Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

A New Interval Type-2 Fuzzy Clustering Algorithm for Interval Type-2 Fuzzy Modelling with Application to Heat Treatment of Steel

Olusayo Obajemu, Mahdi Mahfouf and L. A. Torres-Salomao

Department of Automatic Control and Systems Engineering, The University of Sheffield Sheffield, S1 3JD, UK (e-mail: oobajemu1@sheffield.ac.uk; m.mahfouf@sheffield.ac.uk; latsalomao@ieee.org).

Abstract: In this paper, a new hierarchical data-driven modelling strategy based on Interval Type-2 Fuzzy Clustering is elicited for the Interval Type-2 Takagi-Sugeno-Kang (TSK) Fuzzy Logic System. This framework which we have called the IT2-Squared framework uses interval type-2 fuzzy clustering for initial antecedent parameters and structures determination and least-squares algorithm for deriving initial consequent parameters. To improve the accuracy of the system, we show how the steepest descent algorithm is used to tune the parameters of both the consequent and antecedent parameters. To test the efficacy of this proposed system, the model is used on a real-life engineering project for the prediction of the ultimate tensile strength (UTS) of steel. Results show excellent generalization properties of the IT2-Squared modelling framework when compared to previously elicited models of the same system.

Keywords: Clustering, Steepest Descent, Fuzzy, Modelling, Steel.

I. INTRODUCTION

Fuzzy rule-based modelling has been shown to be particularly effective when expert knowledge of a system is vast and/or when an interpretable system is sought from the data. By expert knowledge, we mean those systems where we have extensive human understanding (Wang *et al*., 2012a). There exists a plethora of techniques for modelling type-1 fuzzy logic systems (T1 FLSs) in the literature. However, Type-2 fuzzy modelling is a relatively new area of research because it was not until relatively recently that the complete theory of type-2 fuzzy sets ($T2\,FSs$) and systems was developed (Karnik *et al.*, 1999 and Mendel, 2001) even though they had been introduced as far back as 1975 (Zadeh, 1975) to tackle the paradox of handling uncertainties with crisp fuzzy membership functions (MFs) in Type-1 Fuzzy Sets (T1 FSs). A T2 FS has a MF that is itself a fuzzy set and a type-2 fuzzy logic system (T2 FLS) is a FLS with at least one of the membership functions (MFs) being a T2 FS (Karnik *et al.,* 1999).

This extra dimension allows T2 FLSs to potentially provide better modelling capabilities than T1 FLSs especially in the presence of system noise and uncertainties. The price to pay for this better uncertainty handling is that T2 FLSs are more computationally demanding than their T1 counterparts mainly because of the added block in the output processing stage called the type-reducer (TR) (Fig. 1). However, this increased computational burden might be a small expense to pay when a more robust system is desired. Moreover, computational feasibility of T2 FLS is greatly enhanced by using the interval type-2 approach (Liang *et al*., 2000).

The purpose of this paper is to provide a new relatively computationally inexpensive and systematic approach to modelling an Interval Type-2 Fuzzy System (IT2 FLS) from data. We start by determining the structure of the system by clustering the data following a similar approach used by Delgado *et al*., 1997. However, here, we take into consideration a source of uncertainty involving the fuzzifier m which regulates the degree of overlap amongst the clusters when fuzzy c-means (FCM) clustering technique is used to elicit the initial structure of fuzzy models.

To achieve this, we follow a similar procedure introduced in Choi *et al.,* 2009 by making use of the interval type-2 fuzzy clustering algorithm developed by Rhee *et al*., 2007 to elicit the initial structure of the antecedent structure of the interval type-2 fuzzy logic system (IT2 FLS). This is a detour from the popular method of randomly initialising the width of the IT2 Fuzzy Sets (IT2 FSs) after having used FCM for the initial structure determination with the fuzzifier value set to a constant value (usually set to 2) (Delgado *et al*., 1997 and Wang *et al*., 2012b).

It is argued in this report that this systematic approach to finding these initial parameters has the potential to helping one build a more optimal fuzzy system especially when no further parameter tuning is performed. Even when the initial system is further optimised, local optima methods may return solutions that are equal or not far off from global optimum solutions. (Park *et al*., 2001) i.e. by making use of IT2 fuzzy clustering, good clustering results are returned and invariably good modelling performances.

Two most popular FLSs used today are the Mamdani and Takagi-Sugeno-Kang (TSK) FLSs (Mendel, 2001). Our paper is focused on the latter. Methods based on orthogonal least-squares (Mendel, 2001) and fuzzy clustering (Babuska *et al.*, 1995) have particularly found interesting applications in modelling TSK fuzzy systems.

Our approach follows that obtainable in the method proposed in Delgado *et al*., 1997 and Babuska *et al*., 1995 which involves first finding the structure of model using FCM clustering and further optimising these parameters. They applied this technique to a T1 FLS while we apply it to an IT2 FLS. The modelling paradigm involves first selecting the initial antecedent structure using IT2 fuzzy clustering and then the least-squares approach is used to derive initial consequent parameters by making use of the Nie-Tan (NT) defuzzification method (Nie *et al.,* 2008). Finally, the whole system is optimised using steepest descent.

In building this framework, we have made some necessary assumptions. For example, in deriving the steepest descent algorithm, we have used the product t-norm to find the firing interval of each input rules antecedent. This will be made clearer by some sets of equations in the succeeding sections.

The rest of this paper is organised as follows: Section II briefly introduces IT2 FLSs and the methodologies involved in IT2 fuzzy clustering. The methodologies involved in this new data modelling framework are presented in section III. Results and findings are then succinctly presented in section IV. And the conclusion of this paper and future works presented in section V.

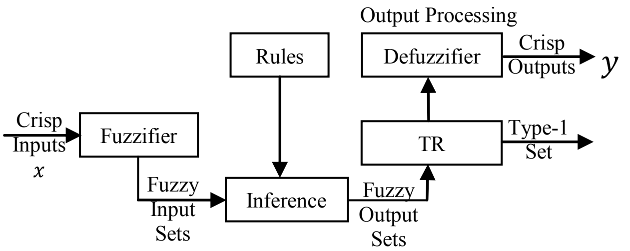

Fig. 1. Conventional Type-2 FLS.

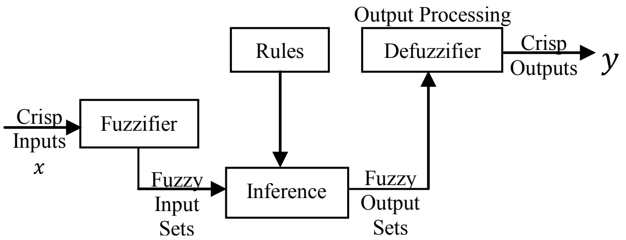

Fig. 2. Type-2 FLS with Nie-Tan Defuzzification.

II. INTERVAL TYPE-2 FUZZY SYSTEMS

A. Interval Type-2 Fuzzy Logic Systems

(i) Interval Type-2 Fuzzy Sets

Because of the much reduced computational burden, we have used the IT2 FLS. The IT2 FLS has the same structure as that of the general T2 FLS, but instead of using general T2 FSs, the fuzzy sets are IT2 FSs. An Example of an IT2 FS is given in fig. 3a. IT2 FSs are still able to handle uncertainties as their membership functions are no longer crisp but a type-1 interval as shown in fig. 3b. Mathematically, an $IT2\ FS$ can be expressed as follows:

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x \subseteq [0,1]} 1/(x,u) \qquad (1)$$

Where x is the primary variable and its measurement domain denoted by X; u is the secondary domain variable $u \in J_x$ at each $x \in X$, J_x is the primary membership of x. The footprint of uncertainty (FOU) is the union of all the embedded T1 FSs which is marked by the grey area in fig. 3a. This grey area is bounded by an upper membership function (UMF) and a lower membership function (LMF). The primary MFs of an IT2 FLS may be any of the convex IT2 FSs such as the triangular, trapezoidal or Gaussian. However, the latter is usually utilised in fuzzy modelling as it meets the requirements of the continuity and smoothness of mapping when using FLSs as a universal approximator (Wang *et al*., 2012b).

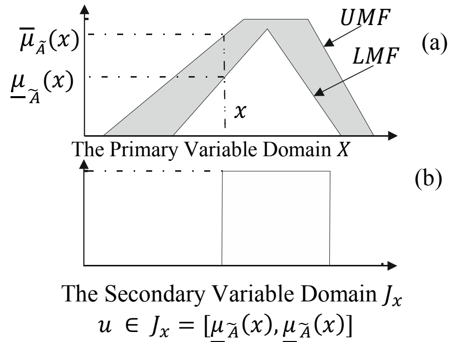

Fig. 3. An example of an IT2 Fuzzy Set.

In this research, our proposed modelling framework uses the Gaussian primary membership function (MF) of fixed mean and uncertain spread i.e.

$$\mu_{\tilde{A}}(x) = exp\left[-\frac{1}{2}\left(\frac{x-v}{\sigma}\right)^2\right] \quad \sigma \in [\underline{\sigma}, \bar{\sigma}] \qquad (2)$$
$$\underline{\mu} = N(v, \underline{\sigma}; x) \quad and \quad \bar{\mu} = N(v, \bar{\sigma}; x)$$

Where $\underline{\mu}$ and $\bar{\mu}$ are its associated LMF and UMF.

(ii) Inference

Given an IT2 TSK FLS with n inputs, $x_1 \in X_1$, $x_2 \in X_2, \dots, x_n \in X_n$, and one output $y \in X$, and the rule based composing of c rules, with the ith rule $\tilde{R}^i$ expressed as:

$\tilde{R}^i$: IF x_1 is $\tilde{A}_1^i$ and x_2 is $\tilde{A}_2^i$ ... and x_n is $\tilde{A}_n^i$, THEN y_i is $h_i(\boldsymbol{x}) \quad \boldsymbol{x} \in R^n$

$\tilde{A}_j^i$ and h_i represent the jth antecedent IT2 membership function and the consequent of the ith rule for $j = 1, 2, \dots n$ and $i = 1, 2, \dots c$ respectively. The primary MF of the jth antecedent is denoted by $\mu_{\tilde{A}_j^i}(x)$. $\tilde{A}_1^i, \tilde{A}_2^i \dots \tilde{A}_n^i$ are IT2 antecedent fuzzy sets. $h_i(\boldsymbol{x})$ is a crisp value defined as follows:

$$h_i(\boldsymbol{x}) = \beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 + \cdots + \beta_{in}x_n \qquad (3)$$

Where $\beta_{ip}(p = 0, 1, 2, \dots n)$ are the consequent parameters for $i = 1, 2, \dots c$.

(iii) Type Reduction and Defuzzifcation

This is the last block of an IT2 FLS and it is where the more tasking computational effort of an IT2 FLS is incurred. In a TSK FLS (T1 or T2) each fired rule, denoted by f_i, represents a scaled FS in the output domain. The defuzzifier obtains a single set by using a certain weighted average method.

Conventionally, an ITS FLS is that obtained in fig. 1. The TR stage involves reducing the T2 FSs into a T1 FS at the consequent. If IT2 FSs are used in the system, then this reduces to a T1 interval after TR. To find this type reduced set, the KM algorithms are used to find the left (L) and right (R) end points. These two points are then averaged to find the final output of the system. Closed form of an IT2 FLS is practically impossible. (Karnik *et al.,* 1999).

The second approach which allows finding a closed form solution of the IT2 FLS was developed by Nie *et al,* 2008 and is used in this research. Using this method the output processing block reduces to a single block of output processing. The closed form for this IT2 FLS may be expressed by (4). Unfortunately, this simple closed form of an IT2 FLS loses its ability to ascertain uncertainty involved in the final computed output values.

$$\hat{y} = \left(\sum_{i=1}^{c} h_i \bar{f}_i + \sum_{i=1}^{c} h_i \underline{f}_i\right) \Big/ \left(\sum_{i=1}^{c} \bar{f}_i + \sum_{i=1}^{c} \underline{f}_i\right) \qquad (4)$$

Where $\hat{y}$ is the final defuzzified value, h_i is as prior defined.

For the Gaussian primary MF of fixed mean and uncertain spread and product T-norm, $\underline{f}_i$ and $\bar{f}_i$ are defined as:

$$\underline{f}_i = \prod_{j=1}^{n} \underline{\mu}_{ij}, \quad \bar{f}_i = \prod_{j=1}^{n} \bar{\mu}_{ij} \qquad (5)$$

or $j = 1, 2, \dots n$ and $i = 1, 2, \dots c$. $\underline{\mu}_{ij}$ is the primary membership grade of the jth antecedent of the ith rule.
In matrix form, (4) may be rewritten as follows:

$$\hat{y} = \left(\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{H} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{H}\right) / \left(\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}\right) \qquad (6)$$

Where ***r*** is a cx1column vector of 1's. ***F*** is a cx1 vector.

B. Interval Type-2 Fuzzy Clustering

The IT2 fuzzy clustering proposed by Rhee *et al.*, 2007 presents a systematic way of choosing the widths of the initial IT2 FSs. A brief illustration of this technique as shown in fig. 4, involves 'determining LMF and UMF' which automatically helps one select the initial width, centre and spread of the antecedent MFs.

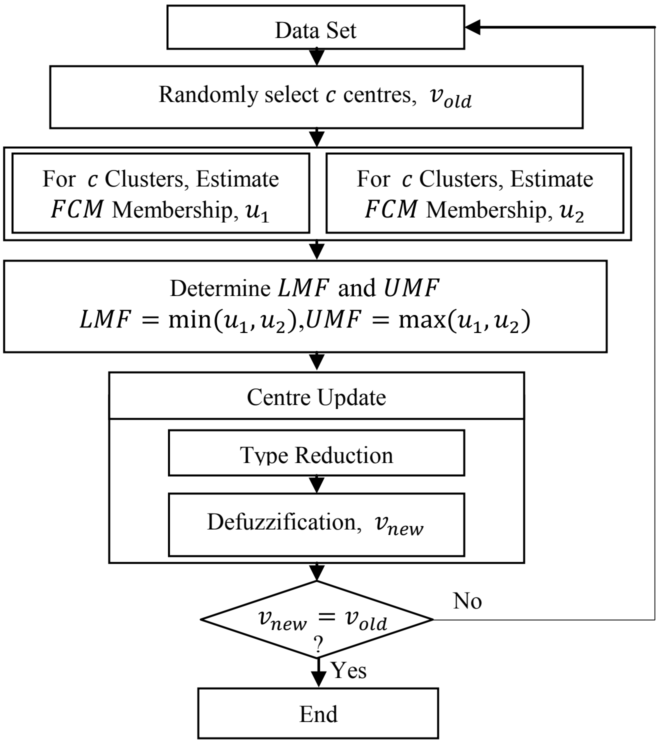

Fig. 4. Interval Type -2 FCM Algorithm.

III. METHODOLOGY

The IT2-Squared has 3 stages as shown in fig. 5. We now describe each of the stages in this new IT2 FLS modelling framework.

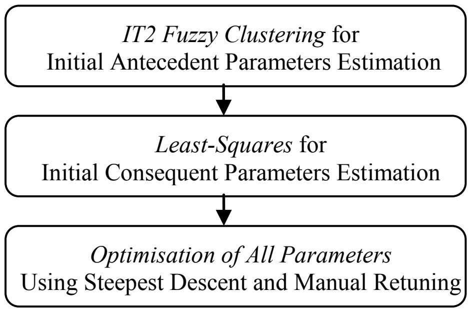

Fig. 5. Stages involved in our IT2 FLS Modelling framework.

A. Stage 1- Initial Antecedent Structure and Parameters

As mentioned earlier, in this stage of the modelling process, the initial structure and parameters of the antecedent MFs are obtained through the use of the IT2 FCM algorithm. The parameters of these antecedent MFs are found by projecting the UMF and LMF found from the fuzzy clusters to each input subspaces. This approach follows directly from that obtainable in Babuska *et al.*, 1995 which uses this approach to find the antecedent initial parameters of a T1 FLS. The centres and widths (consequently the variances) are gotten from the fuzzy means and fuzzy variances respectively of each input dimension after clustering. For detailed discussions of how these antecedent parameters may be found from the input subspaces projections, the reader is advised to consult Babuska *et al.*, 1995 and Delgado *et al*, 1997. Fig. 6 shows a synthetic data in two dimensions and the projections in each dimension for a fixed mean and uncertain standard deviation IT2 FS after using the IT2 FCM to cluster the data (2 clusters/2 rules).

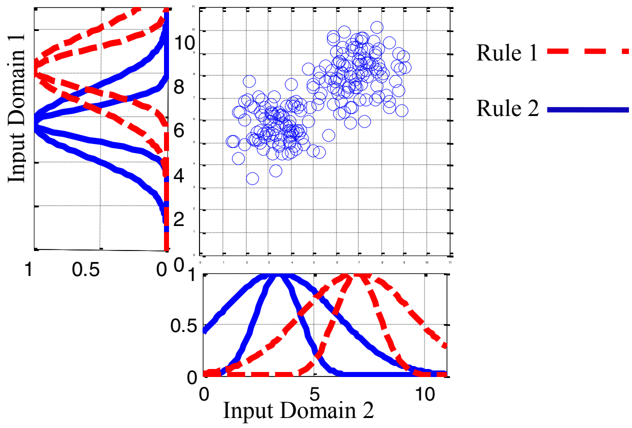

Fig. 6 Synthetic data projections.

B. Least-Squares for the Consequent Part

The least-squares and its variants are popular choices for finding the consequent parameters of a T1 TSK FLS. However, its use in finding that for a T2 FLS has been somewhat limited owing the non-closed-form-compatibility of T2 FLS when we are using the iterative KM algorithm in finding the final output of the T2 FLS. N-T method of defuzzification may help one solve this problem since it is very similar to that of T1 FLS. Below, we show how, when using the N-T method of defuzzification, and after having found the initial antecedent parameters, we may find the consequent parameters of an IT2 FLS using least-squares. (4) may be re-expressed as follows:

$$\hat{y} = \left(\sum_{i=1}^{c} h_i\left(\bar{f}_i + \underline{f_i}\right)\right) / \left(\sum_{i=1}^{c}\left(\bar{f}_i + \underline{f_i}\right)\right) \tag{7}$$

Where
$$h_i = B_i^T \tilde{X} \tag{8}$$

Let
$$\tilde{f}_i = \bar{f}_i + \underline{f_i} \tag{9}$$

Also Let
$$\omega_{ip} = \tilde{x}_p \tilde{f}_i \Big/ \sum_{i=1}^{c} \tilde{f}_i \tag{10}$$

for $p = 0, 1, \ldots, n$, then

$$\hat{y} = B^T \varphi \tag{11}$$

Where φ is a column is vector of ω_{ip} for $i = 1 \ldots c, p = 0, 1, \ldots, n$ and B is a column vector of associated parameters. Least- squares parameter estimation leads to,

$$\hat{B} = (\Phi^T \Phi)^{-1} \Phi^T Y \tag{12}$$

Where Y is a column vector of measured output data is Φ is the design matrix.

C. Stage 3- Steepest Descent Optimisation

In Mendel 2004, the derivatives for an IT2 FLS when making use of the KM algorithm to type- reduce the FLS were derived. In this paper, following similar procedures, we derive the steepest descent algorithm for an IT2 FLS but this time using the NT method of defuzzification. The disadvantage of using this type of defuzzification is that the measure of output uncertainty we get when using the KM algorithm is absent. All derivations may be found in the Appendix. We summarise our final answers in this section. For the particular antecedent or consequent parameter τ, its update formula is:

$$\tau(t+1) = \tau(\mathrm{t}) - \alpha_\tau \frac{\partial \mathrm{e}}{\partial \tau} \tag{13}$$

$$e = \frac{1}{2}[\hat{y} - y]^2 \tag{14}$$

Where $\hat{y}$ is the model output and y is the real output. The challenge is to find the derivatives of e with respect to the design parameter we wish to update.

(i) Antecedent Parameters

The update formula for the antecedent parameters (Gaussian primary MF of fixed mean and uncertain spread) is:

$$\frac{\partial e}{\partial \theta_{ij}^l} = (\hat{y} - y)\left(\frac{hi - \hat{y}}{\overline{\boldsymbol{F}}^T\boldsymbol{r} + \underline{\boldsymbol{F}}^T\boldsymbol{r}}\right)\left(\left[\prod_{\substack{q=1\\q\neq j}}^{n} \bar{\mu}_{iq}\right]\frac{\partial \bar{\mu}_{ij}}{\partial \theta_{ij}^l} + \left[\prod_{\substack{q=1\\q\neq j}}^{n} \underline{\mu}_{iq}\right]\frac{\partial \underline{\mu}_{ij}}{\partial \theta_{ij}^l}\right) \tag{15}$$

$$\frac{\partial \underline{\mu}_{ij}}{\partial v_{ij}} = \frac{(x_j - v_{ij})\boldsymbol{N}(v_{ij}, \underline{\sigma}_{ij}; x_j)}{\underline{\sigma}_{ij}^2},$$
$$\frac{\partial \bar{\mu}_{ij}}{\partial v_{ij}} = \frac{(x_j - v_{ij})\boldsymbol{N}(v_{ij}, \bar{\sigma}_{ij}; x_j)}{\bar{\sigma}_{ij}^2},$$
$$\frac{\partial \underline{\mu}_{ij}}{\partial \underline{\sigma}_{ij}} = (x_j - v_{ij})^2 \frac{\boldsymbol{N}(v_{ij}, \underline{\sigma}_{ij}; x_j)}{\underline{\sigma}_{ij}^3}, \quad \frac{\partial \bar{\mu}_{ij}}{\partial \underline{\sigma}_{ij}} = 0,$$
$$\frac{\partial \bar{\mu}_{ij}}{\partial \bar{\sigma}_{ij}} = \frac{(x_j - v_{ij})^2\boldsymbol{N}(v_{ij}, \bar{\sigma}_{ij}; x_j)}{\bar{\sigma}_{ij}^3}, \quad \frac{\partial \underline{\mu}_{ij}}{\partial \bar{\sigma}_{ij}} = 0.$$

θ_{ij}^l is the lth parameter of the jth antecedent of the ith rule. for $j = 1, 2, \ldots n$, $i = 1, 2, \ldots c$ and $l = v, \underline{\sigma}, \bar{\sigma}$.

(ii) Consequent Parameters

$$\frac{\partial e}{\partial \beta_{ip}} = (\hat{y} - y)\left(\frac{\bar{f}_i + \underline{f_i}}{\overline{\boldsymbol{F}}^T\boldsymbol{r} + \underline{\boldsymbol{F}}^T\boldsymbol{r}}\right)\tilde{x}_p \tag{16}$$

β_{ip} is the pth consequent parameter of the ith rule, for $p = 0, 1, 2, \ldots n$ and $i = 1, 2, \ldots c$.

(iii) Manual Adjustment

When validating the elicited model, some data points may fall way off outside the error band (10%). To improve generalization, the response surface close to where there is 'output outlier' is adjusted manually. This technique involves feeding in the input of this 'outlier' and then adjusting the fired rules values in an iterative way until the difference between the predicted and the real output at this 'outlier' point is reduced as much as possible. For example in fig. 7c1, predicted output was way more than the real output for one of the data points (outside the error band). We took that data point and fed it into the trained model. We checked for rules which were fired and adjusted the MFs in each input space so that the particular rules which were fired have less firing strength. Since we were using a Gaussian primary MF, we reduced the spread of the some input space components so that the fired rules have less firing strength. This was done on a trial and error basis. It should be emphasized that the manual adjustment which may slightly reduce the overall performance of the system with respect to the training data may be necessary as we are concerned primarily with eliciting a model with excellent generalization capabilities

IV. RESULTS

The proposed modelling framework is tested with a real world engineering application associated with the mechanical property prediction of hot rolled steels. The performance index considered to evaluate the obtained fuzzy models is the root square-mean error ($RMSE$) defined by (18):

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(\hat{y}_k - y_k)^2}{N}} \tag{18}$$

Where N is the number of testing data, y_k and $\hat{y}_k$ are the kth actual output and corresponding model output respectively.

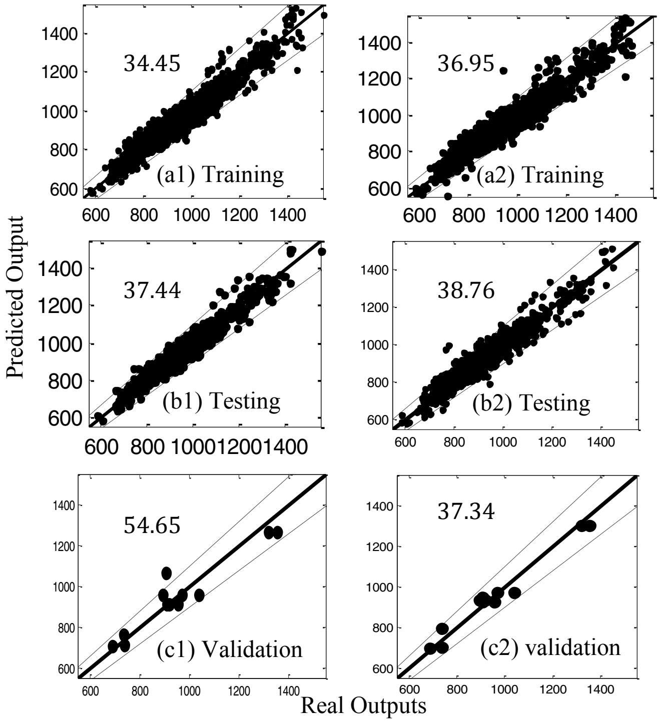

Fig. 7. Performances of the IT2-Squared framework.
(1) Before Manual Adjustment (2) After Manual Adjustment

The heat treatment of hot-rolled steel is a complex high dimensional non-linear process lacking in theoretical analysis which may help to accurately predict a specific mechanical property of steel alloys. However, there is extensive human knowledge. In the past, some data-driven models were developed to assist metallurgists to design alloys (Wang *et al.*, 2012a). One of the goals of these models is to predict the Ultimate Tensile Strength (UTS) which is a common measure of metal strength. A total of 3760 data samples collected from the industry are used in these models and in our research. These data samples include 15 inputs and one output (UTS). 12 data points which are completely removed for the initial data set were used for validation of the final model. The 12 data points were used because, according to the experts, this is the main test of generalisation properties of any elicited model. Six rules (clusters) were found from data using our elicited framework which follows the number of clusters found in the literature. The input and output variables of interest are given in table I. Prediction performance of our elicited model is shown in fig. 7. Our results are also compared with those available in the literature (Panoutsos *et al.*, 2005 and Wang *et al.*, 2012b).

The IT2-Squared framework shows consistency throughout the three data sets which is what is desired of a model with good generalization capability. There was a slight decrease in training and testing RMSE after adjusting the response surface but this is not critical as the performance is still better than previously elicited models as shown in table II. It should be noted that the main strength of the elicited model is the ability to generalise across the training, testing and validation data sets.

Table I Input and Output Variables.

	Input Variables	Output Variable
1	Test Depth (mm)	Ultimate Tensile Strength (UTS) in MegaPascal (MPa)
2	Size (mm)	
3	Site Number	
4	Carbon (%)	
5	Silicon (%)	
6	Manganese (%)	
7	Sulphur (%)	
8	Chromium (%)	
9	Molybdenum (%)	
10	Nickel (%)	
11	Aluminium (%)	
12	Vanadium (%)	
13	Hardening Temperature (^{o}C)	
14	Cooling Medium	
15	Temperature (^{o}C)	

Table II Comparison of our methods with those found in the Literature for the prediction of UTS of steel (RMSE).

Method	Training	Testing	Validation
IT2_Squared	34. 45	38.76	37.34
MOIT2FM	36.33	40.52	34.77
IMOFM_M	46.47	45.12	49.87

V. CONCLUSION

In this paper, a new framework for data-driven IT2 FLS is developed. Stages 1 and 2 of this new framework involve using the IT2 fuzzy clustering for finding the initial parameters of the antecedent MFs and then using least-squares to find the consequent parameters. We also showed how to derive the steepest descent algorithm for an IT2 FLS when the NT defuzzification method is used. To improve generalisation, we manually adjusted the response surface to fit in those data points that were not predicted well on a trial and error basis. Future works will involve automating this last stage of the process.

REFERENCES

Babuska, R and H.B. Verbruggen (1995). New Approach to contructing fuzzy relational models from data. *Proc. 3rd European Congr. on Intelli. Techniq. and Soft Comput*. EUFIT '95. Aachen, Germany, pp. 583-587.

Choi, B. and F. C. Rhee (2009). Interval type-2 fuzzy memberships function generation methods. *Inform. Sci*,. vol. 179, pp. 2102-2122.

Delgado M., A. F. Gomez-Skarmeta and F. Martin (1997). A Fuzzy Clustering-Based Rapid Prototyping of fuzzy rule-based modelling. *IEEE Trans. on Fuzzy Systems*, vol. 5, pp. 223–233.

G. Panoutsos and M. Mahfouf, 2005. Granular computing and evolutionary fuzzy modelling for mechanical properties of alloy steels, *Proc. 16th IFAC World Congress*, Prague, Czech Republic.

Hwang C. And F.C Rhee, 2007. Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means. *IEEE Trans. on Fuzzy Systems*, vol. 15, pp. 107–120.

Liang, Q. and J. M. Mendel, 2000. Interval Type-2 Fuzzy Logic Systems: Theory and Design. *IEEE Trans. on Fuzzy Systems*, vol. 8, pp. 535–550.

Karnik, N. N., J. M. Mendel and Q. Liang, 1999 .Type-2 Fuzzy Logic Systems. *IEEE Trans. on Fuzzy Systems*, vol. 7, pp. 643-658.

L. A. Zadeh, 1975. The concept of a linguistic variable and its application to approximate reasoning–i, *Inform. Sci*., vol. 8, pp. 199–249.

Mendel, J. M., 2001. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice-Hall, Upper-Saddle River, NJ.

Mendel, J. M., 2004. Computing Derivatives in Interval Type-2 Fuzzy Logic Systems. *IEEE Trans. on Fuzzy Systems*, vol. 12, pp. 84-98.

Nie, M and W. W. Tan, 2008. Towards an Efficient Type-Reduction Method for Interval Type-2 Fuzzy Logic Systems. *Proc. IEEE FUZZ Conference*, Paper # FS0339, Hong Kong, China.

Park, S. and H. L.-Kwang, 2001. A designing Method for Type-2 Fuzzy Logic Systems Using Genetic Algorithms. *Proc. of Joint 9th IFSA World Congress and 20th NAFIPS Intl. Conference*, Vancouver, Canada, pp. 2567-2572, 2001.

Wang, S. and M. Mahfouf, 2012a. A New Computationally Efficient Mamdani Interval Type-2 Fuzzy Modelling Framework. *Proc. IEEE FUZZ Conference*, Brisbane Australia.

Wang, S. And M. Mahfouf, 2012b. Multi-Objective Optimisation for Fuzzy Modelling using Interval Type-2 Fuzzy Sets. *Proc. IEEE FUZZ Conference*, Brisbane Australia.

APPENDIX

(a) Derivatives of Antecedent Parameters

$$\hat{y} = \left(\sum_{i=1}^{c} h_i\bar{f}_i + \sum_{i=1}^{c} h_i\underline{f}_i\right) \Big/ \left(\sum_{i=1}^{c}\bar{f}_i + \sum_{i=1}^{c}\underline{f}_i\right) \qquad (A-1)$$

Where

$$h_i = \sum_{i=0}^{n} a_{ij}x_j$$

In matrix form. A-1 may be re-expressed as:

$$\hat{y} = \left(\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{H} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{H}\right)/\left(\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}\right) \qquad (A-2)$$

r is a cx1column vector of 1's

$$\frac{\partial e}{\partial \theta_{ij}^{l}} = \frac{\partial e}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial \theta_{ij}^{l}} = (f_{s2} - y)\frac{\partial \hat{y}}{\partial \theta_{ij}^{l}} \qquad (A-3)$$

$$since\ \frac{\partial e}{\partial \hat{y}} = \hat{y} - y$$

$$\frac{\partial \hat{y}}{\partial \theta_{ij}^{l}} = \sum_{i=1}^{c}\left(\frac{\partial \hat{y}}{\partial \bar{f}_i}\frac{\partial \bar{f}_i}{\partial \theta_{ij}^{l}} + \frac{\partial \hat{y}}{\partial \underline{f}_i}\frac{\partial \underline{f}_i}{\partial \theta_{ij}^{l}}\right) \qquad (A-4)$$

For $j = 1, 2, \dots n,\ i = 1, 2, \dots c$ and $l = v, \underline{\sigma}, \bar{\sigma}$.
It can be shown that

$$\frac{\partial \hat{y}}{\partial \bar{f}_i} = \frac{\partial \hat{y}}{\partial f_i} = \frac{\partial \hat{y}}{\partial f_i} = \frac{hi - f_{s2}}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}} \qquad (A-5)$$

Therefore,

$$\frac{\partial f_{s2}}{\partial \theta_{ij}^{l}} = \sum_{w=1}^{c}\left(\frac{hi - f_{s2}}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}}\left(\frac{\partial \bar{f}_i}{\partial \theta_{wj}^{l}} + \frac{\partial \underline{f}_i}{\partial \theta_{wj}^{l}}\right)\right) \qquad (A-6)$$

$$\frac{\partial \bar{f}_i}{\partial \theta_{wj}^{l}} = \begin{cases} \frac{\partial \bar{f}_i}{\partial \theta_{ij}^{l}} & i = w \\ 0 & i \neq w \end{cases}; \qquad \frac{\partial \underline{f}_i}{\partial \theta_{wj}^{l}} = \begin{cases} \frac{\partial \underline{f}_i}{\partial \theta_{ij}^{l}} & i = w \\ 0 & i \neq w \end{cases} \qquad (A-7)$$

$$\frac{\partial f_{s2}}{\partial \theta_{ij}^{l}} = \frac{hi - f_{s2}}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}}\left(\frac{\partial \bar{f}_i}{\partial \theta_{ij}^{l}} + \frac{\partial \underline{f}_i}{\partial \theta_{ij}^{l}}\right) \qquad (A-8)$$

$$\bar{f}_i = \prod_{q=1}^{n}\bar{\mu}_{iq} \qquad \underline{f}_i = \prod_{q=1}^{n}\underline{\mu}_{iq}$$

$$\frac{\partial \bar{f}_i}{\partial \theta_{ij}^{l}} = \left|\prod_{q=1}^{n}\bar{\mu}_{iq}\right| * \frac{\partial \bar{\mu}_{ij}}{\partial \theta_{ij}^{l}}; \qquad \frac{\partial \underline{f}_i}{\partial \theta_{ij}^{l}} = \left|\prod_{q=1}^{n}\underline{\mu}_{iq}\right| * \frac{\partial \underline{\mu}_{ij}}{\partial \theta_{ij}^{l}} \qquad (A-9)$$

For Gaussian primary MF

$$\mu_{ij}(x_j) = exp\left[-\frac{1}{2}\left(\frac{x_j - v_{ij}}{\sigma_{ij}}\right)^2\right], \quad \sigma_{ij} \in [\underline{\sigma}_{ij}, \bar{\sigma}_{ij}] \qquad (A-10)$$

$$\frac{\partial \underline{\mu}_{ij}}{\partial v_{ij}} = (x_j - v_{ij})\,\boldsymbol{N}(v_{ij}, \underline{\sigma}_{ij}; x_j)/\underline{\sigma}_{ij}^{2}, \qquad (A-11)$$

$$\frac{\partial \bar{\mu}_{ij}}{\partial v_{ij}} = \frac{(x_j - v_{ij})\boldsymbol{N}(v_{ij}, \bar{\sigma}_{ij}; x_j)}{\bar{\sigma}_{ij}^{2}}, \qquad (A-12)$$

$$\frac{\partial \underline{\mu}_{ij}}{\partial \underline{\sigma}_{ij}} = (x_j - v_{ij})^2\,N(v_{ij}, \underline{\sigma}_{ij}; x_j)/\underline{\sigma}_{ij}^{3}, \qquad \frac{\partial \bar{\mu}_{ij}}{\partial \underline{\sigma}_{ij}} = 0,$$

$$\frac{\partial \bar{\mu}_{ij}}{\partial \bar{\sigma}_{ij}} = (x_j - v_{ij})^2 N(v_{ij}, \bar{\sigma}_{ij}; x_j)/\bar{\sigma}_{ij}^{3}, \qquad \frac{\partial \underline{\mu}_{ij}}{\partial \bar{\sigma}_{ij}} = 0,$$

Q.E.D.

(b) Derivatives Consequent Parameters

$$\frac{\partial e}{\partial \beta_{ip}} = \frac{\partial e}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial \beta_{ip}} = (\hat{y} - y)\frac{\partial \hat{y}}{\partial \beta_{ip}} \qquad (A-13)$$

for $p = 0, 1, 2, \dots n$ and $i = 1, 2, \dots c$.

$$\frac{\partial \hat{y}}{\partial \beta_{ip}} = \sum_{w=1}^{c}\frac{\partial \hat{y}}{\partial h_w}\frac{\partial h_w}{\partial \beta_{ip}} = \sum_{w=1}^{c}\frac{\bar{f}_w + \underline{f}_w}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}}\frac{\partial h_w}{\partial \beta_{ip}}$$

$$\frac{\partial h_w}{\partial \beta_{ip}} = \begin{cases} \frac{\partial h_i}{\partial \beta_{ip}} & i = w \\ 0 & i \neq w \end{cases} \qquad (A-14)$$

$$\frac{\partial \hat{y}}{\partial \beta_{ip}} = \frac{\bar{f}_i + \underline{f}_i}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}}\frac{\partial h_i}{\partial \beta_{ip}} \qquad (A-15)$$

$$\frac{\partial h_i}{\partial \beta_{ip}} = \tilde{x}_p \qquad (A-16)$$

$$\frac{\partial e}{\partial \beta_{ip}} = (\hat{y} - y)\left(\frac{\bar{f}_i + \underline{f}_i}{\bar{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r} + \underline{\boldsymbol{F}}^{\boldsymbol{T}}\boldsymbol{r}}\right)\tilde{x}_p \qquad (A-17)$$