The Charpy Impact Transition Temperature for some Ferritic Steel Welds

S. H. Lalam, H. K. D. H. Bhadeshia and D. J. C. MacKay†

University of Cambridge

Department of Materials Science and Metallurgy

Pembroke Street, Cambridge CB2 3QZ, U.K.

†Cavendish Laboratory

Madingley Road, Cambridge CB3 0HE, U.K.

ABSTRACT

Body–centered cubic iron undergoes a ductile–brittle fracture transition as a function of temperature. A common way of describing the Charpy toughness is to measure the temperature T_J corresponding to a particular value of absorbed energy. Extensive data on variations in T_J as a function of the microstructure and weld metal composition have recently been published along with a linear regression analysis. In the present work we show that it is possible to infer more meaning from the data by using a neural network (non–linear regression) analysis.

INTRODUCTION

The Charpy toughness of a steel weld is one of the important quality control parameters, widely specified in industry and used as a ranking parameter in consumable research and development programmes. Body–centered cubic iron undergoes a ductile–brittle transition as the test temperature is reduced. Consistent with international norms, the toughness is therefore frequently characterised by a transition temperature corresponding to a particular value of the absorbed impact energy. In a recent paper, French [1] conducted a careful series of experiments in which the temperature T_{27J} corresponding to a measured Charpy impact energy of 27 J was characterised as a function of the yield strength, oxygen content and the microstructure. The latter included the fraction of acicular ferrite in the as–deposited microstructure, but since the work was done on multipass welds, an overall percentage of reheated microstructure was also measured. Three different welding processes were used: flux–cored arc welding (FCAW), gas metal arc welding (GMAW) and manual metal arc welding (MMAW).

The resulting data were analysed using linear regression as follows:

$$T_{27J} = 0.007(YS) + 550(O) + 0.034(R) - 0.31(AF) - 74 \qquad ^\circ\mathrm{C} \tag{1}$$

where YS is the yield strength in MPa, (O) is the concentration of oxygen in wt%, and the reheated microstructure (R) and acicular ferrite (AF) are as area percentages. The range of applicability of the equation can be gauged from Table 1, which contains information from 59 separate measurements.

The analysis indicated a standard error of $\pm 12\,^\circ$C, with a correlation coefficient of 0.78. It is possible that a better interpretation of the data and associated uncertainties can be obtained using a non–linear regression method, which does not have an *a priori* assumption of the relationship between the variables, which accounts for the interactions between the variables, and which comments not only in the perceived level of noise in the output, but also on how the uncertainty of fitting depends on the particular region of input space where the prediction is being made. We begin with a brief introduction to the method of neural network analysis [2,3].

Input element	Minimum	Maximum	Mean	Standard Deviation
Yield Strength (MPa)	360	630	516	55
Oxygen (wt%)	0.03	0.12	0.06	0.02
Reheated Material (%)	20	79	41	13
Acicular Ferrite (%)	5	86	54	15
Temperature at 27J ($^\circ$C)	-88.0	-13	-54	18

Table 1: Characteristics of the measured parameters in the experiments conducted by French [1].

THE METHOD

A neural network is a general method of regression analysis in which a flexible non–linear function is fitted to experimental data, the details of which have been extensively reviewed [2]. It is, nevertheless, useful to present some salient features in order to place the technique in context.

The flexibility of the non–linear function scales with the number of hidden nodes i. Thus, the dependent variable y is given in the present work by

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

where

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

where x_j are the j variables on which the output y depends, w_i are the weights (coefficients) and θ_i are the biases (equivalent to the constants in linear regression analysis). The combination of equation 2 with a set of weights, biases, value of i and the minimum and maximum values of the input variables defines the network completely. Notice that the complexity of the function is related to the number of hidden units. The availability of a sufficiently complex and flexible function means that the analysis is not as restricted as in linear regression where the form of the equation has to be defined explicitly before the analysis.

The neural network can capture interactions between the inputs because the hidden units are nonlinear. The nature of these interactions is implicit in the values of the weights, but the weights may not always be easy to interpret. 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. A better method is to actually use the network to make predictions and to see how these depend on various combinations of inputs.

Error Estimates

The input parameters are generally assumed in the analysis to be precise and it is normal to calculate an overall error 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 analysis. Whereas E_D gives an overall perceived level of noise in the output parameter, it is, on its own, an unsatisfying description of the uncertainties of prediction.

MacKay has developed a particularly useful treatment of neural networks in a Bayesian framework [2], 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 [3].

The error bars presented throughout this work therefore represent a combination of the perceived level of noise in the output (T_{27J}) and the fitting uncertainty as described above.

Overfitting

A potential difficulty with the use of powerful non–linear regression methods is the possibility of overfitting data. To avoid this, 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. 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 on unseen data [2].

The discussion of overfitting is rather brief because the problem does not 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].

THE ANALYSIS

The aim of the neural network in this case was to predict T_{27J} as a function of the variables shown in Table 1. All the input variables and the output were normalised within the range ± 0.5 as follows:

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

where x is the original value from the database, x_{max} and x_{min} are the respective maximum and minimum of each variable in the original data and x_N is the normalised value. This step is not essential to the running of the neural network but later allows a convenient way to compare the results of the output.

For several runs of the neural network, Fig. 1 shows the model perceived noise σ_ν in T_{27J}. It is very interesting that the level of noise in the normalised output parameter T_{27J}, as perceived by the network, is $\sim 0.15 - 0.18$. This amounts to $\pm 11 - 14\,^\circ$C, which compares favorably with the $\pm 12^\circ$C deduced in by French using linear regression analysis. It is also worth noting that the error, irrespective of the model, is quite large when considering the physical meaning of T_{27J}. Furthermore, one standard error corresponds to a 68% confidence limit whereas two standard errors give the more acceptable 95% error bound. The important point is that the noise level is not reduced by using a non–linear analysis, giving evidence that the problem is not well specified; there are missing variables which clearly affect the toughness. We shall not speculate on what these missing variables could be, but factors such as the hydrogen and nitrogen concentrations, the scale of the microstructure *etc.* come to mind. Note also that the nature of the welding process is not explicitly taken into account.

Fig. 2 shows the predictions for the training and test data for the best model identified as the one with the highest log predictive error [2]. It is clear that the model is reasonably well behaved in the sense that the test data are predicted to a similar level of accuracy as the training data. It is important to note that the error bars plotted in Fig. 2a,b do not include σ_ν, but only the fitting error which depends on the position in the input space. Fig. 2c, shows the corresponding plot for the test data where the error bars contain both the σ_ν and the fitting error. All subsequent plots also include both components since it is logical to consider both the

perceived level of noise in the output and the fitting error. As will be seen subsequently, the latter is particularly important when extrapolating or interpolating, since large fitting errors are calculated in regions where the experimental knowledge is sparse or noisy.

It is possible that a committee of models can make a more reliable prediction than an individual model [2]. The best models are ranked using the values of the test errors. Committees are then formed by combining the predictions of the best L models, where $L = 1, 2, \ldots$; the size of the committee is therefore given by the value of L. A plot of the test error of the committee versus its size gives a minimum which defines the optimum size of the committee, as shown in Fig. 3.

The test error associated with the best single model is clearly greater than that of any of the other committees. It was determined in this case that a committee of thirteen models would be the best choice, being the committee of the lowest test error. The committee was then retrained on the entire data set without changing the complexity of any of its members.

The predictions of the committee trained on the entire data set can be compared with the original dataset as shown in Fig. 4.

Another parameter, σ_w, indicates the importance of an input in terms of its variation having an effect on the output of the model. Fig. 5 compares the values of σ_w for each of the inputs for the thirteen models in committee. A high value of σ_w for a specific input can be caused by the corresponding variable inducing a large variation in the output, but it can be seen from Fig. 5 that different models can assign varying significance to the same input. This is one of the reasons why a committee of models can be more reliable than the single model judged to be best on the basis of a parameter such as σ_ν.

USE OF THE MODEL

It is worth illustrating a few predictions, to emphasise the point that the error bars will not be constant as in [1]. It is important to note that as in equation 1, the predictions are for the case where just one input variable is altered, keeping all other fixed. This may not be possible when conducting experiments, the variables used for analysis were shown in Table 2. Fig. 6a shows that T_{27J} increases with the oxygen concentration; this is expected since the oxygen is inevitably present in the form of oxide inclusions which, for a constant microstructure, are detrimental to toughness.

It is not surprising that Fig. 6b shows that acicular ferrite improves the toughness. However, the neural network model shows that the results are not certain at large fractions of acicular ferrite when all the other variables are kept constant.

Input element	
Yield Strength (MPa)	516
Oxygen (wt%)	0.042
Reheated Material (%)	40
Acicular Ferrite (%)	63

Table 2: Input parameters used for the predictions. These correspond to a FCAW weld studied in [1].

Fig. 7 shows contour plots of T_{27J} as a function of the acicular ferrite and oxygen concentrations. A simple interpretation of the linear regression model (Fig. 7b) indicates that for optimum toughness, the acicular ferrite must be maximised at a zero oxygen concentration. However, there are no weld in the dataset with zero oxygen concentration and such a suggestion is probably not justified since oxides are needed to nucleate acicular ferrite. The neural network analysis, on the other hand, correctly indicates an optimum combination of acicular ferrite and oxygen concentration.

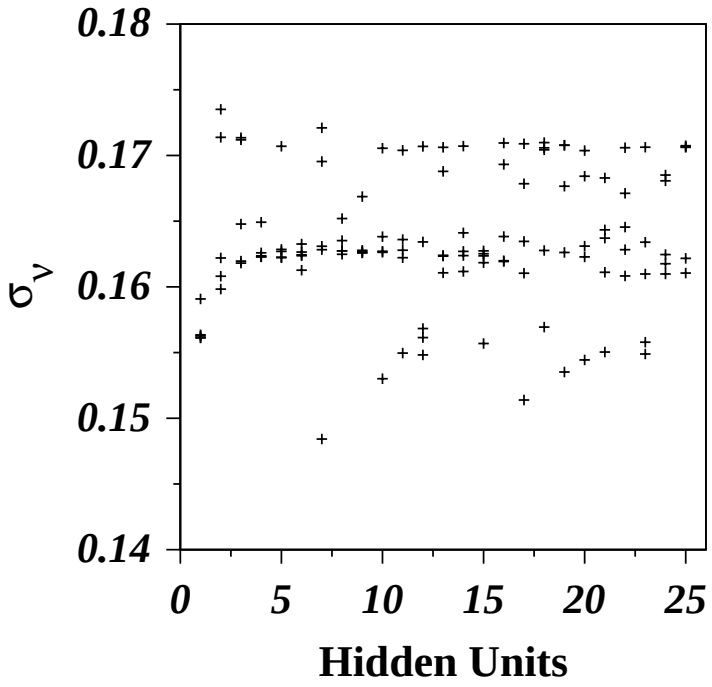

Fig. 1 Variation in σ_ν as a function of the number of hidden units. Several values are presented for each set of hidden units because the training for each network started with a variety of random seeds.

SUMMARY

An important conclusion from this work is that the use of non–linear regression analysis in the form of a neural network does not reduce the rather large perceived level of noise in the measured values of T_{27J}. This is expected with hindsight, since there are many more variables which control toughness when compared with the restricted set studied.

The second conclusion is that the standard error quoted for the linear regression model must be regarded as an underestimate of the real uncertainty, since there will be regions of the input space where the fitting function itself has great uncertainty. This is relevant in both extrapolation and interpolation.

Finally, even though the non–linear model does not help in reducing the perceived noise in the output, it is clear that the dependence of T_{27J} on a particular variable is a function of all the other input parameters. Therefore, unlike linear regression analysis, the neural network correctly predicts that there is a combination of acicular ferrite and oxygen which optimises toughness.

In a further comparison between neural networks and linear regression analysis, the latter it becomes clear that has the advantage of simplicity. However, neural network calculations can be done easily on a popular computer. The software capable of doing these calculations can be obtained freely from

http : *//www.msm.cam.ac.uk/map/mapmain.html*

REFERENCES

1. French, I. E. : *Australasian Welding Journal* **44** (1999) second quarter, 44–46.
2. MacKay, D. J. C.: *Mathematical Modelling of Weld Phenomena 3, Ed. by H. Cerjack* (1997) 359–389.
3. Bhadeshia, H. K. D. H.: *ISIJ International* **39** (1999) 10, 966–979.

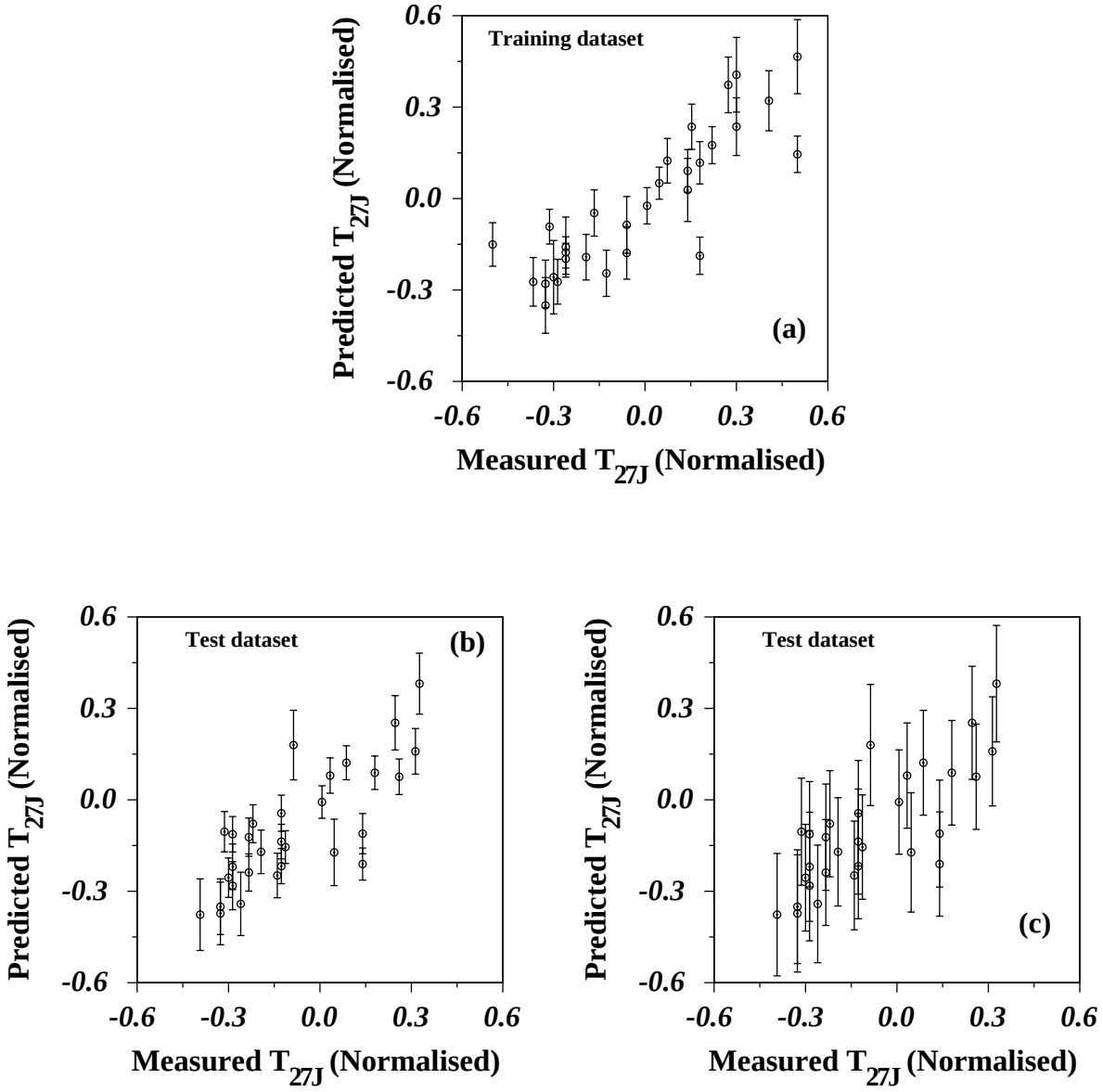

Fig. 2 Comparison of the predictions made using the best model and measured values of T_{27J}, (a) training data plotted with the fitting error, (b) test data plotted with the fitting error, (c) test data with the error bars representing both the fitting error and σ_ν.

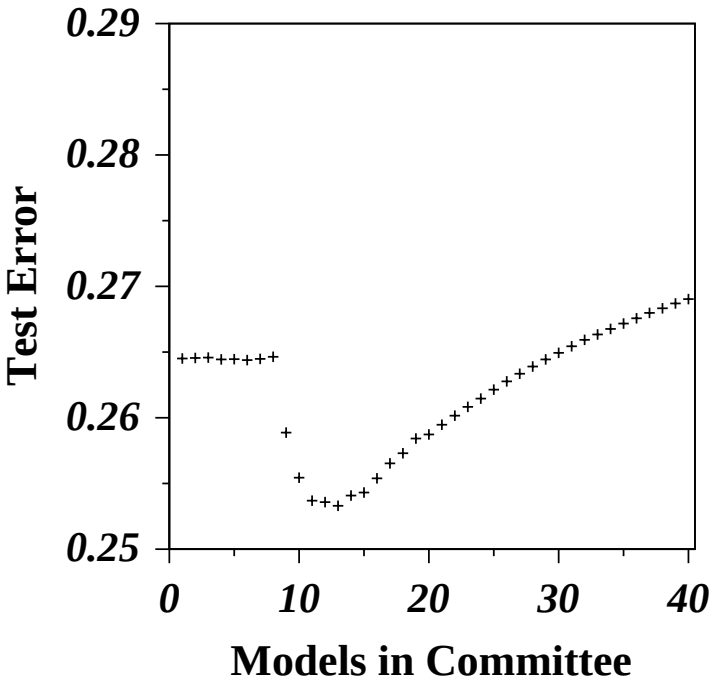

Fig. 3 Comparison of test error of increasing size of committees.

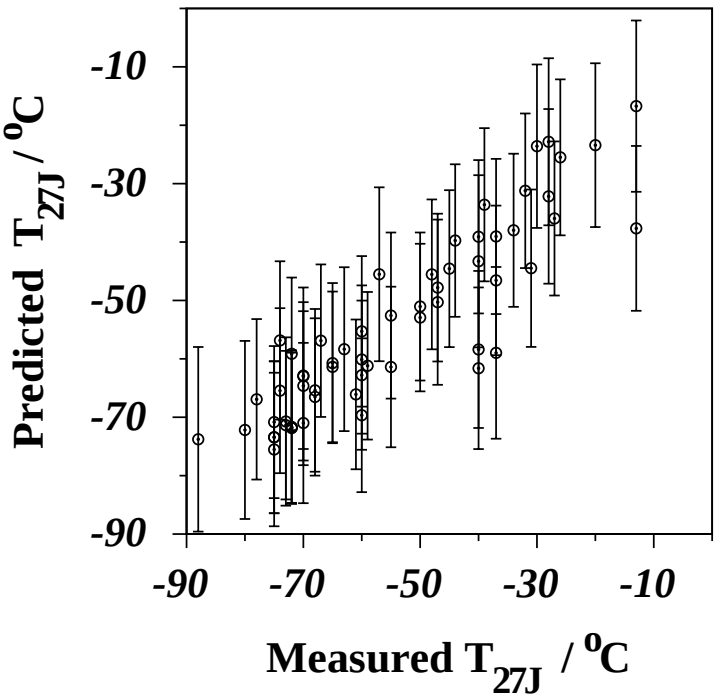

Fig. 4 Comparison of predicted values and experimental values for the optimum committee.

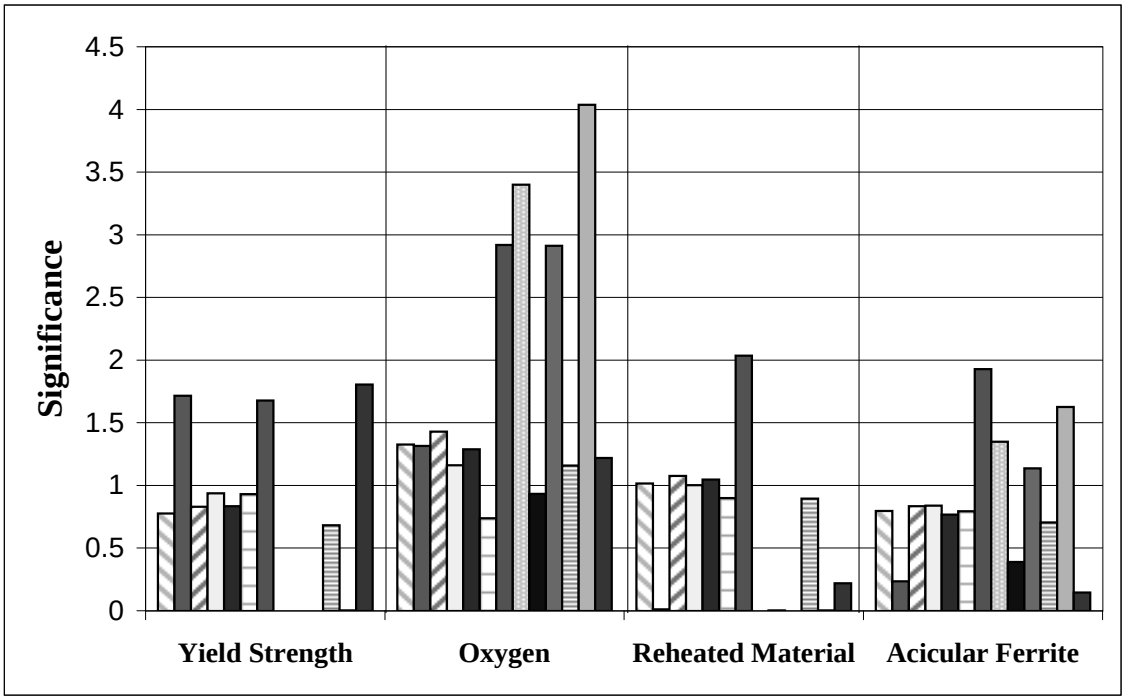

Fig. 5 Bar chart showing the perceived significance (σ_w) for each input variable. There are thirteen bars plotted per input, corresponding to each of the thirteen members of the optimum committee.

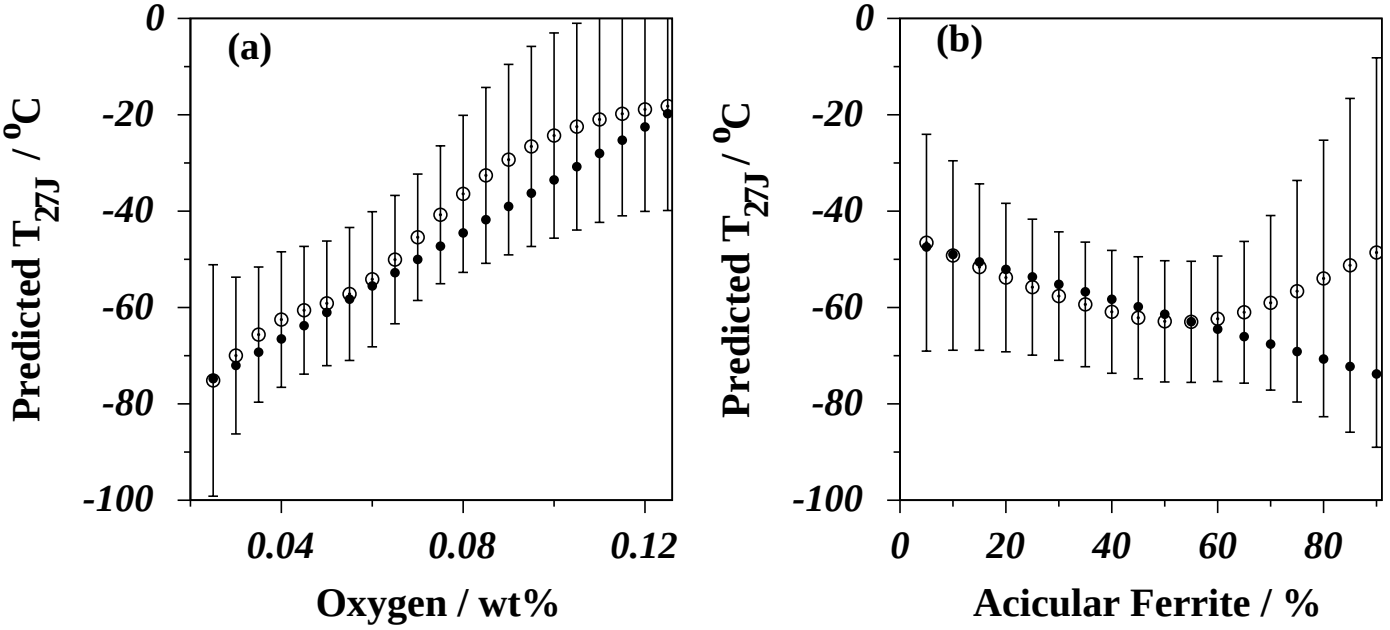

Fig. 6 Calculations as a function of the oxygen and acicular ferrite contents. In each case, the values of the remaining input variables are as listed in Table 2. The open circles with error bars are represent neural network model predictions whereas the filled circles are from equation (1).

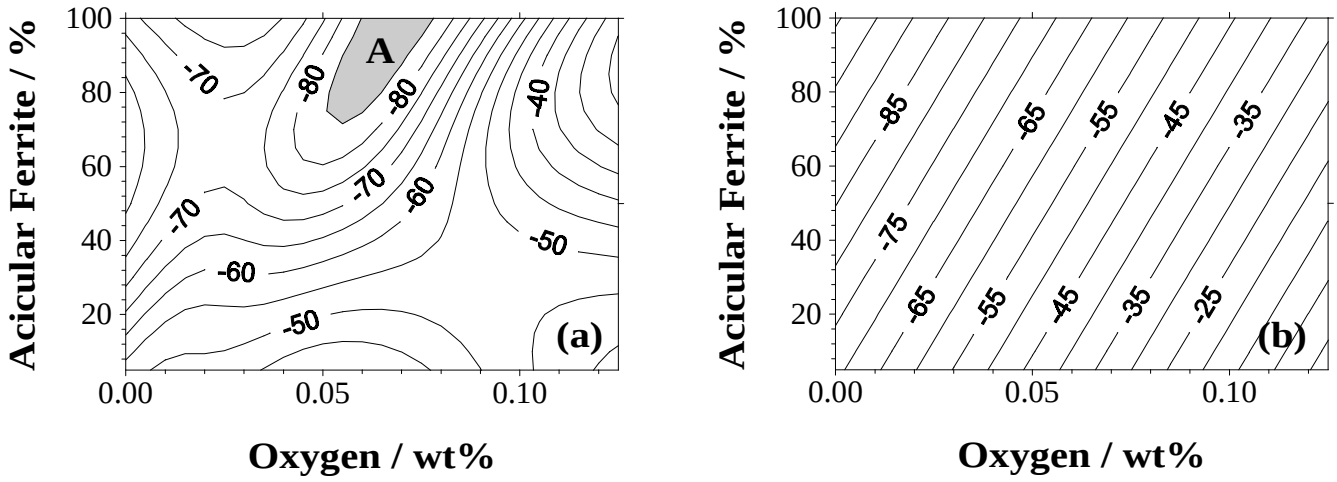

Fig. 7 Contour plots for calculations made using the following inputs: 510 MPa yield strength and 20% reheated material. The contour lines are expressed in °C, a) neural network predictions, here the error bars have been omitted for clarity but range from ±15–75 °C. The region marked 'A' shows that an optimum value of T_{27J} occurs at finite oxygen concentrations, b) using equation (1).