Applied Soft Computing 48 (2016) 1–12

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Hybrid-modelling of compact tension energy in high strength pipeline steel using a Gaussian Mixture Model based error compensation

Guangrui Zhang[a], Mahdi Mahfouf[b,*], Musa Abdulkareem[c], Sid-Ahmed Gaffour[b], Yong-Yao Yang[b], Olusayo Obajemu[b], John Yates[e], Sabino Ayvar Soberanis[d], Christophe Pinna[e]

[a] College of Information Science and Engineering, Ocean University of China, 23 Xianggang Road East, Qingdao 266100, China
[b] Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S1 3JD, UK
[c] Department of Engineering, University of Leicester, University Road, Leicester LE1 7RH, UK
[d] Advanced Manufacturing Research Centre with Boeing, The University of Sheffield, Advanced Manufacturing Park, l Wallis Way, Catcliffe, Rotherham S60 5TZ, UK
[e] The University of Sheffield, Department of Mechanical Engineering, Mappin Street, Sheffield S1 3JD, UK

ARTICLE INFO

*Article history:*
Received 28 October 2015
Received in revised form 25 February 2016
Accepted 9 June 2016
Available online 28 June 2016

*Keywords:*
Pipeline
Steel
Gaussian Mixture Model
Fuzzy
Neural networks
Prediction

ABSTRACT

In material science studies, it is often desired to know in advance the fracture toughness of a material, which is related to the released energy during its compact tension (*CT*) test to prevent catastrophic failure. In this paper, two frameworks are proposed for automatic model elicitation from experimental data to predict the fracture energy released during the *CT* test of *X100* pipeline steel. The two models including an adaptive rule-based fuzzy modelling approach and a double-loop based neural network model, relate the load, crack mouth opening displacement (*CMOD*) and crack length to the released energies during this test. The relationship between how fracture is propagated and the fracture energy is further investigated in greater detail. To improve the performances of the models, a Gaussian Mixture Model (*GMM*)-based error compensation strategy which enables one monitor the error distributions of the predicted result is integrated in the model validation stage. This can help isolate the error distribution pattern and to establish the correlations with the predictions from the deterministic models. This is the first time a data-driven approach has been used in this fashion on an application that has conventionally been handled using finite element methods or physical models.



## 1. Introduction

High strength steel is one of the most commonly used materials in engineering works and the modelling, prediction and prevention of failure of steel materials is a key issue in engineering because of safety concerns and to prevent the huge costs incurred during failures. It is thus no surprise that there is a plethora of materials science studies aim at developing new methods of analysis as well as improving existing techniques.

Fracture toughness relates to the ability of a material with intrinsic cracks to resist failure.

Existing analysis on the fracture toughness of steel used in the design of pipeline steel is the calibrated empirical method based on finite element analysis. This method, although returning good modelling results on the test set, have unfortunately been found to have poor generalisation results across steel specimens. As illustrated in [1], using the charpy upper shelf energy which is predicted by the old application ultimately leads to a large error in determining the pipeline fracture resistance.

Physical based-modelling combined with the Finite Element Methods (*FEM*) are popular for ascertaining fracture characteristics in metals. For example, [2] used the Gurson-Tvergaard-Needleman (*GTN*) model for the prediction of the ductile failure of 22NiMoCr37 and SA-333 Gr-6 Carbon steel. Also, Karabin et al. in [3], developed a constitutive model based on the Gurson-Tvergaard (*GT*) and Leblond-Perrin-Devaux (*LPD*) model [4] for 7085-T7X aluminum alloy plate samples.

Unfortunately as found in [5], the very high dimensionality and complexities of the process variables may incur high computational cost when trying to analyse the models from first principles.

As illustrated in [6,7], mathematical models which are based on data-driven approaches may prove a better solution to this problem. These modelling approaches include fuzzy systems, artificial

* Corresponding author. Tel.: +44 0114 222 5607.
*E-mail addresses:* m.mahfouf@sheffield.ac.uk (M. Mahfouf), docjry@gmail.com (J. Yates).

http://dx.doi.org/10.1016/j.asoc.2016.06.007

**Table 1**
Composition of the steel specimens used in the CT experiment.

| | | | | | | |
|---|---|---|---|---|---|---|
| Element | C | Si | Mn | P | S | Cu |
| Wt % | 0.06 | 0.18 | 1.84 | 0.008 | 0.001 | 0.31 |
| Element | Ni | Cr | Mo | Nb | Ti | Al |
| Wt % | 0.5 | 0.03 | 0.25 | 0.05 | 0.018 | 0.036 |

neural networks, Gaussian processes and support vector machines among others. These approaches have proved to be popular in materials engineering because of their interpolating and generalising capabilities.

For example, [8] predicted the impact energy of API X65 micro alloyed steel using the Artificial Neural Networks (*ANN*) The fuzzy modelling approach was used for modelling the hysteretic behaviour of CuAlBe wire from experimental data in [9]. The literature is replete with different types of computational intelligence techniques applied to materials modelling. They have shown to provide good accuracy on the specific experimental data. However, these methods tend to be 'biased' and are not able to provide a high degree of confidence in predictions. In this work, we provide a data-driven approach of modelling and consequently predicting materials failure in high strength X100[1] pipeline steel. The research examines two types of modelling framework on the steel crack propagation process during the compact tension test on the steel prototypes. The first is based on fuzzy modelling with hierarchical clustering for initial structure determination and the gradient descent optimisation to improve on the accuracy of the model. This method follows directly from that developed in [7]. The second framework is based on a double loop neural networks. The accuracies in predictions of both methods are compared. To further improve on the accuracy of the two elicited models, an error compensation scheme based on Gaussian Mixture Models was developed for the two techniques. The careful design of this error compensation scheme is not only shown to improve on the performances of the two modelling paradigm but also provides a confidence band in the predictions of each model systematically. Finally, the modelling performance of the proposed modelling framework is compared with 55 that of the adaptive neuro-fuzzy inference system modelling framework (*AN-FIS*). The remainder of the paper is organised as follows: Section 2 analyses the X100 steel data used in the paper explaining the input variables the composition of the steel prototypes. Section 3 briefly describes the proposed fuzzy modelling approach. Section 4 discusses the Neural Network approach used in 60 the paper before the error compensation scheme is used on both models which is described in Section 5. Section 6 concludes the paper and recommends direction for future research.

## 2. Data and analysis

The experimental data used in this research originated from the works carried out in the Department of Mechanical Engineering, the University of Sheffield [10]. At room temperature, tests were carried-out on six compact tension specimens with longitudinal direction initial crack. This is the direction of shear fracture in cases of real burst pipelines. The steel specimens were side-grooved on each side by up to 20% of the original thickness of the specimen. This ensures a straight crack front and that shear lip formation are reduced. A low displacement control rate of $0.01 mm/s$ was used during the tests. Table 1 shows the composition of the *X100* pipeline steel used in the experiments.

[1] X100 are high grade steel with yield strength greater than 690 MPa and are usually used for high distance engineering projects.

In the experiments the explanatory variables are the load, *CMOD* and crack-length. The output variable is the released flat fracture energy during the tests, which is indicative of the strength of the steel. Six test data sets contain a total of 432 data points which were used in developing the models. Of the 432 data points, 70% was used in the training the two models (fuzzy and neural networks) and the remaining 30% for testing the generalization capabilities of the elicited models. Fig. 1 shows the distributional characteristics of the data.

It is worth noting that the figure shows that the same load value corresponds to two different released energies. This is because the experiment was carried out using a crack speed controlling procedure, meaning that when the elastic property of the metal was broken in the middle of the crack propagation, the load was lowered to maintain the crack speed. Additionally, the figure only shows the released energy as a function of only the load variable. The released energy have been influenced by other input variables.

### 2.1. Correlation coefficient analysis

Table 2 shows the correlation coefficient analysis for the variables (input and output) to identify the effects the inputs have on the outputs.

The corresponding analysis shows that the correlation between the load and the energy is negative. This is due to decreasing load in the middle of fracture which is caused by the crack controlling procedure. The correlation between the crack length and *CMOD* is high which agrees with the intuition of crack length and *CMOD* increasing simultaneously during fracture. Finally, it may also be concluded that *CMOD* and crack length affect energy more than load.

## 3. Fuzzy model of compact tension energy

The use of fuzzy logic modelling in material science is widespread because of its ability to find very accurate linguistic representation of very complex non-linear systems thus enhancing interpretability (transparency) and simplicity of the process [11]. Fig. 2 shows a typical structure of a fuzzy logic system (*FLS*). The fuzzifier component maps a real input in $R^D$ into a fuzzy set. A fuzzy set (*FS*) extends the capabilities of a crisp set by allowing elements have degree of membership in the set. So the **fuzzifier** provides the degree of membership that the real input belongs to a particular fuzzy set. The **Fuzzy inference engine** (*FIS*) is the heart of the *FLS* and it determines how the fuzzified input is combined with the rules contained in the **Rule Base** to produce a fuzzified output. Finally the **Defuzzifier** produces a crisp output.

The rules of a fuzzy system is usually of the form:

$$Rule^m: \text{IF } x_1 \text{ is } A_1^m \text{ AND} \ldots \text{AND } x_n \text{ is } A_n^m \text{ THEN } y^m \text{ is } B^m.$$

where $m$ is the number of rules, $n$ is the number of inputs. $A_1^m, \ldots$ and $A_n^m$ are fuzzy sets in the input space and $B^m$ is a fuzzy set in the output space. In a Takagi-Sugeno-Kang (*TSK*) *FLS*, the $B^m$s are replaced by $g^m(x_1, \ldots, x_n)$ which represents a function of the inputs variables. Usually this function is just a linear function of inputs.

Expert knowledge is required to build a fuzzy model but mechanisms for automatic rule generation from data may be used when only data is available. Several types of adaptive fuzzy modelling may be found in the literature, [12,13]. The approach used in this work in eliciting the first part of the fuzzy model is similar to that of [14,15], whereby hierarchical clustering is used to determine the initial number of clusters (rules) and then the initial structure of the fuzzy logic model. Data clustering has been shown to be an effective initial fuzzy logic model generation. To improve