

# Discovering Knowledge and Modelling Systems using Granular Computing and Neurofuzzy Structures

George Panoutsos and Mahdi Mahfouf  
Institute for Microstructural and Mechanical Process Engineering:  
The University of Sheffield (IMMPETUS)  
Department of Automatic Control and Systems Engineering  
Mappin St., Sheffield, S1 3JD, UK  
Phone: +44 (0)114 222 5607, Fax: +44 (0)114 222 5624  
Email: g.panoutsos@sheffield.ac.uk, m.mahfouf@sheffield.ac.uk

**ABSTRACT:** In this paper a new systematic modelling approach using Granular Computing (GrC) and Neurofuzzy modelling is presented. In this study a GrC algorithm is used to extract relational information and data characteristics out of the initial database. The extracted knowledge is then translated into a linguistic rule-base of a fuzzy system. This rule-base is finally realised via a Neurofuzzy modelling structure. The proposed methodology is then applied to the challenging environment of a multi-dimensional, non-linear and sparse data space consisting of mechanical properties of heat treated steel.

**KEYWORDS:** Granular Computing, Computing with Words, Fuzzy - Neurofuzzy Systems, Knowledge Discovery

## INTRODUCTION

Extracting knowledge out of data collections is the first and very critical step towards designing data driven soft-computing models. Data clustering techniques offer a simple way of finding relationships between data sets and grouping data together. Fuzzy C-Means (FCM) [1], the Mountain method [2] and probabilistic measures in clustering [3] are among the many clustering techniques that have been used to aid the design of fuzzy and neurofuzzy systems. The main drawback of such methods is that the quality of the solutions (partitions) depends on parameters like the initial values (initial cluster centres – FCM), on estimating the number and location of initial clusters' centres (mountain method) and on several other statistical considerations (such as: probabilistic measures for fuzzy clustering). Granular Computing (GrC) offers a simple and effective way of extracting information out of data sets, inspired by the human perception of grouping similar featured items together. By using GrC it is possible to group data together based on similar features and additionally retain information and data characteristics like granule size, density and orientation in the data space. The transparency offered by GrC and the additional information collected during the grouping process make this methodology ideal for combining it with fuzzy systems modelling. In this paper a neurofuzzy structure is proposed as the facet of the modelling process. The proposed modelling approach is demonstrated using a multi-dimensional, non-linear and sparse data space. This data space consists of mechanical properties data of heat treated steel. The aim of the modelling effort is for the model to be able to accurately predict steel properties (Tensile Strength – TS, Elongation, and Impact Energy) based on a transparent and interpretable structure.

## KNOWLEDGE DISCOVERY USING GRANULAR COMPUTING

Granular computing tries to mimic the perception and the societal instinct of humans when grouping similar items together. Data granulation [4], [5] is achieved by a simple two step iterative process that involves the following steps:

- Find the two most 'compatible' information granules and merge them together as a new information granule containing both original granules.
- Repeat the process of finding the two most compatible granules until a satisfactory data abstraction level is achieved.

The most important concept of the above process is the definition of the compatibility measure. This can be purely geometrical (distance between granules, size of granules, volume of granules), density driven (ratio of cardinality versus granule volume) or similarity driven. In this case the compatibility measure is a function of the distance between the granules and a function of the information density of the newly formed granule. A mathematical representation of the compatibility criterion is given in equation 1.

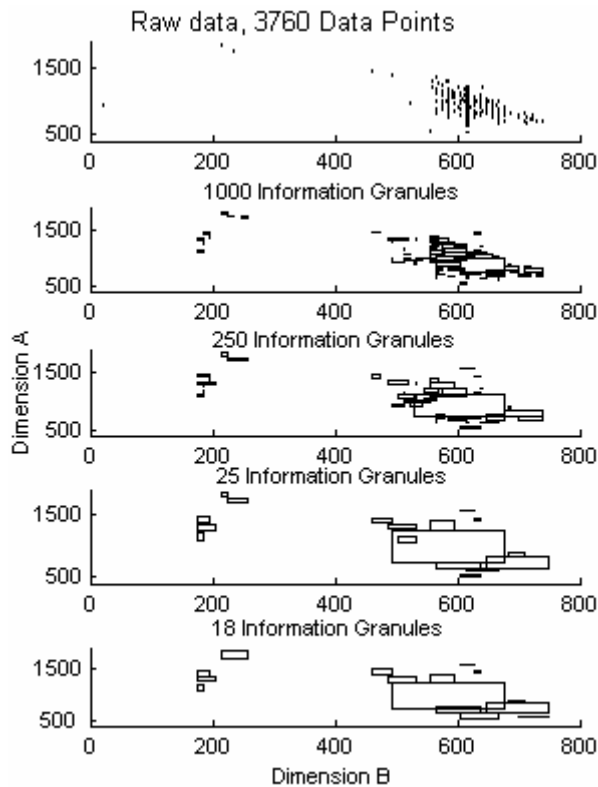
$$\text{Compatibility} = f(w_1 \cdot \text{Distance}, w_2 \cdot \text{Density})$$

$$\text{Distance} = \sum_{i=1}^k (\text{position of granule}_b - \text{position of granule}_a)_i \quad (1)$$

$$\text{Density} = \frac{\text{granule cardinality}}{\text{granule volume}} = \frac{\text{no. of sub-granules}}{\prod_{i=1}^k \text{granule length}_i}$$

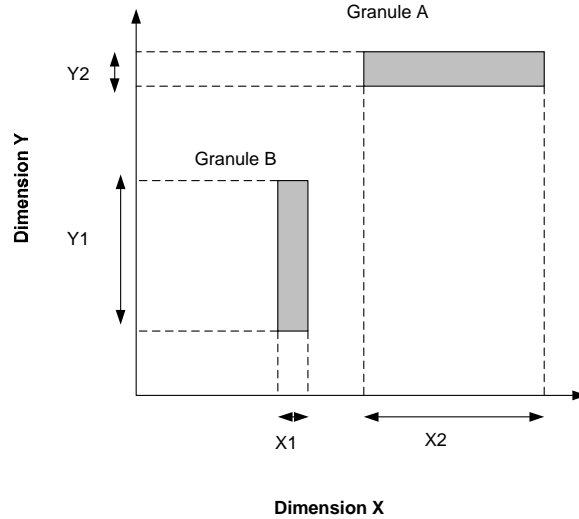
Where  $w_1, w_2$  are weights for balancing the distance/density requirements and  $k$  the dimensionality of the data space.

Even though this process can be accidentally identified as hierarchical clustering there is a major difference; each granule consists of the same objects (sub-granules). In hierarchical clustering new objects are created and the boundaries of the new clusters can be in an area where no data are present. The growth of clusters allows strong linkage between the original data set (transparency) and it allows visual monitoring methods for terminating granulation. By considering the merging of each set of granules as some information condensation or information loss, it is possible to link information loss to the 'merged distance' and therefore plot an information loss graph. This graph can be used on-line or off-line as a criterion for terminating the granulation process.



**Figure 1: Iterative Information Granulation**

Figure 1 shows snapshots of a 2-dimensional data granulation. Dimension A has units between 0-2000 and Dimension B between 0-800. The first snapshot, shown at the top of the figure, is the representation of the raw, pre-granulated, data, consisting of 3760 data points. As the iterative granulation algorithm progresses, snapshots of the granules are shown, consisting of 1000 granules (2<sup>nd</sup> snapshot of data), 250 granules, 25 and finally 18 granules. The added information collected during the granulation process is stored, and it consists of the cardinality and the multidimensional length of each granule. These data are going to be used for creating a linguistic rule-base, as it is shown in the next section.



**Figure 2: Orientation of granules**

The following granulation example shows how the transparency and the additional information of the GrC process can assist the modelling process. Granules A and B (figure 2) have same size, cardinality and density, but are of different orientation (90deg difference). Considering the fuzzy modelling GrC structure, every granule consists of a linguistic rule; in this case:

$$\begin{aligned} GrA: & \text{ if } X = X2 \text{ then } Y = Y2 \\ GrB: & \text{ if } X = X1 \text{ then } Y = Y1 \end{aligned}$$

Rule GrA has more output sensitivity (narrow space) and less input sensitivity (wide space). On the other hand, GrB has more input sensitivity rather than output one. By directing the algorithm towards solutions of type A or type B granules it is possible to enhance the model's performance, i.e. by increasing the sensitivity of an important variable (this has been practically demonstrated using simulation experiments).

## FORMATION OF A FUZZY RULE-BASE

Consider the granulation of a database provided by a multi-input single-output (MISO) system. By granulating across each input dimension individually and at the same time across the whole input space it is possible to identify relational information (rules) similar to a Mamdani FIS rule-base:

$$\begin{aligned} \text{Rule 1:} \\ \text{if (inputA = } A_1 \text{ and inputB = } B_1 \text{ and...)} \\ \text{then(output = } O_1 \text{)} \\ \text{Rule 2:} \\ \text{if (inputA = } A_2 \text{ and inputB = } B_2 \text{ and...)} \\ \vdots \\ \text{then(output = } O_2 \text{)} \end{aligned} \quad (2)$$

Where  $A_i, B_i, \dots, O_i$  are information granules discovered during the GrC process and  $i$ : is the number of granules.

The information captured using GrC defines the initial structure of the fuzzy rule-base, number of rules and initial location and width of membership functions (MFs). For extracting the location of each MF the centre of the each granule can be used, and the width can be determined using the density of each granule (cardinality, length). This rule-base is going to be used as the initial structure of a neurofuzzy system.

## NEUROFUZZY STRUCTURE

Consider a MISO fuzzy logic system as described by equation 2. The consequent part of each linguistic rule (...then: the output is  $O$ ) can be:

- a)  $O$ : Fuzzy Set (Mamdani rule-base)
- b)  $O$ : Singleton (Mamdani Singleton)
- c)  $O$ : Lineal Function (Takagi-Sugeno-Kang, TSK)

Fuzzy logic systems having centre of gravity (COG) defuzzification, product inference rule and singleton fuzzy output space can be expressed with the following mathematical equation:

$$y = \sum_{i=1}^p z_i \left[ \frac{\prod_{j=1}^m \mu_{ij}(x_j)}{\sum_{i=1}^p \prod_{j=1}^m \mu_{ij}(x_j)} \right] \quad (3)$$

Where  $\mu_{ij}(x_j)$  is the Gaussian membership function of  $x_j$  that belongs to the  $i$ -th rule.

$$\mu_{ij}(x_j) = e^{-\left( \frac{(x_j - c_{ij})^2}{\sigma_{ij}^2} \right)} \quad (4)$$

Where  $c_{ij}$  and  $\sigma_{ij}$  are the centre and the width of each membership function respectively,  $m$  the number of inputs and  $p$  number of rules.

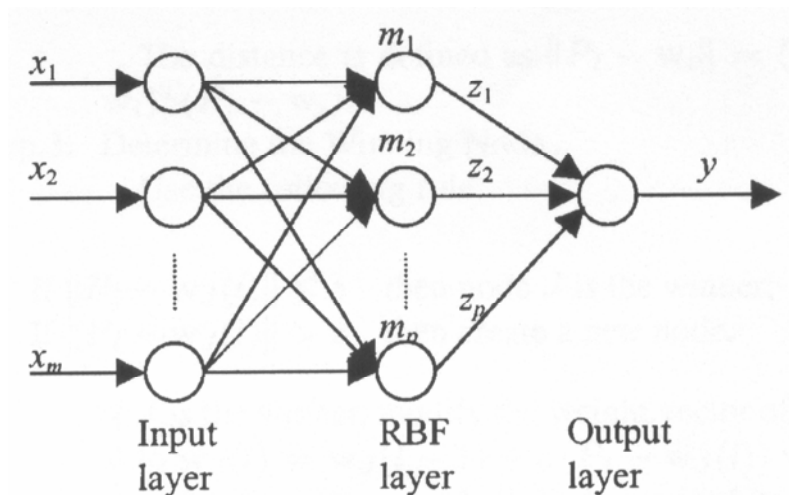


Figure 3: Neurofuzzy Structure

Equation (3) can be written as:

$$y = \frac{\sum_{i=1}^p z_i m_i(\underline{x})}{\sum_{i=1}^p m_i(\underline{x})} \quad (5)$$

Where  $m_i(\underline{x}) = e^{(-\|\underline{x}-c_i\|^2/\sigma_i^2)}$  is the degree of membership of the current input vector  $\underline{x}$  to the  $i$ -th fuzzy rule. Finally, using the radial Basis Function (RBF) definition:

$$g_i(x) = \frac{m_i(x)}{\sum_{i=1}^p m_i(x)} \quad (6)$$

The neurofuzzy input-output relationship is:

$$y = \sum_{i=1}^p z_i g_i(x) \quad (7)$$

The neurofuzzy training procedure starts by initialising the centres and widths of each fuzzy weight. This can be achieved by taking the equivalent values from the GrC rule-base structure. The neurofuzzy system is then trained using an adaptive weighted back-error-propagation (BEP) algorithm [6]. This modelling procedure is tested against real data of an industrial process and the results are presented in the next section.

## EXPERIMENTAL STUDIES

Determining the optimal heat treatment regimes and weight percentage of composite materials of steel is a common but not trivial task in the steel industry. By predicting correctly the optimal conditions, it is possible to obtain the required steel grade with accuracy and at a reduced cost. The modelling procedure is based on both chemical composition and heat treatment data. This is a multiple-input single-output (MISO) process that is difficult to model due to the following reasons:

- a) Non-linear behaviour of the process,
- b) High interaction between the multi-variable input spaces,
- c) Measurement uncertainty of the industrial data,
- d) High complexity of the optimisation space, and
- e) Sparse data space

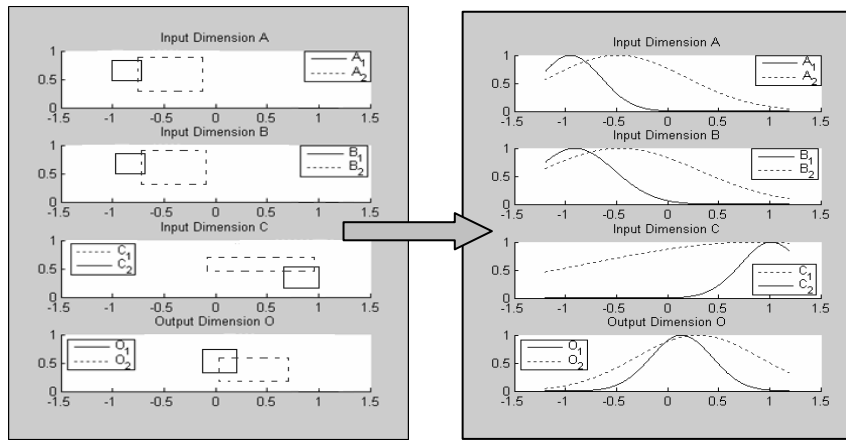
Black box modelling techniques [7] are usually employed to tackle this problem with an acceptable level of performance but fail to make use of experts' (metallurgists) knowledge that will prove to be very valuable. The technique presented in this paper offers a good performance level while maintaining the high transparency of the system during the modelling process. System transparency is often desirable in these kinds of systems as a better understanding of the physical process can be achieved and experts' knowledge can be used along with the model or embedded into the model.

A highly dimensional data set taken from the steel industry is used for modelling purposes. Each set of points represents 15 input variables and 1 output variable. The input variables include both: a) the chemical composition of steel (i.e. % content of C, Mn, Cr, Ni etc.) and b) the heat treatment data (Tempering temperature, Cooling medium etc.). The output variable is the desired steel property to be modelled – predicted. The steel properties that are going to be modelled are: Tensile strength (TS), Elongation, and Impact Energy.

### TENSILE STRENGTH

The TS data set consists of 3760 data vectors. This data set has been cleaned for spurious or inconsistent data points. The dimension of the data space is 16 (15-inputs 1-output). The data space, apart from being highly non-linear and

complex, is also very sparse. This is because the industrial data are focused towards specific grades of steel; therefore there are some discontinuities in most of the input dimensions. The modelling procedure that is followed is the one described in this paper, hence the first step consists of the knowledge extraction out of the database. Using the GrC algorithm presented in the first section, the rule-base is obtained in the form of equation 2.

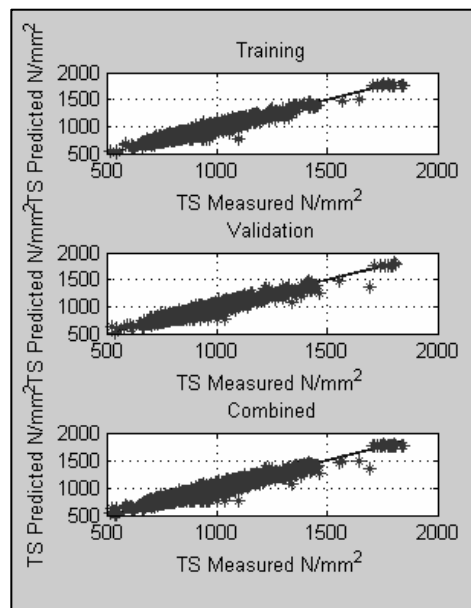


**Figure 4: GrC Rule-Base / Fuzzy Rule-Base**

Figure 4 shows a graphical representation of how the GrC rule-base is translated into a fuzzy rule-base. One rule is shown (one granule) having 4 dimensions – for simplicity – ( 3-inputs 1-output). The following laws apply:

- The GrC and Fuzzy MFs centres are the same,  $C_{GrC} = C_{MFs}$
- The width of each fuzzy membership function is obtained using the granule's cardinality and size,  $\sigma_{MFs} = f(\text{cardinality}, \text{size})$
- One granule (multi-dimensional) corresponds to one fuzzy rule

After obtaining the rule-base (9-rules), the neurofuzzy model is initialised and optimised using the adaptive BEP algorithm shown in the previous section. The prediction performance of the model, and the fit plot (predicted vs. measured) is shown below, consisting of both training and validation data sets.



**Figure 5: Data Fit - Tensile Strength**

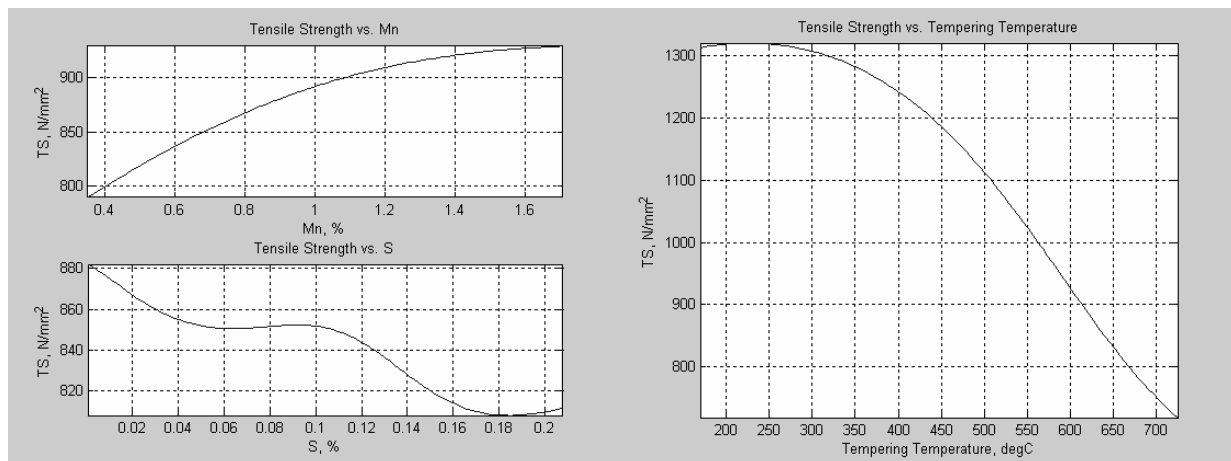
The performance of the model is shown below, in RMSE:

**Table 1: Tensile Strength Model's Performance**

	RMSE – Tensile Strength Model	Confidence Band (+/-)
Training	38.62	33.43
Validation	41.87	35.12
Combined	40.28	34.01

As it was expected, the performance level of this GrC-Neurofuzzy model is better than the performance of equivalent pure fuzzy models, as presented in [5], but at the same time this modelling structure maintains its transparency and simplicity. As compared to black-box modelling methodologies, for an equivalent process [7], the GrC-Neurofuzzy modelling performance is of the same level or better than previous results, with the considerable advantage over black-box models of the added transparency and simplicity.

Additionally, the transparency and the physical interpretation of the model can be verified using variable effect plots. This can be achieved by keeping  $(m-1)$  input variables constant ( $m$ : no of total input variables), and plotting the remaining varying variable against the output. The  $(m-1)$  variables that are kept constant are usually set to their respective *average* values, but in this case this is not possible. As it was mentioned at the beginning of this section the data set is very sparse, and the industrial data collection was based on *steel grades*. Therefore the average value of a particular input dimension might not correspond to an actual (meaningful) value. Hence, it was decided to use the average values of the dominant steel grade (as in [7]), which is the 1%CrMo steel grade. Some of the variable effect plots are shown in figure 6.



**Figure 6: Variable Effect Plots, TS**

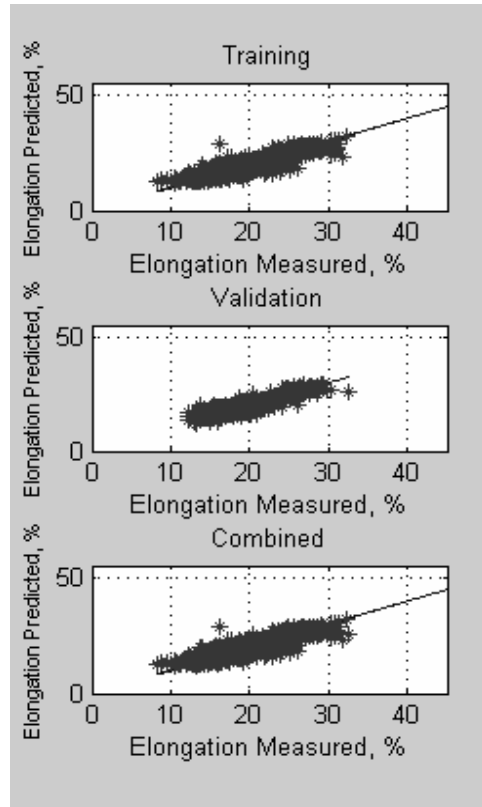
For simplicity, three (Mo, S, Tempering Temperature) out of the possible fifteen variable effect plots are shown. All variable effect plots are verified that are following the expected behaviour (within reasonable tolerance limits) as predicted by theory or by expert's (metallurgist's) knowledge. For example, the Tempering Temperature (TT) is inversely proportional to the Tensile Strength (TS) as it is predicted by theory and experts and shown in figure 6, right hand side plot.

#### ELONGATION AND IMPACT ENERGY

For verifying the generalisation of the modelling structure, three further steel properties have been investigated, Elongation, Impact Energy and Reduction of Area (ROA). Sample performance results and variable effect results are shown in the following pages.

## Elongation

The modelling performance of the elongation model follows the same performance pattern as the TS model. The model achieves better performance than the equivalent pure Fuzzy models (as in [5]) and achieves same or better performance with added transparency and simplicity as compared to equivalent black-box modelling structures (as in [7]). The variable effect investigation is performed in the same manner as in the previous section (Tensile Strength) and the interpretability of the results follows the theory i.e. the Tempering temperature is proportional to the Elongation. Figures 7,8 and table 2 demonstrate these results. It has to be noted that the input space dimensionality of the Elongation data set is increased by one, as the 'gauge length' had to be included in the data set.



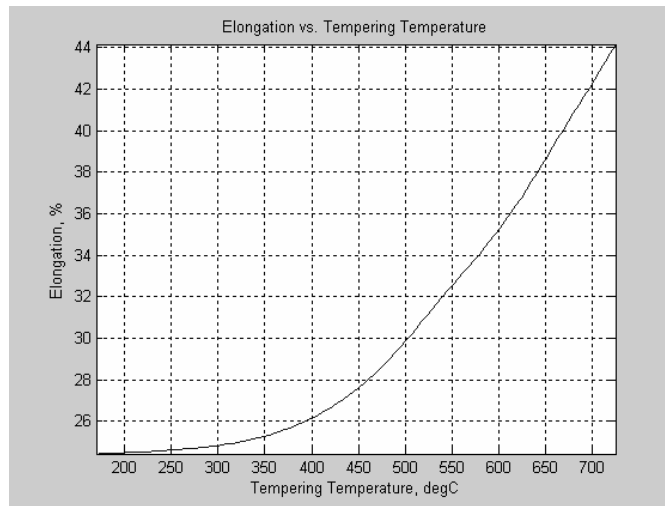
**Figure 7: Data Fit - Elongation**

And the performance in RMSE, is:

**Table 2: Elongation - Model's Performance**

	RMSE – Tensile Strength Model	Confidence Band (+/-)
Training	1.53	1.20
Validation	1.46	1.25
Combined	1.49	1.22

The inconsistency between the training and validation performance (Validation < Training) is not a feature of the modelling structure but a result of the inconsistent data distribution of the industrial data set. This behaviour was also experienced in [7] with the same data set.

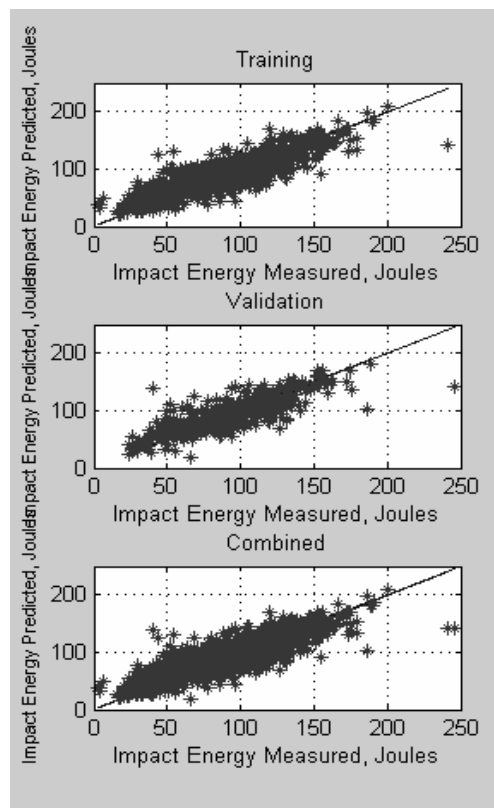


**Figure 8: Variable Effect, Elongation Vs. TT**

The Elongation Vs. TT variable effect shows that the elongation is proportional to the TT. This result is consistent with the theory, and combined with the equivalent TS result it is possible to realise the contradictory behaviour between Elongation and TS as predicted by theory.

### *Impact Energy*

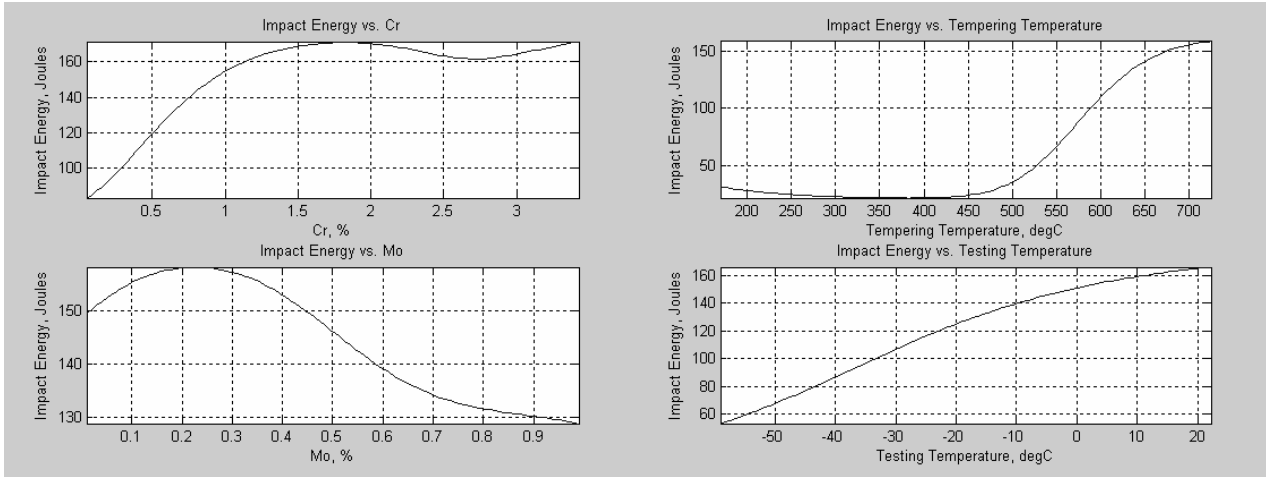
Impact energy is one of the most difficult properties in the steel making process to model, and this is not only because the impact property is highly non-linear in relation to the steel composition and heat treatment regime, but also because the test types used are often incompatible. Despite the difficult nature of the problem the performance of the GrC-Neurofuzzy model is very good and the interpretability of the model gives good insight into the non-linear interaction of the variables. The results are shown below:



**Figure 9: Data Fit – Impact Energy**

**Table 3: Impact Energy - Model's Performance**

	RMSE – Tensile Strength Model	Confidence Band (+/-)
Training	16.45	14.12
Validation	19.19	16.36
Combined	17.18	15.92



**Figure 10: Impact Energy - Variable Effect**

The highly non-linear interaction between the variables and the impact energy is shown. The variables shown are: Cr, Mo, Tempering Temperature and Testing Temperature.

## CONCLUSION

Neurofuzzy modelling and knowledge discovery, within databases containing complex relationships, using Granular Computing is discussed in this paper. By taking advantage of the ability of GrC to capture knowledge in a transparent and linguistic way and combining this knowledge with a Fuzzy inference engine (Neurofuzzy Structure) it is possible to build very transparent systems capable of modelling various complex processes.

The initial database has to be cleaned (spurious points, inconsistent points etc) before it is processed by a GrC algorithm. The algorithm then works iteratively to capture all the required relational data and extra data features, so that the rule-base is assembled. The final rule-base is expressed in a linguistic form, which is ideal for use with fuzzy systems. After the rule-base is translated from granules to fuzzy membership functions the data are assigned to a Neurofuzzy structure. This structure can be of Mamdani, Singleton or TSK types. One can select between the possible structures depending on the performance / transparency-interpretability requirements. Finally, the Neurofuzzy structure is optimised using an adaptive weighted BEP algorithm.

The proposed modelling methodology has been tested against real industrial data of high dimensionality and complex nature. The resulting models achieved better performances than pure fuzzy models. Comparable or better performance was experienced (depending on structure selection) as compared with black-box modelling techniques (i.e. Neural Networks), but higher transparency and interpretability were possible during the modelling process. Due to the transparency of the process, experts' knowledge can be used during the modelling process for deciding various modelling or optimisation parameters, which lead to improved performance of the final model. Additionally, the transparency of GrC will always help to develop in the area of incremental learning, and system adaptation via fuzzy systems. By granulating and capturing information of new data sets using GrC it is possible to explore the technique of combining new knowledge with the existing rule-base without any significant loss of performance.

## ACKNOWLEDGMENTS

The authors would like to thank the Engineering and Physical Sciences Research Council (EPSRC), UK for their financial support under grant GR/R70514/01 and Corus Engineering Steels (CES) Stocksbridge, Sheffield, UK for their help and in providing the heat treatment data.

## REFERENCES

- [1] Bezdek, J. C., 1981, "Pattern Recognition With Fuzzy Objective Function Algorithms", Plenum Press, New York, NY.
- [2] Yager R. R. and D. P. Filev, 1994, "Approximate Clustering via the mountain method", IEEE Transactions on Systems, Man, and Cybernetics, Vol 24, No. 8, pp 1279-1284, 1994
- [3] Brailovsky, V. L., 1991. "A probabilistic approach to clustering. Pattern Recog. Lett.12, 4 (Apr. 1991), 193–198.
- [4] Bargiela, A and W Pedrycz , 2003," *Granular Computing An Introduction*", Kluwer Academic Publishers, 2003
- [5] Panoutsos, George and Mahdi Mahfouf, 2005, "Granular Computing and Evolutionary Fuzzy Modelling for Mechanical Properties of Alloy Steels", IFAC 2005, Proceedings of, 16<sup>th</sup> IFAC World Congress, July 4-8 2005, Prague Czech Republic, International Federation of Automatic Control (IFAC)
- [6] Chen, Min-You and D.A. Linkens, 2001, "A Systematic Neuro-Fuzzy Modelling Framework With Application to Material Property Prediction", IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol.31, No.5, October 2001, pp. 781-790
- [7] Tenner, J., 1999, "*Optimisation of the Heat Treatment of Steel using Neural Networks*", Ph.D. Thesis, Department of Automatic Control and Systems Engineering, University of Sheffield, UK.