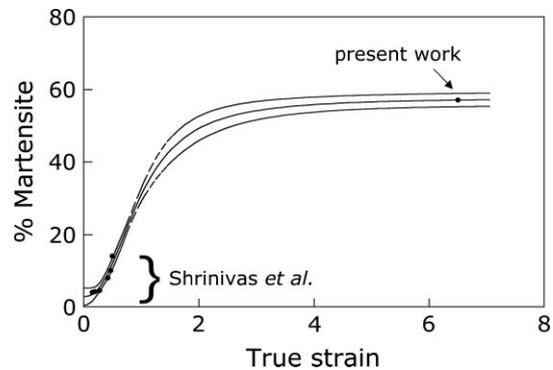


Fig. 6 Experimental data (points) and a neural network model (solid curve). The bounding curves represent  $\pm 1\sigma$  modeling uncertainties. For details see [[18]].

## 4. UNEXPECTED OUTCOMES

### 4.1. Welding Alloys

To discover something novel using a neural network is particularly rewarding because the method has its background in empiricism.

It is now some 62 years since it was recognized that ferritic steels have a ductile-brittle transition temperature beyond which the mode of fracture changes from cleavage to one capable of absorbing considerable energy during fracture [31,32]. It is a long-standing view that the addition of nickel to ferritic steel improves its toughness [33,34]. It has been suggested that nickel, in effect, enhances the mobility of dislocations in ferrite, making it more amenable to plastic deformation and hence avoiding cleavage [35].

The problem of ensuring adequate toughness is acute in the case of welding alloys where the properties have to be achieved in the cast-state. In spite of the perceived beneficial influence of nickel, early attempts at making manual metal arc weld metals containing large nickel concentrations failed to achieve toughness and the reasons for this were not understood [36].

However, research in high-nickel ferritic welds was stimulated with Muruganath and coworkers [37,38] who discovered, using a neural network model, that at large concentrations, nickel is only effective in increasing toughness when the manganese concentration is small. This is illustrated in Fig. 7(a) for tests done at  $-60^{\circ}\text{C}$ . Since this relationship had not been recognized previously, weld metals A (7Ni–2Mn), B (9Ni–2Mn) and C (7Ni–0.5Mn) were manufactured and tested; the results are illustrated in Fig. 7(b) and confirm the predictions.

The neural network does not of course indicate a mechanism for the degradation of toughness when both the nickel and manganese concentrations are large. However, characterization experiments revealed that a coarse phase (Fig. 8), previously unknown in welds, is induced to form in

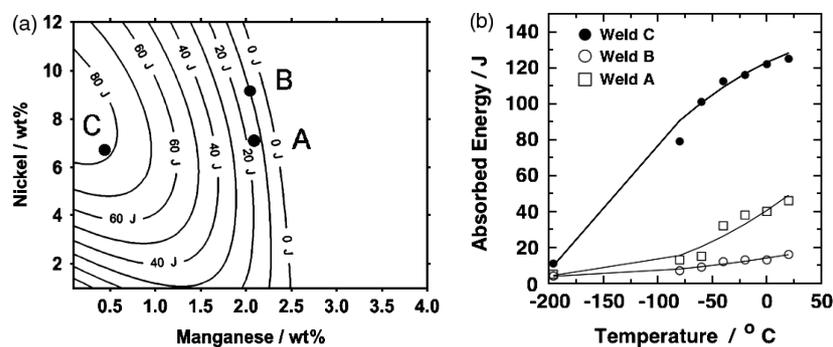


Fig. 7 (a) Combined effect of solutes on the calculated toughness for  $-60^{\circ}\text{C}$ , of weld metal produced using arc welding with a heat input of  $1\text{ kJ mm}^{-1}$ , with a base composition (wt%) 0.034 C, 0.25 Si, 0.008 S, 0.01 P, 0.5 Cr, 0.62 Mo, 0.011 V, 0.04 Cu, 0.038 O, 0.008 Ti, 0.025 N, and an interpass temperature of  $250^{\circ}\text{C}$ . (b) Full results for welds A, B, and C.

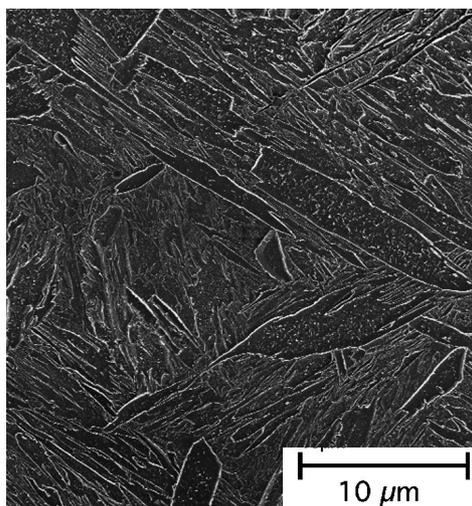


Fig. 8 The coarse plates are coalesced bainite in a 7Ni-2Mn wt% weld metal [39–41]. The coalescence process is visible in some cases, for example the large plate in the top left hand corner of the image.

these circumstances [39–42]. This *coalesced bainite* occurs when the transformation temperatures are suppressed by alloying such that there is only a small difference between the bainite and martensite-start temperatures.

Coalesced bainite is the consequence of the merger of crystallographically identical but adjacent platelets of bainite to form a single, larger plate [43]. The coalescence can only occur if there is sufficient driving force to sustain the greater strain energy associated with the coarser plate, and during the early stages of the transformation of austenite so that impingement between different crystallographic variants is minimized [43]. The first condition is satisfied by the large undercooling. The reason why the coalesced bainite is detrimental to toughness is because it presents a crystallographically homogeneous microstructure, which does not deflect the propagation of cleavage cracks. To summarize, the discovery of the coalesced bainite in weld metals

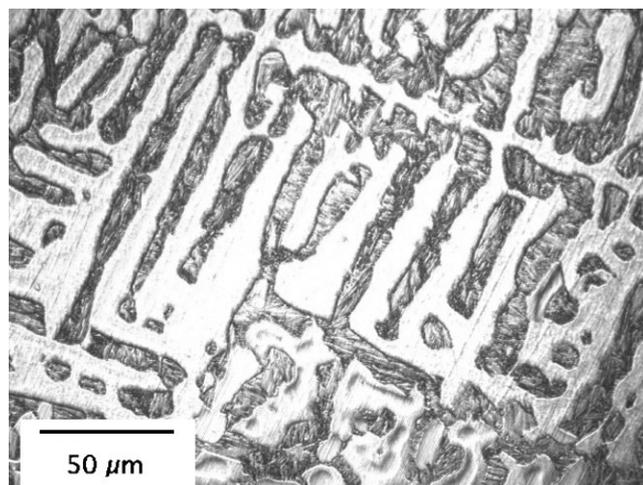


Fig. 9 The microstructure of  $\delta$ -TRIP steel [45].

is a direct and unexpected consequence of neural network modeling.

#### 4.2. Case 2: $\delta$ -TRIP and Genetic Algorithms

A design problem becomes interesting when the desired result can be achieved in many different ways. Genetic algorithms [44] can be combined with neural networks to search for domains of inputs which lead to a similar output. The method has been used to pose the following question: for a given set of properties, can a transformation-induced plasticity (TRIP)-assisted steel be designed with a minimum of silicon and a maximum of retained austenite?<sup>1</sup>

A combination of neural networks and genetic algorithms has in this way been used to design a novel low-silicon

<sup>1</sup> TRIP stands for transformation-induced plasticity. TRIP-assisted steels are strong and yet have a large uniform ductility because of the fact that retained austenite transforms into martensite during deformation.

TRIP-assisted steel with a strange microstructure consisting of  $\delta$ -ferrite dendrites and a mixture of bainitic ferrite and carbon-enriched retained austenite Fig. 9. The steel has been manufactured and tested to reveal a tensile strength of about 1 GPa and a uniform elongation of 23% [45].

## 5. EXPLANATION OF DIVERSE OBSERVATIONS

One of the major advantages of neural networks and similar nonlinear methods is that they have the ability to indicate the noise associated with the output when dealing with complicated problems.

Dilatometric techniques are used widely in measuring the onset of solid-state phase transformations in metals when they lead to a change in density, for example, [46–49]. When the transformation occurs during cooling, the highest temperature at which it is detected is known as the ‘start-temperature’. Such temperatures are frequently quoted without error bars.

The analysis of dilatometric data is subjective when precise measurements are necessary, for reasons discussed elsewhere [50]. To resolve this problem, a method has been developed which is fundamentally justified and is reproducible, the so-called *offset method*. The details of this are not relevant here, but in spite of the use of high-precision instruments, the martensite-start ( $M_S$ ) temperature was surprisingly found to vary in repeated experiments by as much as  $\sigma = \pm 12^\circ\text{C}$ .

When this was compared with published neural network models of the  $M_S$  temperature, created using vast experimental datasets [51–54], the perceived error was  $\sigma = \pm 15^\circ\text{C}$ . Although somewhat larger than that associated with the offset method [50], a greater error is expected in the neural models because the datasets they use have been compiled for diverse sources.

These comparisons are valuable because they suggest that the observed variations in the  $M_S$  temperatures are not due to instrumental problems or alloy homogeneity [50]. Possible explanations, which remain to be explored, include nonuniform temperatures within the dilatometer samples which are of finite size. More fundamentally, the early stages of martensitic transformation are sensitive to the initial austenite grain structure [55–57]. There is no guarantee that identical grain structures are generated in every sample. Martensitic nucleation can occur from arrays of dislocations associated with grain boundaries and therefore may be sensitive to the nature of those boundaries [58].

## 6. HYBRID MODELS

There are many kinds of mathematical models developed in the context of materials science, including *ab initio*

electron theory, molecular dynamics, thermodynamics, fluid dynamics, kinetics, finite element modeling, genetic algorithms *etc.* It is possible that benefits accrue when neural networks are combined with one or more of these methods.

There is sophisticated theory available for predicting the steady-state shape of the pool of liquid metal that forms during welding. Included in the calculations is heat and fluid flow, Marangoni effects, gravitational and electromagnetic effects; the transfer of mass from the welding electrode to the pool may also be considered [59,60]. The calculations are time-consuming but have been used to form the inputs of a neural network model [61,62]. Apart from the practical advantage of economy in computation times, once a trained set of networks is created, it is claimed that such a network should be robust since it is based on computational data, which satisfy the basic law of mass, momentum and energy conservation. Whereas this might be an advantage, it is probable that these basic laws provide an incomplete description of weld pool phenomena. For example, it is common practice when dealing with turbulence to simply increase the viscosity and using this value in streamlined fluid flow calculations. The effective viscosity is determined by matching with experimental data [63]. For these reasons, it remains to be demonstrated whether a neural network model based on the outputs of other models are fundamentally better than directly training the network on experimental data. In particular, it is not clear how a neural network trained on the output of physical models actually captures the principles of the physical models and more tests need to be done to study the long-range extrapolation behavior of such models.

Attempts have also been made to incorporate the outputs of thermodynamic models as additional data in the creation of neural networks on the basis of experimental datasets. Examples include work in the context of creep deformation [38,64] and phase transformation theory [65,66].

## 7. EXPRESSION OF DATA AND CONTROL ALGORITHMS

Neural networks are a convenient way of quantitatively representing complex data and the resulting mathematical functions from a trained network can be evaluated very rapidly, typically within milliseconds of computing time. This makes trained networks extremely useful in control algorithms. For example, the rolling process in the production of steel plates involves a very large number of variables including the chemical composition, the dimensions and temperature of the slabs when they enter the mill, the rolling reduction, the time between passes, the temperature at each stage of the process and the final coiling conditions. In one case, the neural network model created to calculate the

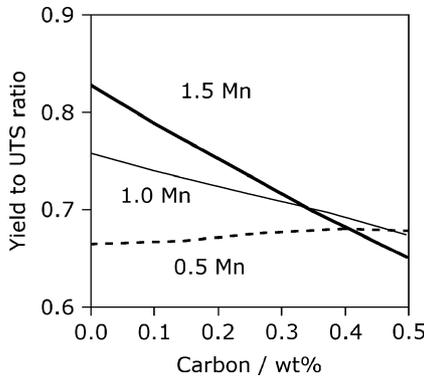


Fig. 10 The yield to ultimate tensile strength ratio as a function of the carbon and manganese concentrations of hot-rolled steels [67].

tensile properties of the steel following hot-rolling included 108 variables [67]. It is an amazing achievement that models like these are now used for the on-line control and automation of entire rolling mills [68–71].

Whereas the primary purpose of this particular kind of work is in control engineering, the analysis is complex and nonlinear, which means that unexpected and elegant relationships may emerge which are not obvious from limited experiments. For example, Fig. 10 shows how the yield to ultimate tensile strength ratio varies as a function of the key solutes, carbon and manganese. A low ratio is essential in safe design against calamities such as earthquakes, because it is necessary then to sustain a lot of plasticity prior to fracture during dynamical loading [72,73].

The trends illustrated in Fig. 10 have in hindsight been metallurgically interpreted, in terms of the fraction of pearlite in the microstructure [67].

## 8. SOME QUIRKS

### 8.1. Bounding the Output

A neural network like any other empirical method carries the risk of making predictions, which are physically impossible because neither its inputs nor its outputs are in general bounded. The fraction  $\psi$  of a phase can only take continuous values between 0 and 1; to cope with this, Yescas *et al.* [74,75] used a double logarithm function to ensure  $0 \leq \psi \leq 1$ . This form is consistent with the theory of solid-state transformations [76–78] with  $\psi = 1 - \exp\{-kt^n\}$  where  $k$  and  $n$  are constants and  $t$  is the time. It follows that  $\ln\{-\ln(1 - \psi)\}$  should vary with  $n \ln t$ . It turns out that it is only when  $\ln t$  is used as an input and the double logarithm as the output, rather than the raw fraction, that the variation in the latter can be correctly captured by the neural network [74,75]. In another context, Sourmail *et al.* in their work on the elevated temperature properties of steels

modeled the logarithm of creep rupture life rather than the life itself [79] on the grounds that in a plot of stress versus life, the life extends dramatically as the stress is decreased towards zero. The life should be infinite at zero stress, but it clearly cannot be negative and this is consistent with the use of a logarithm of the life.

It is unlikely, however, that such methods can be applied generally when the logarithmic form is not physically justified. Consider the single logarithm which confines the output  $y'$  to be greater than zero in the function  $\ln\{(y - y_{min}) / (y_{max} - y_{min})\}$  where  $y_{max}$  and  $y_{min}$  are the maximum and minimum values of  $y$ . There arises a difficulty in setting the value of  $y_{max}$ , which in principle can take any value in excess of the maximum in the training set. Figure 11 shows how two different choices for  $y_{max}$  give dramatically different predictions of the Charpy properties when the model is deliberately extrapolated over a long range. The reason why, in the second case, the extrapolation ends up at about zero for high temperatures (Fig. 11(b)) is that the training of the neural network is confined to values much less than  $y_{max}$  so the model perceives that most values should be confined to the lower end of the scale. In contrast, setting  $y_{max}$  to a value close to that in the experimental set allows the curve to increase slowly towards  $y_{max}$ , but this

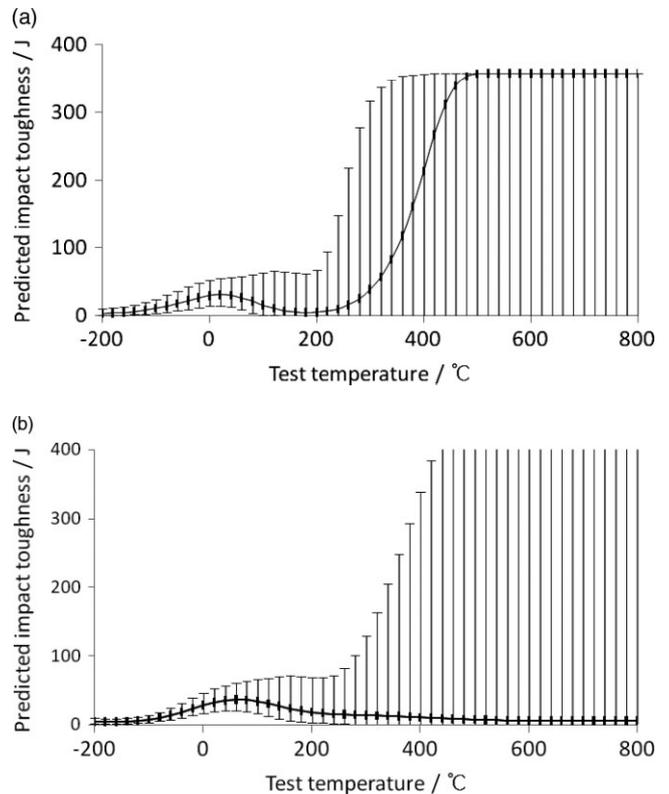


Fig. 11 Extrapolation of models in which the toughness (output) was modeled as a single logarithmic function  $\ln\{(y - 0) / (y_{max} - 0)\}$ , with (a)  $y_{max} = 357$  J and (b)  $y_{max} = 3570$  J, [80].

increase cannot then be extrapolated indefinitely and hence the saturation.

## 9. SIMPLIFICATION

The design of steels for the fusion reactor programme requires predictions to be made well-beyond the range of current knowledge and the experimental facilities needed to validate the calculations will not be available for more than a decade. There are therefore efforts to estimate the influence of large doses of energetic neutrons on the properties of candidate steels [81–84]. Most of these efforts include as many variables in the analysis as the availability of the data will permit, on the grounds that Bayesian methods are robust enough to minimize the influence of irrelevant variables, thus avoiding the risk of preconceived bias.

A recent analysis [85] has taken a different approach aimed at ‘dimensionality reduction’, i.e. simplification of the problem prior to creating models. Steels contain a myriad of solutes and an attempt was made to use linear combinations of solutes to reduce the number of input variables. The procedure involved keeping aside high dose-per-atom data in order to assess the behavior of networks created in this way. The residue between the calculated values of these test data and actual measurements was used as a measure of success.

The physical significance of the linear combination of solute concentrations is not clear, and perhaps a better assessment of the model would involve the study of a large number of trends in the output as a function of the inputs, including large extrapolations; a study of the residue alone is not sufficient to decide whether the technique has achieved better performance.

## REFERENCES

- [1] J. M. Ziman, *Models of Disorder*, Cambridge, Cambridge University Press, 1979.
- [2] M. Kranzberg and C. S. Smith, *Mater Sci Eng* 37 (1979), 1–39.
- [3] R. S. Claassen and A. G. Chynoweth, *Mater Sci Eng* 37 (1979), 41–102.
- [4] H. K. D. H. Bhadeshia, *Mater Sci Technol* 24 (2008), 128–135.
- [5] M. Hillert, Prediction of iron base phase diagrams, in *Hardenability Concepts with Applications to Steels*, D. V. Doane and J. S. Krikaldy (eds), Warrendale, PA, The Metallurgical Society of AIME, 1977, 5–27.
- [6] T. G. Chart, J. F. Counsell, G. P. Jones, W. Slough, and P. J. Spencer, *Int Met Rev* 20 (1975), 57–82.
- [7] H. K. D. H. Bhadeshia, in *Introduction to Materials Modelling*, Z. Barber (ed.), Thermodynamics: Maney, London, 2006, Ch. 4, 45–72.
- [8] H. K. D. H. Bhadeshia, *ISIJ Int* 39 (1999), 966–979.
- [9] M. Militzer, *ISIJ Int* 47 (2007), 1–15.
- [10] D. J. C. MacKay, *Neural Comput* 4 (1992), 448–472.
- [11] D. J. C. MacKay, *Neural Comput* 4 (1992), 415–447.
- [12] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge, UK, Cambridge University Press, 2003.
- [13] D. J. C. MacKay, *Darwin College Magazine*, 1993, 81–85.
- [14] W. Sha, *J Mater Process Technol* 171 (2006), 283–285.
- [15] W. Sha and K. L. Edwards, *Mater Des* 28 (2007), 1747–1752.
- [16] W. Sha, *J Hydrol* 340 (2007), 119–121.
- [17] W. Sha, *Geosci Remote Sens, IEEE Trans* 45 (2007), 1896–1897.
- [18] H. S. Wang, J. R. Yang, and H. K. D. H. Bhadeshia, *Mater Sci Technol* 21 (2005), 1323–1328, Software: <http://www.msm.cam.ac.uk/phase-trans/2005/thread.html>.
- [19] J. F. Breedis and W. D. Robertson, *Acta Metall* 11 (1963), 547–559.
- [20] R. Lagneborg, *Acta Metall* 12 (1964), 823–843.
- [21] M. J. C. J. R. Strife and G. S. Ansell, *Metall Trans A* 8A (1977), 1471–1484.
- [22] K. Tsuzaki, S. Fukasaku, Y. Tomota, and T. Maki, *Trans JIM* 32 (1991), 222–228.
- [23] C. Maier, O. Blaschko, and W. Pichl, *Phys Rev B* 52 (1995), 9283–9290.
- [24] J. R. Yang, C. Y. Huang, W. H. Hsieh, and C. S. Chiou, *Mater Trans JIM* 37 (1996), 579–585.
- [25] X. Song, N. Gu, and H. Peng, *Defect Diffus Forum* 148–149 (1997), 165–167.
- [26] S. Chatterjee, H. S. Wang, J. R. Yang, and H. K. D. H. Bhadeshia, *Mater Sci Technol* 22 (2006), 641–644.
- [27] M. C. Tsai, C. S. Chiou, J. S. Du, and J. R. Yang, *Mat Sci Eng A* A332 (2002), 1–10.
- [28] A. Matsuzaki, H. K. D. H. Bhadeshia, and H. Harada, *Acta Metall Mater* 42 (1994), 1081–1090.
- [29] H. K. D. H. Bhadeshia, *Mater Sci Eng A* 378A (2004), 34–39.
- [30] R. Mehta, S. S. Sahay, A. Datta, and A. Chodha, *Mat Manuf Processes* 23 (2008), 204–209.
- [31] C. F. Tipper, *The Brittle Fracture Story*, Cambridge, Cambridge University Press, 1962.
- [32] A. H. Cottrell, *Int J Press Vessels Piping* 64 (1995), 171–174.
- [33] W. Jolley, *J Iron Steel Inst* 206 (1968), 170–172.
- [34] L. A. Norström and O. Vingsbo, *Met Sci* 13 (1979), 677–684.
- [35] W. C. Leslie, *The Physical Metallurgy of Steels*, New York, McGraw Hill, 1982.
- [36] M. Lord, Design and modelling of ultra-high strength steel weld deposits, PhD thesis, University of Cambridge, 1998.
- [37] M. Muruganath, H. K. D. H. Bhadeshia, E. Keehan, H. O. Andrén, and L. Karlsson, Strong and tough steel welds, in *Mathematical Modelling of Weld Phenomena 6*, H. Cerjak, H. K. D. H. Bhadeshia (eds), London, UK, Institute of Materials, 2002, 205–230.
- [38] M. Muruganath, Design of welding alloys for creep and toughness, PhD thesis, University of Cambridge, Cambridge, 2002.
- [39] E. Keehan, L. Karlsson, and H.-O. Andrén, *Sci Technol Weld Joining* 11 (2006), 1–8.
- [40] E. Keehan, L. Karlsson, H.-O. Andrén, and H. K. D. H. Bhadeshia, *Sci Technol Weld Joining* 11 (2006), 9–18.
- [41] E. Keehan, L. Karlsson, H.-O. Andrén, and H. K. D. H. Bhadeshia, *Sci Technol Weld Joining* 11 (2006), 19–24.
- [42] H. K. D. H. Bhadeshia, E. Keehan, L. Karlsson, and H. O. Andrén, *Trans Indian Inst Met* 59 (2006), 689–694.

- [43] L. C. Chang, and H. K. D. H. Bhadeshia, *Mater Sci Technol* 12 (1996), 233–236.
- [44] N. Chakraborti, *Int Mater Rev* 49 (2004), 246–260.
- [45] S. Chatterjee, M. Muruganath, and H. K. D. H. Bhadeshia, *Mater Sci Technol* 23 (2007), 819–827.
- [46] M. Takahashi and H. K. D. H. Bhadeshia, *J Mater Sci Lett* 8 (1989), 477–478.
- [47] G. K. Prior, *Mater Forum* 18 (1994), 265–276.
- [48] M. Onink, F. D. Tichelaar, C. M. Brackman, E. J. Mittemeijer, and S. van der Zwaag, *Z Metallkunde* 87 (1996), 1–32.
- [49] J. Z. Zhao, C. Mesplont, and B. D. de Cooman, *Mater Sci Eng A A332* (2002), 110–116.
- [50] H.-S. Yang and H. K. D. H. Bhadeshia, *Mater Sci Technol* 23 (2007), 556–560.
- [51] C. Capdevila, F. G. Caballero, and C. G. de Andrés, *ISIJ Int* 42 (2002), 894–902.
- [52] C. Capdevila, F. G. Caballero, and C. G. de Andrés, *Mater Sci Technol* 19 (2003), 581–586.
- [53] J. Wang, P. J. van der Wolk, and S. van der Zwaag, *Mater Trans JIM* 41 (2002), 761–768.
- [54] A. P. d. W. W. G. Vermeulen, P. F. Morris, and S. van der Zwaag, *Ironmaking Steelmaking* 23 (1996), 433–437.
- [55] O. A. Ankara, A. S. Sastri, and D. R. F. West, *JISI* 204 (1966), 509–511.
- [56] T. Maki, S. Shimooka, and I. Tamura, *Metall Trans* 2 (1971), 2944–2955.
- [57] P. J. Brofman and G. S. Ansell, *Metall Trans A* 14A (1983), 1929–1931.
- [58] G. B. Olson and M. Cohen, *Metall Trans A* 7A (1976), 1897–1923.
- [59] T. DebRoy and S. A. David, *Rev Mod Phys* 67 (1995), 85–112.
- [60] T. Zacharia, J. M. Vitek, J. A. Goldak, T. A. DebRoy, M. Rappaz, and H. K. D. H. Bhadeshia, *Model Simul Mater Sci Eng* 3 (1995), 265–288.
- [61] A. Kumar and T. DebRoy, *Sci Technol Weld Joining* 11 (2006), 106–119.
- [62] S. Mishra and T. DebRoy, *Mater Sci Eng A* 454 (2007), 477–486.
- [63] A. De and T. DebRoy, *Sci Technol Weld Joining* 11 (2006), 143–153.
- [64] B. Meyer and H. Gutte, *Steel Res* 72 (2001), 361–385.
- [65] C. Garcia-Mateo, T. Sourmail, F. G. Caballero, and C. G. de Andrés, *Mater Sci Technol* 21 (2005), 934–940.
- [66] S. B. Singh and H. K. D. H. Bhadeshia, *Mater Sci Eng A A245* (1998), 72–79.
- [67] S. B. Singh, H. K. D. H. Bhadeshia, D. J. C. MacKay, H. Carey, and I. Martin, *Ironmaking Steelmaking* 25 (1998), 355–365.
- [68] B. Ortman, *Stahl und Eisen* 115 (1995), 35–40.
- [69] J. Larkiola, P. Myllykoski, A. S. Korhonen, and L. Cser, *J Mater Process Technol* 80–81 (1998), 16–23.
- [70] N. B. Skorokhvatov, *Stal* 5 (2003), 42–45.
- [71] L. E. Zárate and F. R. Bittencout, *J Mater Process Technol* 197 (2008), 344–362.
- [72] A. Elnashai, Design of low-rise steel structures subjected to accidental dynamical loading, in *Building in Steel*, London, UK, Insitute of Metals, 1991, 25–32.
- [73] M. Toyoda, *Welding J* 74 (1995), 31–42.
- [74] M. Yescas, H. K. D. H. Bhadeshia, and D. J. C. MacKay, *Mater Sci Eng A* 311 (2001), 162–173.
- [75] M. Y. Gonzalez and H. K. D. H. Bhadeshia, *Mater Sci Eng A* 333 (2002), 60–66.
- [76] M. Avrami, *J Chem Phys* 7 (1939), 1103–1112.
- [77] M. Avrami, *J Chem Phys* 8 (1940), 212–224.
- [78] M. Avrami, *J Chem Phys* 9 (1941), 177–184.
- [79] T. Sourmail, H. K. D. H. Bhadeshia, and D. J. C. MacKay, *Mater Sci Technol* 18 (2002), 655–663.
- [80] J.-H. Pak, J.-H. Jang, H. K. D. H. Bhadeshia, and L. Karlsson, *Mater Manuf Processes* (2009) (in press).
- [81] R. Kemp, G. A. Cottrell, H. K. D. H. Bhadeshia, G. R. Odette, T. Yamamoto, and H. Kishimoto, *J Nucl Mater* 348 (2006), 311–328.
- [82] G. A. Cottrell, R. Kemp, H. K. D. H. Bhadeshia, G. R. Odette, and T. Yamamoto, *J Nucl Mater* 367–370 (2007), 603–609.
- [83] G. A. Cottrell, R. Kemp, and H. K. D. H. Bhadeshia, *J Nucl Mater* 367–370 (2007), 1586–1589.
- [84] J. A. Wang and N. S. Rao, *J Nucl Mater* 301 (2002), 193–202.
- [85] G. C. C. G. Windsor and R. Kemp, *Model Simul Mater Sci Eng* 16 (2008), 1–14.