Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

Prediction of Yield Stress and Charpy transition temperature in highly neutron irradiated ferritic steels

Colin Windsor, Geoff Cottrell and Richard Kemp

EURATOM/UKAEA Fusion Association, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK, E-mail: colin.windsor@ukaea.org.uk

Abstract
Recent predictions have been made of metallurgical properties of low-activation ferritic/martensitic steels alloys at the high irradiation levels (displacements per atom or dpa) needed for a fusion power plant as based on measurements at low irradiation levels where more data is available. These have been published for the yield stress and for the Charpy ductile to brittle transition temperature shift. The neural network model predictions use training data up to a certain dpa level to predict metallurgical properties above this level. This "extrapolation" mode of neural networks is explored in some detail. Our studies revealed an increasing accuracy of predictions as the test dpa level is increased for both yield stress and Charpy shift predictions. This result suggests that a model exists for these metallurgical properties as a function of dpa level which becomes more accurate as the available irradiation range in the training data is increased. The explanation suggested is that the metallurgical annealing which occurs as the irradiation level is increased simplifies the microstructure and makes prediction more reliable.

Keywords: Metals and alloys, Neutron irradiation, yield stress, Charpy transition temperature
61.82.Bg, 61.80.Hg, 62.20.Fe

To be submitted to "Modelling and Simulation in Materials Science and Engineering"

1. Introduction: the prediction of the metallurgical properties of irradiated steels

In designing a fusion power plant it is of the utmost importance to be able to predict the metallurgical properties of steels at irradiation levels of 20-100 displacements per atom (dpa) due to high-energy (14 MeV) neutron bombardment arising from D-T fusion reactions. Data on the properties of irradiated materials is expensive to gather and there are currently no facilities capable of simulating fusion-spectrum neutron irradiation. Such databases that exist are primarily based on fission-generated neutrons and accelerated ions. The proposed International Fusion Materials Irradiation Facility (IFMIF) [1, 2] is designed to provide the experimental data required to validate materials design for fusion power plant facilities but has yet to be approved. The ITER fusion device [3] is now under construction but will provide predominantly low-dose data. In order to expedite the design of future power-plant facilities it is therefore necessary to assess the accuracy of extrapolating data collected from specimens with low dpa levels to those expected from high levels. This is especially important for the engineering design of demonstration fusion power stations (DEMO) [4]. A limitation of our work is that because nearly all the existing data are from fission-based irradiation sources, possible effects from the 14 MeV neutrons which will be present in fusion-based irradiations will not be taken in to account, in particular the *in situ* generation of helium by α-decay of activated atoms within the material matrix. This problem has been partly addressed by doping with boron atoms, which transmute to helium atoms under irradiation. For example Gaganidze et al [5] demonstrated that EUROFER97 steels doped with ^{10}B showed progressive embrittlement with increasing helium for irradiation temperatures below 340°C.

Databases are available for the yield stress [6] and Charpy transition temperature shift (a measure of fracture toughness) [7] of low activation ferritic alloys as a function of dpa. These contain

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

some data in the required irradiation range but have many more at lower irradiation levels, below 20 dpa. Our first paper [8] evaluated the prediction accuracy of a test dataset containing the yield stress of alloys at irradiations above 30 dpa using neural networks trained below this level. Our second paper [9] performed a similar evaluation for the Charpy ductile to brittle transition temperature shift, but for irradiation levels above 20 dpa. In this third paper we extend both previous studies by considering how the accuracy of both predictions varies with the upper irradiation level included in the training data. Previous neural network studies using the whole of the database for prediction have been published for yield stress [10] and for Charpy shift [11]. Both papers divided up the whole dataset into randomised training and test datasets and made no attempt to investigate the accuracy of predictions extrapolated outside the training data set.

In seeking simple predictive models it is relevant to consider the metallurgical origins of yield stress and fracture toughness as measured by Charpy shift in un-irradiated materials. Both properties can be increased by appropriate modification of precipitate structure and grain size [12]. However, fracture toughness is also highly dependent on grain boundary properties, particularly in irradiated steels where segregation of radiation-produced helium to the grain boundaries and radiation-accelerated precipitation of metastable phases can occur. For irradiated materials, it is common to refer to *hardening-induced embrittlement*, which is strongly correlated with changes in yield stress, and *non-hardening-induced embrittlement*, which can occur in the absence of such changes. The latter is associated with inter-granular failure, and can occur at high irradiation doses [13]. It is also worth considering briefly how radiation damage is measured. A sample which has received 20 dpa radiation damage has, on average, had each atom displaced from its lattice site 20 times. Under these circumstances it is almost remarkable that any semblance of the original crystal structure remains. The greatly enhanced atomic mobility and consequent rearrangement of the microstructure associated with these displacements can be regarded as akin to an annealing process and we might therefore expect convergence of material properties between different alloys at very high damage levels as microstructural differences are destroyed.

The yield stress database [6] contains 1811 example alloys E_i $i=1,1811$ with 37 input variables, including 31 atomic fractions $A_{i,j}$ $j=1,31$ defining the example alloy atom weight percentages, and six non-atomic variables $B_{i,j}$ $j=1,6$ defining other variables such as the cold work, irradiation temperature, radiation dose, helium dose, measurement temperature and a binary "as quenched" switch.

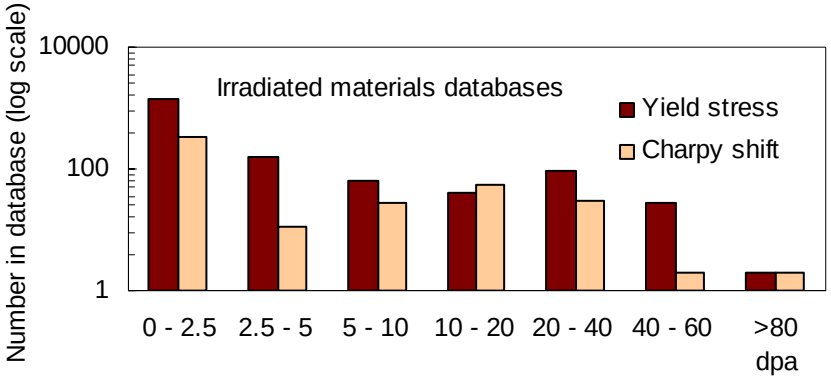

Figure 1: The irradiation level distribution through the database for the Yield stress and Charpy shift (Fracture toughness) databases. A large fraction of the data in both cases is concentrated at low irradiation levels below the 20 to 100 dpa level of interest for a fusion power plant.

The Charpy database [7] is considerably smaller having some 459 data points C_k $k=1,459$ with 26 input variables including 19 atomic fractions inputs $A_{k,j}$ $j=1,19$. All the specimens used were sub-sized, made by machining 3.3x3.3 mm^2 rods. Again there were six non-atomic input variables $B_{k,j}$ $j=1,6$, different from those in the yield stress database. These were: normalising temperature,

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

normalising time, tempering temperature, tempering time, irradiation temperature, and irradiation dose. This database gives the ductile-brittle transition temperature shift (ΔDBTT) as a function of irradiation level, given by the ductile-brittle transition temperature for un-irradiated material less that for the irradiated material. Unirradiated alloys are assigned a zero shift, and the change in transition temperature is measured as a function of irradiation level, dpa.

Figure 1 shows the distribution of the data with irradiation level for these two databases (note the logarithmic scale). It is seen that the data are concentrated below 20 dpa, and are sparse from 20 to 100 dpa. This range contains the levels relevant to a fusion power plant. For example 93.3% of the yield stress data and 92.6% of the Charpy shift data lie below 20 dpa.

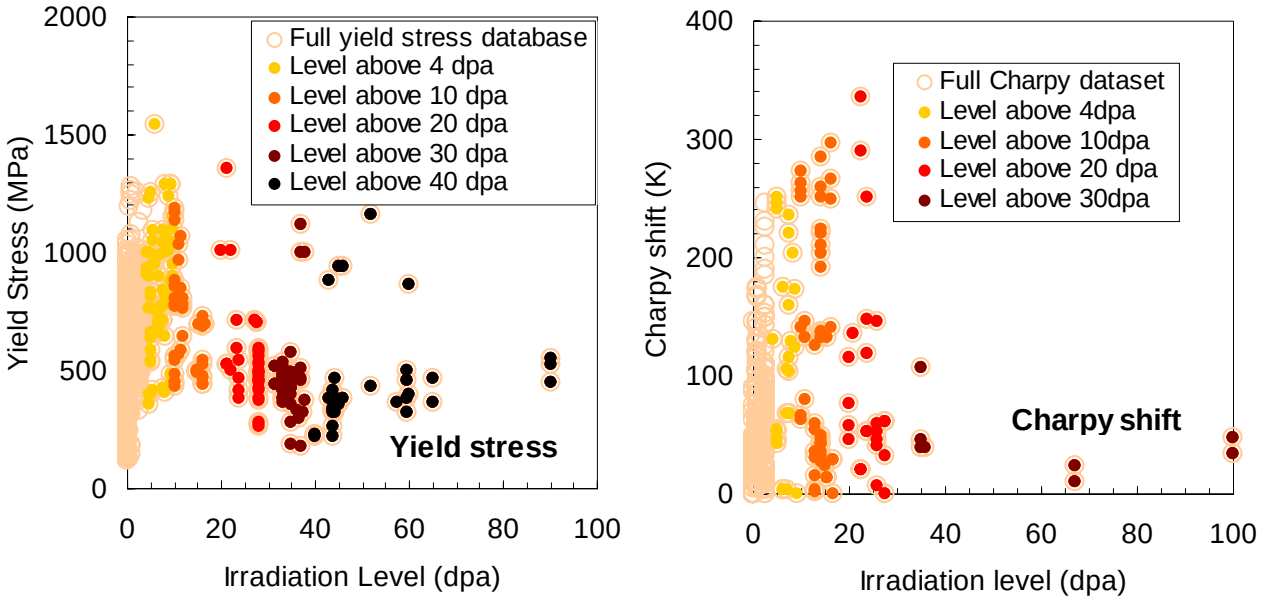

Figure 2: The yield stress and fracture Charpy shift databases plotted against the irradiation level. The darker shades correspond to higher irradiation levels.

Figure 2 shows the distributions of yield stress and Charpy shift as a function of dpa for each database. Both distributions show large spreads at any given irradiation level. This is partly due to the varying sample compositions and partly because of varying irradiation temperature and other non-atomic variables in the databases. However, there is also a reasonable level of scatter in the data. There is an important trend: the mean yield stress and Charpy shift tend to decrease as the dpa increases. This has been described in terms of "irradiation softening" by Yamamoto *et al* [6].

Test (dpa) irradiation	Yield stress		Charpy shift	
	Number	Test (%)	Number	Test (%)
>4	237	13.09	116	25.3
>10	161	8.89	86	18.73
>20	122	6.74	34	7.41
>30	84	4.64	7	1.52
>40	31	1.71	4	0.87

Table 1: The number and percentage of the yield stress and Charpy shift test data above given irradiation levels

For the purposes of assessing the optimal complexity of a neural network model, we need to divide the data sets into two subsets, a "training" set and a "test" set. The information on the test

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

irradiation limits used in this paper is summarized in Table 1, showing the number and percentage of the alloys in each database above a given test irradiation level. It will be clear why the limit of 30 dpa was chosen in our yield stress study [8], since it gives a useful 84 test examples (4.6%). Similarly the lower 20 dpa limit in the Charpy shift study [9] gives a tolerable 34 test examples (7.4%). This paper explores each of the chosen irradiation limits shown in the table. In addition, some data can be held back as a "validation" set, to assess the reliability of the trained model in predicting unseen data.

2. Neural network predictions of yield stress and Charpy shift: the role of extrapolation

The methods used have been described in earlier studies [8, 9]. The datasets are divided into training and test datasets, where the training data comprises those alloys with irradiation up to some defined level, 4, 10, 20, 30 or 40 dpa, and the test data includes those remaining data at higher irradiation levels. The trained neural network thus seeks to extrapolate its knowledge into higher dpa domains. This methodology contrasts with earlier conventional studies [10, 11] which divide the dataset into equal randomly selected training and test datasets, and seek to find the best overall interpolation.

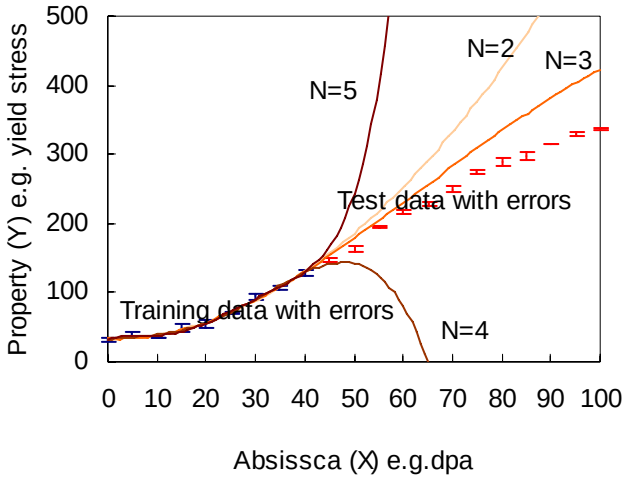

Figure 3: A simple example showing the pitfalls of extrapolation in least squares fitting. The "mock" data are from a third order polynomial showing some mild saturation with added statistical errors. The training data lie below 40 dpa. Best fitting polynomials to these data are shown for the polynomial orders N=2 to 5.

Figure 3 illustrates the potential dangers of extrapolation for the simple case of a network with training data up to 40 units (say dpa) and test data from 40 to 100. The *X* axis may thus be thought of as irradiation level, and the *Y* axis as any variable being fitted to the available data, such as yield stress. The training and test points are from a third order polynomial with the imposed random errors shown. The training data *(X<= 40)* are fitted to polynomials of order two, three, four and five, and the fitted distributions are shown over all the data. All the polynomials shown fit the training data set well, and their differences are not evident on a plot of this scale. However beyond the range of the training data, some of the curves diverge quickly. The simplest polynomial (order *N=2*) fit is smooth but fails to show the tendency for saturation in the data. The *N=3* fit is much better, correctly predicting the general form of the data. For the present purpose, a measure of "complexity" is the order of the fitting polynomial, *N*. Thus the *N=3* curve has a degree of complexity matching that of the data. The higher order polynomial fits with *N=4* and *5* diverge rapidly from the data at larger *X*. The conclusion is that extrapolation can work if the complexity of the extrapolation function correctly mimics that of the underlying data.

In the present problem of predicting mechanical metallurgical properties using neural networks, the principal way by which the network complexity may be adjusted is by varying the

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

number of hidden units. A second way is by adjusting the number of included input parameters. The conventional method of determining the optimum network architecture is to divide the available data into two subsets. One subset (the "training" data) is used as fitting data for a wide number of networks of different complexity. The other subset (the "test" data") is then used to assess how well the network performs on new data. An under- or over-complex network may fit the training data well but will fit the test data poorly and the optimum complexity can hence be identified by a minimum in the "test error". Once the best network architecture is identified, the network can be retrained on the full dataset. Conventionally these subsets are a random fraction of the whole dataset but in the present case the training data consists of alloy data up to some defined irradiation level, and the test data a fraction of the alloys at higher irradiation levels. Any remaining data constitute a validation dataset whose purpose is to provide an independent check on the performance of the network.

The optimisation of the parameters of the network is achieved by defining a cost function to be minimised. These results are assessed through the root mean squared least squares residual R_t. This is defined as the square root of the summed squared difference between the network targets T_i and calculated network outputs O_i divided by the number of examples N:

$$R=\{[\Sigma_{i=1,N}(T_i - O_i)^2]/N\}^{1/2} \quad (1)$$

where the summation i includes only the test examples chosen from the dataset. Here T_i are the network targets, essentially the experimentally-obtained values of the yield stress or Charpy shift. O_i are the outputs of the trained neural network. The physical units of O_i and T_i are the units of the unknown variable (MPa for the yield stress and K for the Charpy shift). In addition the "coefficient of determination" [14] will sometimes be given, defined by

$$C = \{1 - [\Sigma_{i=1,N}(T_i - O_i)^2]/ [\Sigma_{i=1,N} (T_i - T_{ave})^2]\}^{1/2} \quad (2)$$

where T_{ave} is the average value of the experimentally-obtained values. This coefficient is equal to unity for a very good fit, and reduces to zero for a poor fit where deviations of predicted values from experiment are comparable with the spread of experimental values within the dataset. In the present case the deviations of predicted values from experiment are occasionally so large that the coefficient of determination has a negative argument in the square root. Such cases will be indicated by "-". The neural network code used, TDR, is an adapted version of the MAGEOM code written under the direction of Chris Bishop [15].

The objective is to find the network, defined by its number of hidden units and the weights associated with each input parameter, which gives the lowest test. The validation dataset is unseen by the neural net during the training and testing regimes. When data is scarce, as it is in this case of metallurgical data at high irradiation levels, the validation stage may be omitted but it must be understood that any results deduced from the test data will only be valid within the parameter range covered by the test data and no information is available outside this range.

The process of optimising the network complexity is illustrated in figure 4. The networks in this figure represent the computer runs with the lowest residual out of 20 separate runs with different random number seeds (*i.e.* different initial guesses for the network weights at the start of the training process).The training data with irradiation levels below 10 dpa are shown by the open diamonds. The test data, shown by the dark full triangles, are for levels between 10 and 20 dpa. The validation data, shown by the crosses, consist of data with levels above 20 dpa. The first plot with one hidden unit shows predicted Charpy shifts which hardly correlate with the actual value and so give a relatively large training residual R_{train}=48K. The test and validation performance is similarly poor. This

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

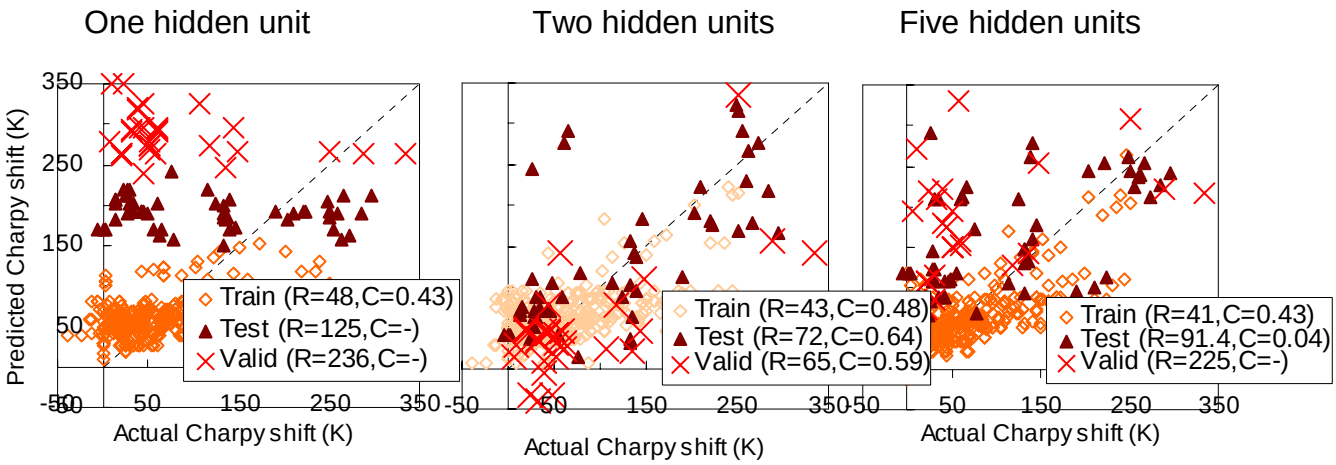

Figure 4: The training, testing and validation data scatter plots for varying hidden unit numbers in Charpy shift data. Training data is below 10 dpa, testing data between 10 and 20 dpa and validation data above 20 dpa. The network was trained using atomic concentration inputs.

network clearly does not capture important trends in the data and is insufficiently complex. Plot three (five hidden units) has a good fit for the training data but the test and validation data show large scatter. However the second plot, with two hidden units, shows greatly improved performance with fair prediction accuracy for train, test and validation dataset. Although both the test and validation residuals are larger than the training residual as might be expected, the validation residual is lower than the test residual. This point will be returned to later. In general, as the network complexity increases, the training residual *R* decreases monotonically but the test residual has a minimum at a particular number of hidden units. This network then performs dramatically better with the reserved high-irradiation validation data.

The full lines of figure 5 upper left hand plot summarize the results of this procedure by giving the training, test and validation residuals as a function of hidden unit number. It is seen that the training data residual reduces more or less continuously. This is as expected as the network is becoming more complex and better able to assimilate detail in the training data. The test data residual, by contrast, has a rather flat minimum between two and three hidden units. The conclusion is clear. Networks with low hidden units numbers, (and correspondingly low numbers of adjustable parameters) are able to make better predictions when extrapolating from the training data for this dataset. The validation data residual has an even lower minimum value of 65 K (compared to 72 K) and at the same hidden unit number of two. However it has more scatter and rises more rapidly at both lower and high hidden unit numbers.

We believe that these results are sufficiently important and unusual to warrant further investigation. The same procedure has therefore been followed for figure 5 (ii) training <20 dpa, testing 20-30 dpa and validation ≥ 30 dpa; (iii) training <30 dpa, testing 30-40 dpa and validation ≥ 40 dpa; and (iv) training <30 dpa, testing 30-50 dpa and validation ≥ 50 dpa. All these results give a model for the Charpy prediction at high irradiation levels which varies smoothly with the irradiation limit of the training data. The training residuals are very similar as might be expected since comparatively few alloys were being added to the training data as the upper irradiation limit is increased (table 1). However a clear trend is that, as the extent of the extrapolation decreases, the test and validation residuals decrease. This is illustrated in figure 6. It is seen that both sets of residuals decrease as the test range of irradiation levels increases. The main difference is that the validation minimum is much sharper than the test residual minimum. This decrease in residual with dpa corresponds to a reduction in the test and validation ranges compared with the training range. We are therefore led to the (expected) result that the prediction accuracy tends to improve when the range of extrapolation is reduced for the optimal choice of hidden unit number.

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

An important conclusion may be drawn. For this Charpy shift dataset, the performance of the validation data set suggests that extrapolation from lower irradiation levels is possible and is not greatly different from the test data performance, provided that the hidden unit number is optimal. This conclusion is not common in the neural network literature. For example Lohninger [16] states: "Neural networks exhibit a major drawback when compared to linear methods of function approximation: they cannot extrapolate." Clearly this result is formally correct as we cannot know in advance that a particular data set behaves in a way that allows any functional dependence of a target property over low values of a variable to be maintained at high values. However in the case of mechanical properties such as a shift in the DBTT due to exposure to radiation, it is reasonable to conjecture an underlying functionality which describes the metallurgy controlling the variation with irradiation level. That is, we assume the mechanical property extrapolates with some degree of smoothness and we expect no discontinuities.

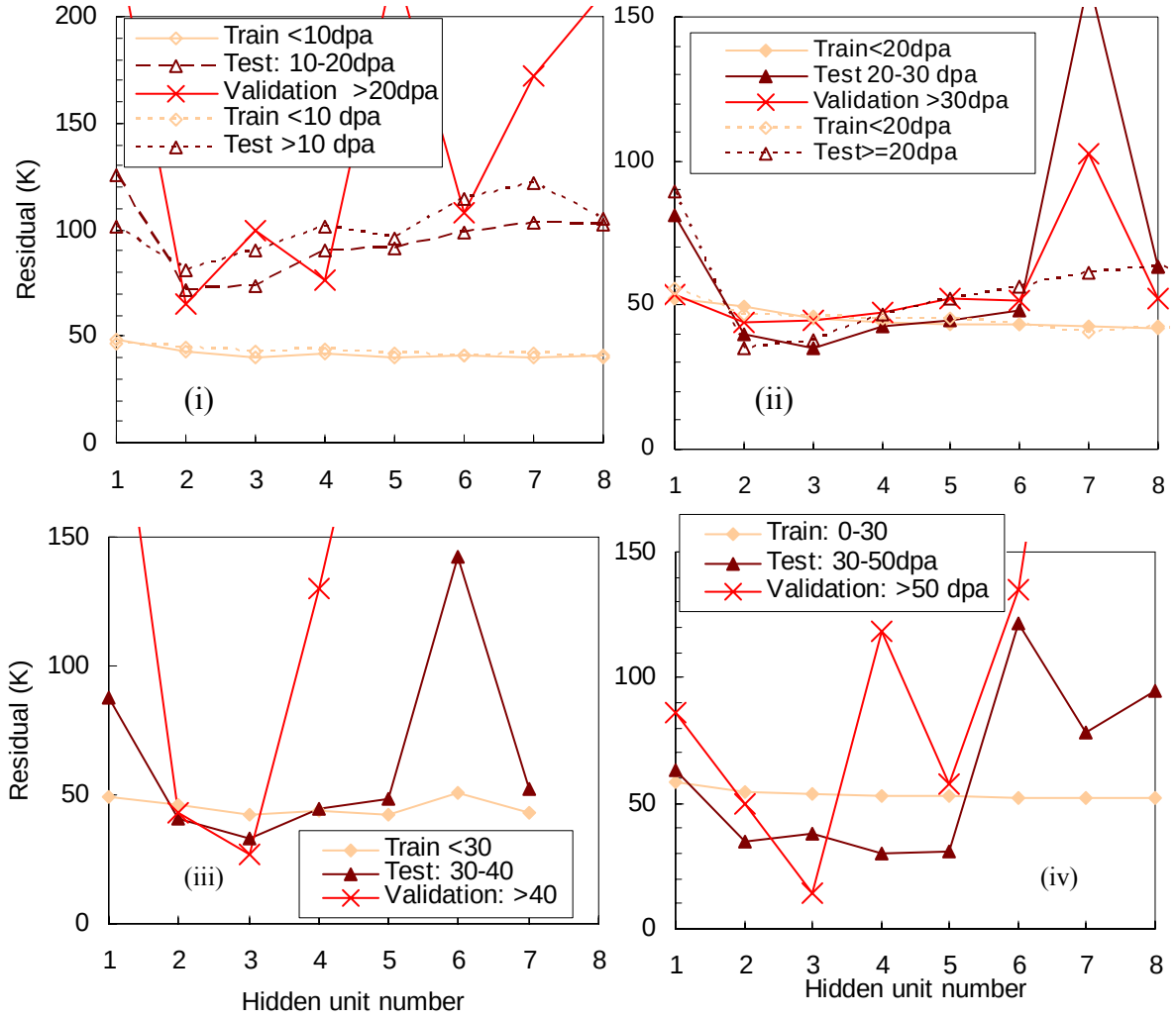

Figure 5: The dependence of the residual fit to Charpy shift data on hidden unit number for various extrapolations where data is trained at low irradiation levels (feint filled diamonds) and tested and validated at higher irradiation levels. The full lines and filled symbols represent situations where there is a test irradiation range followed by a higher validation range shown by crosses. The dashed lines and open symbols correspond to the case of no validation data, where the test and validation ranges have been combined together. The four cases (i)-(iv) correspond to the irradiation level ranges indicated.

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

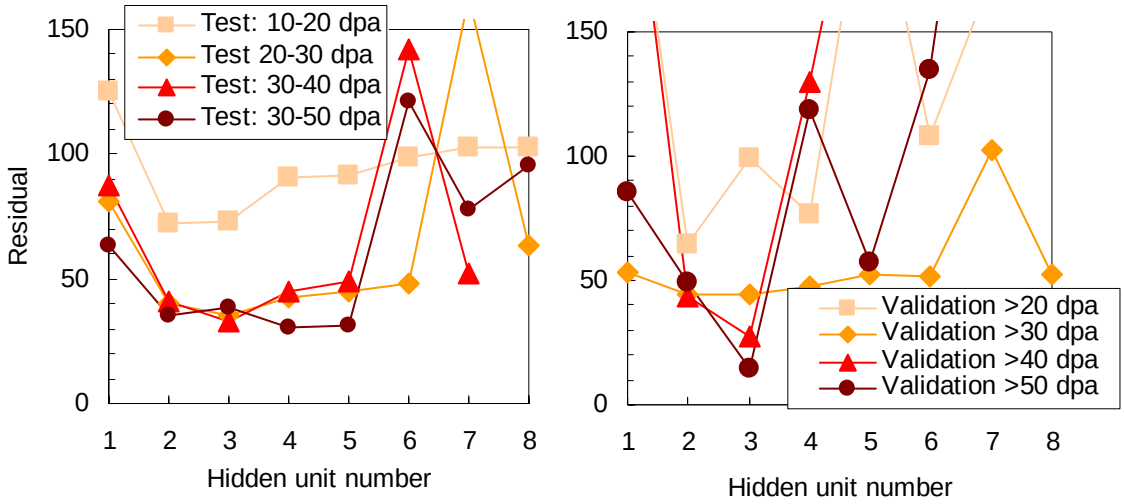

Figure 6: The dependence of the residual fit to Charpy shift data with hidden unit number for test and validation data at varying irradiation ranges. It is clear that the minimum residual decreases for both test and validation data, as the irradiation level increases.

The dashed lines in figure 5 show similar results with the same Charpy shift dataset, but with the data divided into just two fractions, a training dataset below 20 dpa for the top left hand graph and a test dataset above 20 dpa, and no validation dataset. These results differ in that they were averages over 20 runs with different initial random seeds and so have smaller statistical fluctuations. The results shown by the open symbols compared with the earlier closed symbols are very similar, with the training data residuals being hardly distinguishable. The lowest test residual is occasionally different by one hidden unit but generally the variation with hidden unit number is very similar.

Again a conclusion may be drawn for this dataset. The dependence of network performance on hidden unit number when the data are divided into training, test and validation sets is very similar to that when there are no data reserved as validation data. This result will be used in defining all the results of section 4.

3. Bayesian predictions of the prediction error: committees of selected complexity

Further insights into these results are found by using the "BIGBACK" code written by David MacKay [17]. This was run using the interface code "Model Manager" [18]. It includes a full Bayesian error treatment so that values of the property to be predicted are given an error bar defining the prediction uncertainty. This uncertainty represents the spread of neural network models which may fit the training data, whilst the nominal prediction represents the most probable model. Also it is able to lower the weightings of irrelevant variables automatically so that all possible input parameters may be included and the program itself is able to select and appropriately weight each of the possible input parameters.

The conventional use of the program uses a "committee" structure where several different networks from runs with different random starting positions are combined together and averaged. The committee choice may also be defined to define the network complexity. A series of BIGBACK runs was performed when the training data was a randomly chosen 2/3 of the alloys within the irradiation range up to 20 dpa, and the testing data the remaining 1/3 of the alloys in this range. The

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

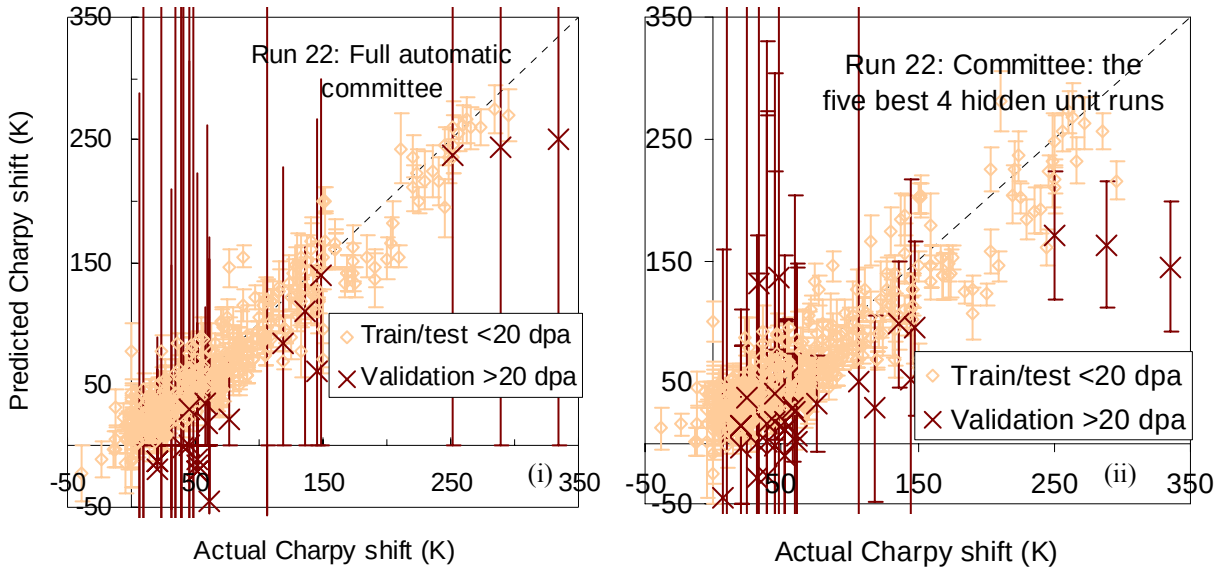

Figure 7: The predicted Charpy shift compared with its actual value for two committee structures generated using the Bayesian BIGBACK code. Both committees used data trained and tested with 100 randomly chosen sets of data below 20 dpa. Data above 20 dpa was unseen and used for validation. In (i) the standard automatically chosen BIGBACK committee structure was used. The fit is good but the prediction uncertainties are large. In (ii) the committee is chosen from the 5 best runs with 4 hidden units.

validation data were the remaining alloys with irradiation levels above 20 dpa. Figure 7 shows the estimated prediction and, by the error bars, the prediction uncertainty, of the Charpy shift compared with its actual value. Figure 7(i), on the left, shows the results of the default committee – consisting of a set of networks with varying hidden unit numbers – generated by Model Manager. The training results are good with a residual R=19.3K (C=0.95). However a large prediction uncertainty is seen for the validation dataset with a residual R=208K, and no real coefficient of determination as the higher dpa examples were very poorly predicted. Nine points lie off the scale of the graph. The prediction uncertainty increases on extrapolation, as expected, as the predictions from all possible models supported by the data diverge. Although the nominal predictions for the most probable model are reasonable, the uncertainties act as a warning that the neural network is not well constrained in this prediction region and so the predictions should be treated with some caution. It is clear that if one wishes to reduce the uncertainties and use the extrapolation mode with confidence, the complexity of the network must be appropriately reduced. In figure 7(ii) are the results from a manually chosen committee generated by deleting the automatically generated committee structure and choosing committees with specific properties, such as having a given number of hidden units. Here we show for the same set of training data runs, the results from a committee chosen from the five best runs of just four hidden units. The training and test data have a larger scatter with R=26.7 (C=0.87) but it is seen that the calculated uncertainties on the validation data are much reduced, as the flexibility of the network has been constrained. Only one point on this scatter plot now lies outside the scale of the plot. The validation residual is R=70.6, (C=0.44) much better than that for the default committee.

The dependence of the validation residual with hidden unit number is shown in figure 8. The validation residual is reduced very sharply as committees of increasing complexity are used. Then for five hidden units and beyond the residual again rises sharply. The minimum is in fact much sharper than was seen in figures 5 and 6 where no committee structure was used.

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

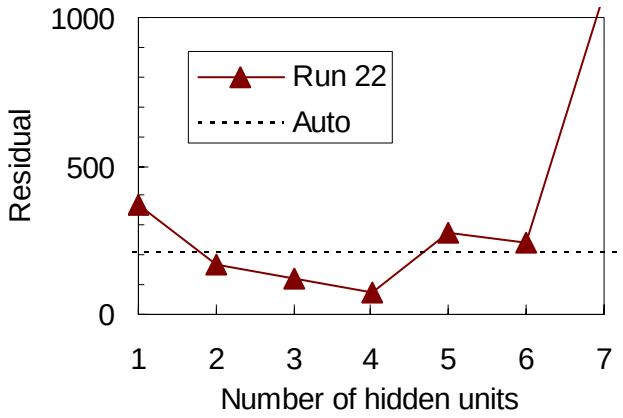

Figure 8: The dependence of the validation residual for BIGBACK committees chosen with varying hidden unit numbers. The results are for Charpy shift data trained with several randomly chosen sets of data below 20 dpa, with data above 20 dpa being used for validation.

4. Dimensionality reduction – when can it aid extrapolation?

Returning to figure 3 we can see that the form of the fitting function, in this case the order of the polynomial must be appropriately matched in complexity to the problem. If the observed changes in Charpy shift and yield stress with irradiation level are relatively smooth then a simple fitting function of low complexity will suffice. However, if there are highly non-linear or discontinuous changes then no simplification may be possible. Highly complex fitting functions (for example neural nets with large hidden unit numbers) are required to fit such data and it is unlikely that extrapolation will be an option. A physical example of such a non-linear change with increasing irradiation level is the appearance of a new phase, such as helium bubbles, within the metal matrix. This new phase may appear suddenly at some particular irradiation level and cause an immediate change in properties. In fact there is little metallurgical evidence for such a change within this dataset. Helium bubbles have not been observed in most of the ferritic steels used in this study. Figure 1 illustrates the yield stress and Charpy transition temperature shift for all the alloys as a function of irradiation level. There is appreciable scatter but there is little evidence in either variable for any discontinuity in these properties as a function of irradiation. This does not, however, preclude the possibility of such behaviour occurring under fusion-spectrum irradiation but this would be a problem for any method attempting to predict properties based on the existing data: it is not a problem unique to neural networks.

Returning to figure 4, we see that validation data over the region of extrapolation indicated an optimal hidden unit number and thus an optimal degree of network complexity. Having identified this point, we can redefine most of the validation data as test data and re-optimise the parameters of the neural network on the basis of minimising the test data residual. At this stage, only very minor changes in the network weights are expected.

An alternative and complementary method of reducing the number of adjustable parameters defining the network is to reduce the number of input variables. Several techniques have been used to achieve this:

(i) training begins with just those non-atomic "fixed" variables which are known to be beneficial in reducing the residual, R_{test}

(ii) the atomic fraction inputs are added in turn for each atom and that input giving the largest reduction in R_{test} noted. This follows the forward sequential selection of features method [19]

(iii) after the selection of the best atomic input, linear combinations of the other atomic inputs are made. That combination which gives the lowest R_{test} is selected for the network. This is our target-driven components method described in more detail in [8]

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

(iv) when the first set of atomic components is complete, in the sense that further atomic components do not give any reduction in R_{test}, a second component beginning with that atomic input giving the lowest R_{test} is constructed. This process continues until there is no further reduction in R_{test}

(v) this process is repeated for a series of hidden unit numbers and the one giving the lowest R_{test} is chosen.

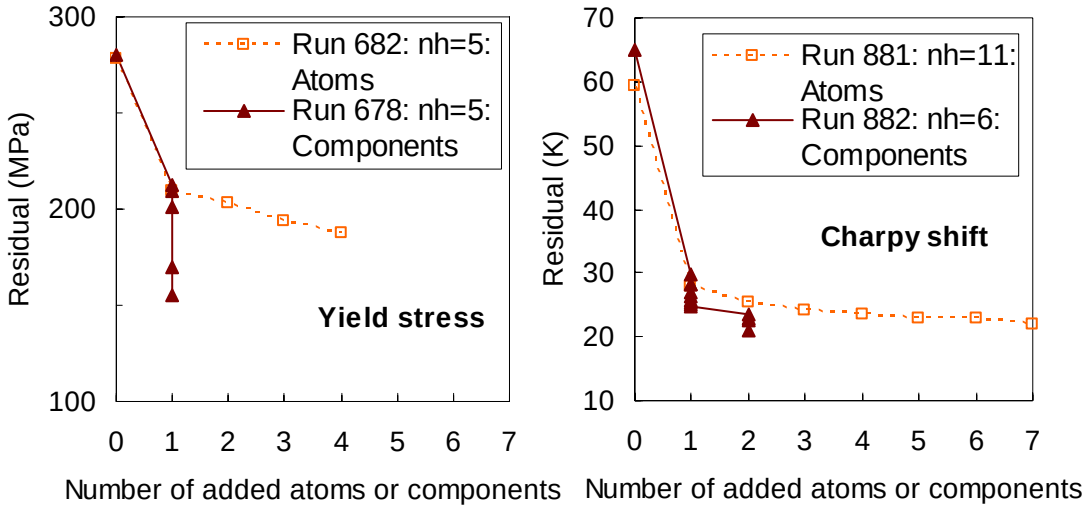

Figure 9: The reduction in the residual as atomic inputs or target-driven components are added to a neural net. For both the Yield Stress and Charpy shift databases the test data were for alloys with irradiations > 20 dpa. The results shown are for the optimal number of hidden units (nh) in each case. The vertical lines of full triangles indicate the reduction in residual with components containing an increasing number of linear combinations of atomic inputs.

Figure 9 shows an example of these methods applied to the yield stress and Charpy shift databases, demonstrating the change in the residual as additional atomic inputs or target-driven components are added to the network. The open squares and dashed lines show the effect of adding additional atomic inputs. A reduction in the residual is seen to occur for four atoms and seven atoms respectively. However the lowest residuals are seen to occur for the addition of target-driven linear components of atomic inputs, as shown by the closed triangles. In this method the additional network inputs are not simply some atomic input, but a linear combination of atomic inputs. On metallurgical grounds it might be expected that some alloying elements act in combination with or similarly to one another. For example tungsten promotes the formation of hard carbide precipitates and hence has a similar effect to manganese. In fact it is found that a linear combination of these two inputs gives a lower residual than either individually and the yield stress component in figure 9 contains the combination $0.73A(\mathrm{W})+0.09A(\mathrm{Mn})$, where the A represents the corresponding atomic input in weight percent. The component method has the advantage of reducing the overall number of inputs to the network and so constraining the range of extrapolated behaviour, leading in turn to more realistic predictions.

5. Yield stress and Charpy shift prediction accuracy against test irradiation level

Runs have been performed for the yield stress and Charpy shift databases using both the added atomic inputs and the target-driven components method as a function of the test irradiation levels as detailed in table 1. In all cases a scan has been made over hidden unit number from 1 to 20 hidden units to find the network complexity which gives the best prediction as defined by the minimum

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

residual. In the case of the Charpy shift prediction, the calculations have been made with and without the inclusion of the irradiation temperature *Tirr* as one of the non-atomic input variables. The results shown in figure 10 are for that case which gives the lowest residual. In all yield stress results the best results were from the target-driven components method. In nearly all Charpy cases the best result was from the target-driven components method excluding the irradiation temperature input *Tirr*.
The occasional improvement from omitting the input *Tirr* is surprising as this is an important physical variable. This can occasionally occur in neural network modelling if inputs are correlated and one can be used as a proxy for another, but in this dataset *Tirr* does not appear to correlate with any other inputs. This result was previously reported [9] and it is likely that there is not enough variation of *Tirr* within the test (high dpa) data to allow any dependence to be clearly identified.

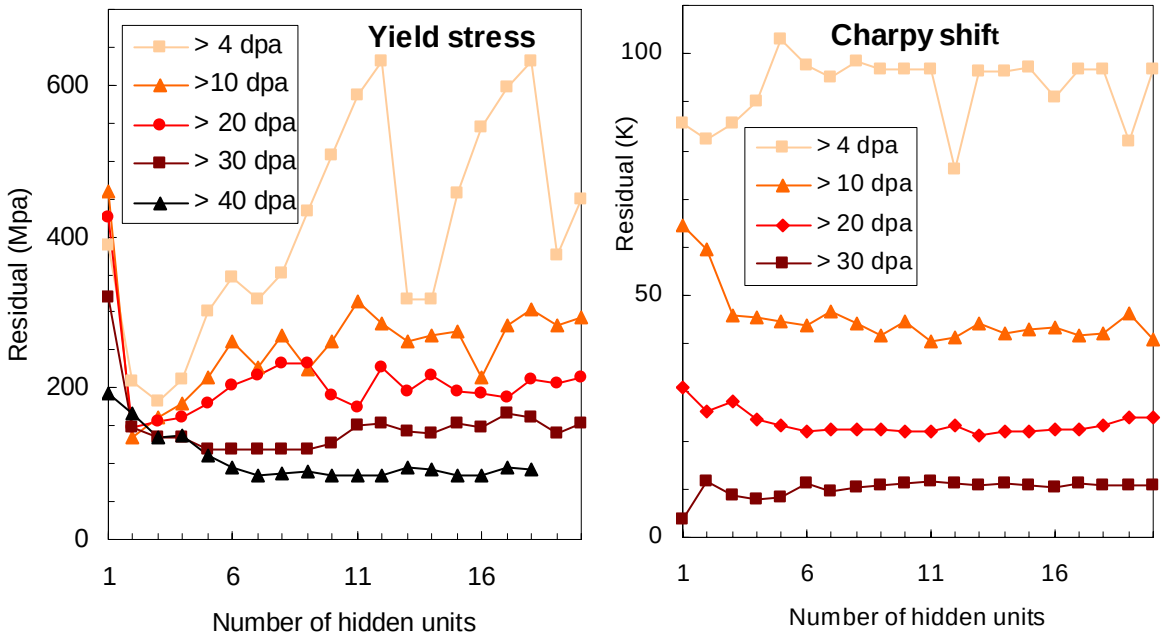

Figure 10: Dependence of residual on hidden unit number, for different sets of neural network test data as described in Table 1. In the case of the Charpy shift, the lowest recorded residual is shown whether or not the irradiation temperature variable *Tirr* was included. The darker shades correspond to training up to higher irradiation levels (and correspondingly fewer test data).

In all these results the datasets have been divided up simply into fixed training and test datasets based on the irradiation level shown. The discussion of section 2 shows that while division into training, test and validation sets is generally effective, this is not the case with validation datasets involving extrapolation. Dimensionality reduction with an informed choice of the complexity of the network is essential. Clearly this is only possible if there is some underlying simple behaviour in the dataset and a smooth extrapolation can be assumed. Figure 10, particularly for the yield stress on the left illustrates very nicely that for large extrapolations (from low dpa levels) a low complexity (low hidden unit number) network is optimal. As the degree of extrapolation diminishes, as at the higher dpa levels, the optimal complexity of the network increases, although for very high irradiation levels with small test datasets, there is no clear minimum in the test residual, indicating that there are probably too few test data to adequately constrain the network output.
Several interesting results are evident from figure 10. There is a clear general trend for the lowest residuals to occur for the highest test irradiation levels. This might have been expected because there is a smaller percentage of the total dataset present in the test data (as noted in Table 1), and a correspondingly larger percentage of the total dataset available for training. However it might also be

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

expected that the data at very high irradiation levels, well outside the range of the training data, would be poorly predicted. That this is not so suggests that the assumption of smoothness in the data is justified over this range of extrapolation. One possible explanation for this may be that the phenomena occurring in alloys at high irradiation levels are similar to those due to annealing. At irradiation levels of many dpa most atoms in the alloy will have had many chances to "anneal" to their lowest energy state from an initial state which may have been far from equilibrium. Such phenomena may lead to the observed reduction in yield stress and Charpy shift with increasing irradiation level (figure 2).

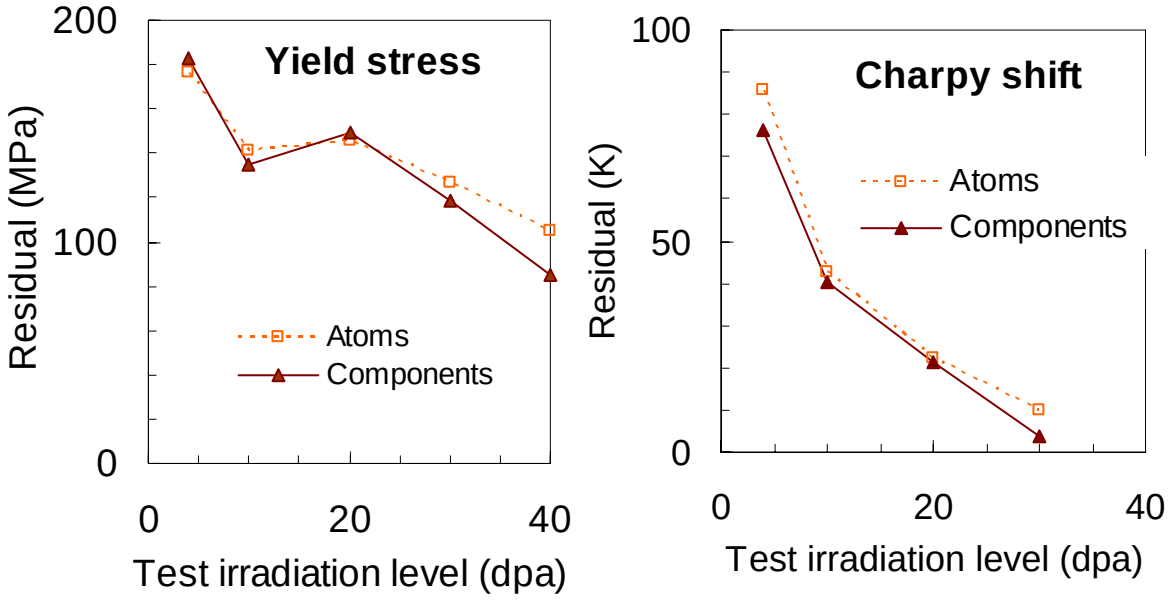

Figure 11: The lowest residual as a function of hidden unit number plotted against the test irradiation level for the Yield stress and Charpy shift databases. The dashed lines and open squares indicate the added atom method, the full lines and closed triangles the target-driven components method.

A second result clear from figure 10 is that yield stress and Charpy shift behave quite similarly, although the yield stress dataset is around four times larger. The hidden unit number for the best performance is however different. For the yield stress set the optimal hidden unit number is low, around two or three, for low test irradiation levels but increases to around 10 for higher test irradiation. For the Charpy shift, the optimum hidden unit number remains around ten, except for the > 30 dpa test irradiation, where it shifts to one.

The main results of this paper are summarized in figure 11. For each of the plots against hidden unit number in figure 10, the minimum residual is determined and is plotted against the test irradiation level. Both plots fall sharply indicating that the predictions above the higher irradiation levels are best. This remains true for inputs consisting of both raw atomic concentrations and target-driven components methods. However for the yield stress the reduction is flatter in percentage terms, showing that the improvement in prediction accuracy with increasing test irradiation level is less for yield stress than for the Charpy shift. The explanation for this in metallurgical terms is suggested by figure 2. At high irradiation levels (>20 dpa) the material has been essentially melted and annealed several times over. The microstructures present in the original alloy, chosen by virtue of their favourable metallurgical properties, have been modified to give an alloy with generally more average properties. The large scatter in the yield stress and Charpy shift data seen in figure 2 at irradiations below 20 dpa changes to lower average levels with less spread.

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

Figure 12 shows the actual performance in the prediction of yield stress and Charpy shift for test data with irradiations > 20 dpa. These results use the target-driven components method, which, in both cases, gave lower residuals than the added atoms method. The conclusion from figure 6 is that the useful predictions of the yield stress and Charpy shift at irradiations greater than 20 dpa, based on data from lower irradiation doses, are quite feasible. This is an important result since it suggests that useful predictions of possible alloys for the fusion power plant regime may be made using existing published data which is dominated by measurements at lower irradiation levels.

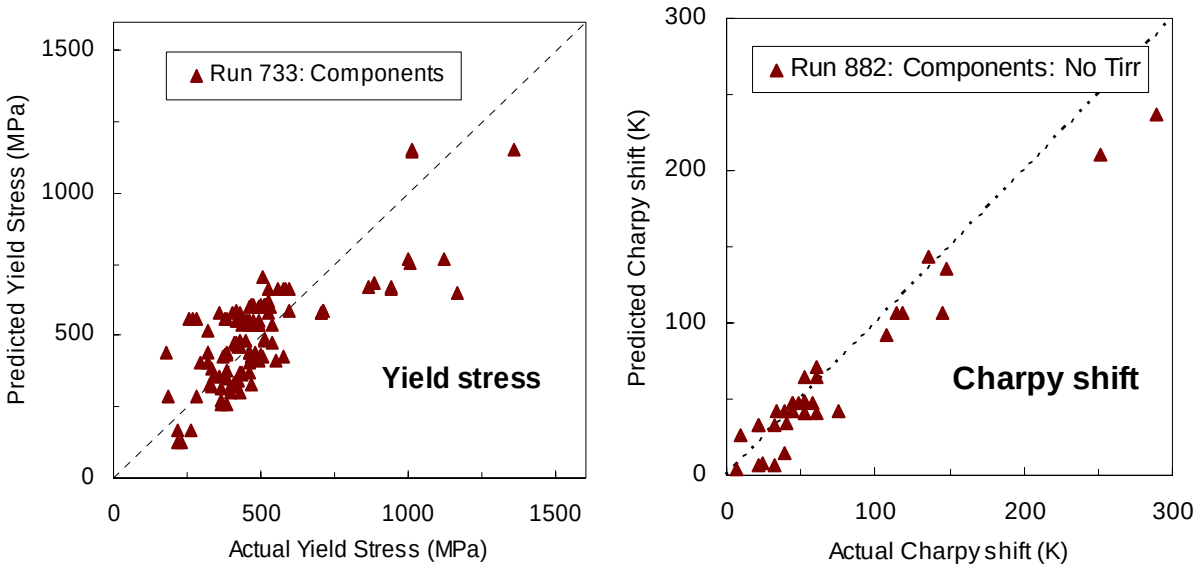

Figure 12: Scatter plots of the predicted yield stress and Charpy shift for test data at irradiations > 20 dpa. In both cases the target driven components method was used.

6. Conclusions

Neural network predictions of both the yield stress and Charpy shift in highly irradiated specimens have been made as a function of the test irradiation level. In particular:

(i) The utility of the extrapolation mode using validation and test datasets in the context of dimensionality reduction has been demonstrated for networks of appropriate complexity.

(ii) The accuracy of both yield stress and Charpy extrapolations increases with the level of the test set irradiation threshold suggesting that results at lower irradiations provide a reasonable predictive model for higher irradiation levels. This also suggests that different radiation damage mechanisms may dominate on short and long irradiation timescales. The data do show a very rapid initial change in yield stress and Charpy shift with irradiation dose, which appear to stabilise to some extent at higher doses.

(iii) At many irradiation levels, networks with a low number of hidden units provide the best predictions.

(iv) The target-driven components method generally provides the best predictions. The reason may again be that this method results in a low dimensional input space for the network.

(v) Useful predictions of the yield stress and Charpy shift in alloys with irradiations above, say 20 dpa, can be made from training data at lower irradiations.

Although the low-dose data appear to have a high degree of scatter, the successful prediction of higher-dose data from a model based on low-dose data also implies that the few high-dose data are representative of the material properties changes expected under irradiation from currently available radiation sources.

Corresponding author: C G Windsor: colin.windsor@ukaea.org.uk

Acknowledgements

This work was funded by the United Kingdom Engineering and Sciences Research Council and the European Communities under the contract of Association between UKAEA and EURATOM. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

[1] A. Moeslang, V. Heinzel, H. Matsui, M. Sugimoto, The *IFMIF test facilities design*, Fusion Engineering and DesignVolume 81, Issues 8-14, , Proceedings of the Seventh International Symposium on Fusion Nuclear Technology - ISFNT-7 Part B, February 2006, Pages 863-871.
[2] R. Andreani, E. Diegele, R. Laesser, B. van der Schaaf, *The European integrated materials and technology programme in fusion*, Journal of Nuclear MaterialsVolumes 329-333, Part 1, , Proceedings of the 11th International Conference on Fusion Reactor Materials (ICFRM-11), 1 August 2004, Pages 20-30. [5] *The ITER project*: http://www.iter.org
[3]*Presentation on the iter website*, http://www.iter.org/Future-beyond.htm
[4] Maisonnier *et al* . *DEMO and fusion power plant conceptual studies in Europe,* (2006), Fusion Eng. and Design, 81, 1123-1130.
[5] Gaganidze E, Petersen C and Aktaa J, *Study of helium embrittlement in boron doped EUROFER97 steels,* (2009) Journal of Nuclear Materials, 386, 349-352.
[6] Yamamoto Y, Odette G R, Kishimoto H and Rensman J W, *Compilation and preliminary analysis of an irradiation hardening and embrittlement database for 8Cr martensitic steels* (2003) Technical Report DOE/ER-0313/35, ORNL (yamataku@engineering.ucsb.edu)
[7] Yamamoto T, Odette G R, Kishimoto H and Rensman J, *On the effects of irradiation and helium on the yield stress changes and hardening and non-hardening embrittlement of ~8Cr tempered martensitic steels: Compilation and analysis of existing data,* (2006), Journal of Nuclear Materials 356, 27–49. linkinghub.elsevier.com/retrieve/pii/S0022311506002418
[8] Windsor C G, Cottrell G A and Kemp R, *Prediction of yield stress in highly irradiated ferritic steels,* Modelling Simul. Mater. Sci. Eng. 16 (2008) 025005
[9] Windsor C G, Cottrell G A and Kemp R, *Prediction of Charpy transition temperature in highly irradiated ferritic steels,* Modelling Simul. Mater. Sci. Eng. 16 (2008) 075008
[10] Kemp R, Cottrell G A, Bhadeshia H K D H, Odette G R, Yamamoto T and Kishimoto H 2006 *Neural network analysis of irradiation hardening in low-activation steels* J. Nucl. Mater. 348 311–28 and http://www.msm.cam.ac/phase-trans/2006/Kemp.JNM.2006.pdf
[11]Cottrell G A, Kemp R, Bhadeshia H K D H, Odette G R, and Yamamoto T, *Neural-Network Analysis of Charpy transition temperature of irradiated low-activation steels*, Journal of Nuclear Materials 367-370 (2007) 603-609
[12] Bhadeshia, H K D H and Honeycombe, R W K, *Steels: Microstructure and Properties*, pub. Butterworth-Heinemann, 3rd Ed. 2006.
[13] R.L. Klueh, K. Shiba, M. Sokolov, *Embrittlement of Irradiated Ferritic/Martensitic Steels in the Absence of Irradiation Hardening*, Proceedings of the 13th International Conference on Fusion Reactor Materials (ICFRM-13), December 2007
[14] Draper, N.R. and Smith, H. *Applied Regression Analysis,* 1998, Wiley-Interscience. ISBN 0-471-17082-8 http://en.wikipedia.org/wiki/Coefficient_of_determination
[15] Bishop C M, *Neural networks for pattern recognition*, 1996, Clarendon Press, Oxford.
[16] H.Lohninger: *Teach/Me Data Analysis*, Springer-Verlag, Berlin-New York-Tokyo, 1999: http://www.vias.org/tmdatanaleng/cc_ann_extrapolation.html
[17] MacKay D J K *Information Theory, Inference, and Learning Algorithms,* 2003, Cambridge University Press: Materials Algorithms Project Program Library: http://www.msm.cam.ac.uk/map/utilities/modules/nnecode.html
[18] Neuromat Ltd, *Model manager* 2004 : http://www.neuromat.com/model_manager.html
[19] Aha D Wand Bankert R L, *A comparative evaluation of sequential feature selection algorithms* , in D. Fisher & H. Lenz, eds, 1995, proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics, Ft. Lauderdale, FL, pp. 1--7. http://citeseer.ist.psu.edu/aha95comparative.html