Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 158 (2007) 2169–2188

www.elsevier.com/locate/fss

Multiple regression with fuzzy data

Andrzej Bargiela[a,*], Witold Pedrycz[b], Tomoharu Nakashima[c]

[a] School of Computer Science, The University of Nottingham, Nottingham NG8 1BB, UK
[b] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alta., Canada T6G 2G6
[c] College of Engineering, Osaka Prefecture University, Gakuen-cho 1-1, Sakai, Osaka 599-8531, Japan

Received 25 July 2005; received in revised form 23 February 2007; accepted 6 April 2007
Available online 20 April 2007

Abstract

In this paper, we propose an iterative algorithm for multiple regression with fuzzy variables. While using the standard least-squares criterion as a performance index, we pose the regression problem as a gradient-descent optimisation. The separation of the evaluation of the gradient and the update of the regression variables makes it possible to avoid undue complication of analytical formulae for multiple regression with fuzzy data. The origins of fuzzy input data are traced back to the fundamental concept of information granulation and an example FCM-based granulation method is proposed and illustrated by some numerical examples. The proposed multiple regression algorithm is applied to one-, three- and nine-dimensional synthetic data sets as well as the 13-dimensional Boston Housing dataset from the machine learning repository. The algorithm's performance is illustrated by the corresponding plots of convergence of regression parameters and the values of the prediction error of the resulting regression model. General comments on the numerical complexity of the proposed algorithm are also provided.

Keywords: Multiple regression; Fuzzy data; Gradient descent; Fuzzy C-means (FCM)

1. Introductory comments

Regression analysis is one of the basic tools of scientific investigation enabling identification of functional relationship between independent and dependent variables. In the classical regression analysis both the independent and dependent variables are given as real numbers. However, in many real-life situations, where the complexity of the physical system dictates adoption of a more general viewpoint, regression variables are given as non-numerical entities such as linguistic variables [6]. Unfortunately, such real-life situations are quite often outside the scope of the classial regression analysis [2,3].

Following the introduction of the concept of fuzzy sets by Zadeh in 1965 [22–24] various researchers attempted extending the regression analysis from crisp to fuzzy domain. The problem statement in this study is diametrically different from the one commonly encountered in the literature. Traditionally, starting from the early work by

[*] Corresponding author. Tel.: +44 7768634741.

E-mail addresses: Andrzej.Bargiela@cs.nott.ac.uk (A. Bargiela), pedrycz@ece.ualberta.ca (W. Pedrycz), nakashi@cs.osakafu-u.ac.jp (T. Nakashima).

0165-0114/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2007.04.011

Tanaka et al. [19], see also [18,20,1,16,21], fuzzy regression was introduced as follows:

- Given numeric experimental data $(\mathbf{x}_k, \mathbf{y}_k)$ $k = 1, 2, \ldots, N$, design a fuzzy regression $Y = A_0 \oplus A^{\mathrm{T}}\mathbf{x}$ where A_0 and A are parameters of the model treated as some fuzzy numbers (in particular described by triangular membership functions). Owing to the character of the model and the form of the assumed parameters of the model, Y becomes also a triangular fuzzy number. Note that the operation of addition (denoted here by $\oplus$) pertains to fuzzy numbers rather than plain numeric entities.

In essence, the fuzziness at the output of the regression model emerged because of the lack of perfect fit of numeric data to the assumed linear format of the relationship under consideration. In other words, through the introduction of triangular numbers (parameters), this fuzzy regression reflects the deviations between the data and the linear model. Computationally, the estimation of the fuzzy parameters of the regression is concerned with some problems of linear programming; refer again to the early works in this area. There have been a number of variants of the underlying optimisation techniques, cf. [11,13,16].

In a nutshell, in spite of the differences in the optimisation, the overall mapping can be captured through the relationship

$$\mathbf{R}^n \to F(\mathbf{R}), \tag{1}$$

where $(F\mathbf{R})$ denotes a family of fuzzy numbers (in our case triangular ones) defined in the space of real numbers $\mathbf{R}$.

The approach advocated in this study marks a departure from the conceptual frameworks governed by (1). For a given collection of input–output fuzzy data, we are concerned in building a numeric regression model that approximates the fuzzy data in an optimal fashion. Referring to (1), the relationship of interest here arises in the form

$$F(\mathbf{R}^n) \to F(\mathbf{R}). \tag{2}$$

To make the problem of building the regression line more manageable from the optimisation standpoint, we can refine (revise) the mapping to read as

$$F(\mathbf{R}) \times F(\mathbf{R}) \times \cdots \times F(\mathbf{R}) \to F(\mathbf{R}) \tag{3}$$

with $F(.)$ denoting the corresponding families of fuzzy numbers.

The conceptual framework (2) was originally adopted by Diamond [7,8] who developed a simple linear regression model for triangular fuzzy numbers. This was subsequently generalised to fuzzy regression models with regression variables expressed as arbitrary fuzzy numbers [10,4,5,13,17]. Another generalisation of the regression model, involving the use of fuzzy random variables, was suggested by Koerner and Nather [15]. However, in all of the above approaches the analytical formulae quantifying the values of the parameters of the regression model had to address the issue of negative spreads [9] which complicates significantly the algorithms and makes them difficult to apply to highly-dimensional data. The consequence of having to consider 2^{n-1} (n is dimensionality of data) of optimisation cones in analytical regression methods meant that most of the examples of use of these methods were confined to low-dimensional data, typically single independent and single dependent variable systems.

The main contribution of this paper is the re-formulation of the regression problem as a gradient-descent optimisation, which enables a natural generalisation of the simple regression model to multiple regression models in a computationally feasible way. Indeed, the new formulation provides a basis for a further generalisation to multiple non-linear regression with fuzzy variables.

In Section 2, we provide some background on the classical regression analysis, fuzzy numbers and fuzzy simple linear regression. In Section 3, we extend the scope of fuzzy regression to multiple variables and provide a gradient-descent optimisation algorithm that provides a practical way of calculating regression coefficients. Section 4 offers some practical considerations of generating fuzzy sets for independent and dependent variables and provides several numerical examples of the application of the algorithm to various data sets.

2. Background discussion

2.1. Classical regression analysis

The general task of regression analysis is defined as identification of a functional relationship between the independent variables $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ and dependent variables $\mathbf{y} = [y_1, y_2, \ldots, y_m]$, where n is a number of independent variables in each observation and m is a number of dependent variables. The regression model is expressed in this case as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \varepsilon, \tag{4}$$

where $\mathbf{f}(\mathbf{x})$ is a vector function $[f_1(\mathbf{x}), \dots, f_m(\mathbf{x})]$ and $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_m]$ is a vector of random error of functional approximation. The general model (1) is frequently simplified by assuming a linear relationship between the independent and dependent variables, thus reducing the task of identification of a functional relationship $\mathbf{f}()$to an identification of parameters of a linear function. Furthermore, the multiple dependent variables $\mathbf{y}$ are considered separately to give m independent regression models. In the simplest case of a single dependent variable the regression models are given as

$$y = a_0 + a_1 x + \varepsilon \tag{5a}$$

or

$$y = a_0 + a_1 x_1 + \cdots + a_m x_m + \varepsilon. \tag{5b}$$

The above are typically referred to as a *simple linear regression* and *multiple linear regression*, respectively.

2.2. Simple linear regression with fuzzy data

In order to generalise the simple linear regression to the case of imprecise (non-numeric) independent and dependent variables we follow the approach proposed by Diamond [7] and adopt a subfamily of fuzzy sets, called fuzzy numbers, as a formal framework for the representation of imprecise data. A fuzzy number can be formally defined as follows [7]:

Definition 1. A fuzzy subset A of the set of real numbers R with membership function $\mu_A : R \to [0, 1]$ is called a fuzzy number if

(i) A is normal, i.e. there exists an element z_0 such that $\mu_A(z_0) = 1$;
(ii) A is fuzzy convex, i.e. $\forall z_1, z_2 \in R \mu_A(\lambda z_1 + (1 - \lambda) z_2) \geqslant \mu_A(z_1) \wedge \mu_A(z_2), \forall \lambda \in [0, 1]$;
(iii) μ_A is upper semicontinuous;
(iv) $\sup(A) = \overline{\{z \in R : \mu_A(z) > 0\}}$ is bounded.

A fuzzy number A can be represented as a family of sets called α-cuts, A_α, defined as

$$A_\alpha = \{z \in R : \mu_A(z) \geqslant \alpha\} \tag{6}$$

and giving rise to a set representation

$$A = \bigcup_{\alpha \in (0,1]} A_\alpha. \tag{7}$$

Based on the resolution identity we get

$$\mu_A(z) = \sup\{\alpha I_{A_\alpha}(z) : \alpha \in (0, 1]\}, \tag{8}$$

where $I_{A_\alpha}(z)$ represents the characteristic function of A_a. From the definition of the fuzzy number it is easily seen that every α-cut of a fuzzy number A is a closed interval $A_\alpha = [A^{\mathrm{L}}(\alpha), A^{\mathrm{U}}(\alpha)]$ where

$$A^{\mathrm{L}}(\alpha) = \inf\{z \in R : \mu_A(z) \geqslant \alpha\}, \tag{9}$$

$$A^{\mathrm{U}}(\alpha) = \sup\{z \in R : \mu_A(z) \geqslant \alpha\}. \tag{10}$$

Consequently, for two fuzzy numbers A and B with α-cuts $A_\alpha = [A^{\mathrm{L}}(\alpha), A^{\mathrm{U}}(\alpha)]$ and $B_\alpha = [B^{\mathrm{L}}(\alpha), B^{\mathrm{U}}(\alpha)]$ we can define a distance between A and B as

$$d(A, B) = \sqrt{\int_0^1 (A^{\mathrm{L}}(\alpha) - B^{\mathrm{L}}(\alpha))^2 \, \mathrm{d}\alpha + \int_0^1 (A^{\mathrm{U}}(\alpha) - B^{\mathrm{U}}(\alpha))^2 \, \mathrm{d}\alpha}. \tag{11}$$

Using the formalism of fuzzy numbers we can express the fuzzy simple linear regression problem as a problem of identification of parameters $b_0, b_1 \in R$ of a fuzzy linear model

$$Y = b_0 + b_1 X. \tag{12}$$

The parameters b_0, b_1 are evaluated by minimising the error measured as a distance between the actual observations and the estimates evaluated from the model

$$\min H(b_0, b_1) = \sum_{i=1}^{k} d^2(Y_i, b_0 + b_1 X_i). \tag{13}$$

It must be noted, however, that the exact form of the error function $H(.)$ depends on the sign of the parameter b_1. If $b_1 > 0$ then

$$H^+(b_0, b_1) = \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{L}}(\alpha) - b_0 - b_1 X_i^{\mathrm{L}}(\alpha))^2 \,\mathrm{d}\alpha + \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{U}}(\alpha) - b_0 - b_1 X_i^{\mathrm{U}}(\alpha))^2 \,\mathrm{d}\alpha \tag{14}$$

and if $b_1 < 0$ then

$$H^-(b_0, b_1) = \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{L}}(\alpha) - b_0 - b_1 X_i^{\mathrm{U}}(\alpha))^2 \mathrm{d}\alpha + \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{U}}(\alpha) - b_0 - b_1 X_i^{\mathrm{L}}(\alpha))^2 \,\mathrm{d}\alpha. \tag{15}$$

Having pre-determined the sign of the parameter b_1 we can calculate exact numerical values of b_0 and b_1 by solving either

$$\frac{\partial H^+(b_{0,}, b_1)}{\partial b_0} = 0 \quad \text{and} \quad \frac{\partial H^+(b_{0,}b_1)}{\partial b_1} = 0 \tag{16}$$

or

$$\frac{\partial H^-(b_{0,}b_1)}{\partial b_0} = 0 \quad \text{and} \quad \frac{\partial H^-(b_{0,}b_1)}{\partial b_1} = 0. \tag{17}$$

In the first case we obtain

$$\hat{b}_0^+ = \tilde{Y} - \hat{b}_1^+ \tilde{X}, \tag{18}$$

$$\hat{b}_1^+ = \frac{SS_{xy}^+}{SS_{xx}}, \tag{19}$$

where

$$\tilde{X} = \int_0^1 \frac{\bar{X}^{\mathrm{L}}(\alpha) + \bar{X}^{\mathrm{U}}(\alpha)}{2} \,\mathrm{d}\alpha, \tag{20}$$

$$\tilde{Y} = \int_0^1 \frac{\bar{Y}^{\mathrm{L}}(\alpha) + \bar{Y}^{\mathrm{U}}(\alpha)}{2} \,\mathrm{d}\alpha, \tag{21}$$

$$SS_{xx} = \sum_{i=1}^{k} \int_0^1 ((X_i^{\mathrm{L}}(\alpha))^2 + (X_i^{\mathrm{U}}(\alpha))^2) \,\mathrm{d}\alpha - 2k\tilde{X}^2, \tag{22}$$

$$SS_{xy}^+ = \sum_{i=1}^{k} \int_0^1 (X_i^{\mathrm{L}}(\alpha) Y_i^{\mathrm{L}}(\alpha) + X_i^{\mathrm{U}}(\alpha) Y_i^{\mathrm{U}}(\alpha)) \,\mathrm{d}\alpha - 2k\tilde{X}\tilde{Y}. \tag{23}$$

In the second case the regression parameters are estimated as

$$\hat{b}_0^- = \tilde{Y} - \hat{b}_1^- \tilde{X}, \tag{24}$$

$$\hat{b}_1^- = \frac{SS_{xy}^-}{SS_{xx}}, \tag{25}$$

where

$$SS_{xy}^- = \sum_{i=1}^{k} (X_i^{\mathrm{U}}(\alpha) Y_i^{\mathrm{L}}(\alpha) + X_i^{\mathrm{L}}(\alpha) Y_i^{\mathrm{U}}(\alpha))\, \mathrm{d}\alpha - 2k\tilde{X}\tilde{Y} \tag{26}$$

and $\tilde{X}$, $\tilde{Y}$, SS_{xx} are evaluated as in (20), (21), (22), respectively.

2.3. Gradient-descent optimisation

Although with the quantification of the regression error, (14) and (15), it is possible to find analytical solution to Eqs. (16) and (17), in the case of a large number of regression variables such an approach is problematic because the number of cost functions that have to be considered grows exponentially as 2^m (where m is a number of regression variables). In other words, the analytical solution to fuzzy linear regression described by (16)–(26) is NP-complete in the number of regression variables. To overcome the above limitation we propose an alternative approach to fuzzy regression that is based on iterative refinement of regression model parameters. We observe that with an initial approximation to b_0 and b_1 it is possible to evaluate partial derivatives of (14) or (15) as indicators of the local gradient of the functional H. The computational advantage of this approach is twofold: first the calculation of the values of partial derivatives of H is much simpler than solving systems of simultaneous equations such as (16)–(17) and second the approximations of the regression variables define uniquely the form of the functional H thus reducing the exponential set of problems to one-problem-per-iteration. Indeed, the proposed approach can be easily applied even if the regression model is non-linear. For the simple regression model given by (12) the gradient-descent optimisation can be summarised as:

Gradient-descent algorithm for simple fuzzy regression.

(a) Make an initial guess of b_0 and b_1 say b_0^0 and b_1^0;

(b) Set the iteration counter $i = 1$;

(c) Evaluate gradient H with respect to regression model parameters as per Eq. (14) or (15);

(d) Update the parameters

$$\Delta b_0 = \mu_0 \frac{\partial H^+(b_0, b_1)}{\partial b_0} \quad \text{and} \quad \Delta b_1 = \mu_1 \frac{\partial H^+(b_0, b_1)}{\partial b_1} \tag{27}$$

or

$$\Delta b_0 = \mu_0 \frac{\partial H^-(b_0, b_1)}{\partial b_0} \quad \text{and} \quad \Delta b_1 = \mu_1 \frac{\partial H^-(b_0, b_1)}{\partial b_1}; \tag{28}$$

(e) Update parameter estimates

$$b_0^i = b_0^{i-1} + \Delta b_0 \quad \text{and} \quad b_1^i = b_1^{i-1} + \Delta b_1; \tag{29}$$

(f) If $\Delta b_0 > \varepsilon$ or $\Delta b_1 > \varepsilon$ then update iteration counter $i = i + 1$ and repeat from (c); otherwise stop.

In order to converge to the solution the algorithm requires appropriate selection of parameters μ_0 and μ_1 and the selection of the termination criteria ε. However, even with a heuristic selection of these parameters (which may require repeated runs of the algorithm) the evaluation of the regression model parameters remains computationally very appealing.

3. Multiple linear regression with fuzzy data

The fuzzy simple linear regression model (12) can now be extended to a fuzzy model with multiple independent variables:

$$Y = b_0 + b_1 X^1 + b_2 X^2 + \cdots + b_m X^m, \tag{30}$$

where $Y, X^1, X^2, \ldots, X^m$ are all fuzzy numbers defined on R and $b_0, b_1, b_2, \ldots, b_m$ are real numbers. The parameters $b_0, b_1, b_2, \ldots, b_m$ are evaluated by minimising the cost function $H(.)$ defined as a squared distance between the fuzzy observations and the corresponding fuzzy dependent variable Y evaluated from (30):

$$\min H(b_0, b_1, \ldots, b_m) = \sum_{i=1}^{k} d^2(Y_i, b_0 + b_1 X_i^1 + b_2 X_i^2 + \cdots + b_m X_i^m) \tag{31}$$

with a distance function $d(.)$ defined as in (11).

Using the α-cut representation of fuzzy numbers it is necessary to ensure that the minimum and maximum values of α-cut intervals are properly matched for both positive and negative values of model parameters $b_0, b_1, b_2, \ldots, b_m$. We can formalise this requirement by introducing the following substitution of variables:

$$\widehat{X}_i^{j\mathrm{L}}(\alpha) = X_i^{j\mathrm{L}} \quad \text{and} \quad \widehat{X}_i^{j\mathrm{U}}(\alpha) = X_i^{j\mathrm{U}} \quad \text{if } b_j \geqslant 0, \tag{32}$$

$$\widehat{X}_i^{j\mathrm{L}}(\alpha) = X_i^{j\mathrm{U}} \quad \text{and} \quad \widehat{X}_i^{j\mathrm{U}}(\alpha) = X_i^{j\mathrm{L}} \quad \text{if } b_j < 0, \tag{33}$$

where L and U denote the corresponding lower and upper bounds of the α-cut intervals and $j = 1, \ldots, m$. With these substitutions the cost function $H(.)$ can be written explicitly as

$$\begin{aligned} \widehat{H}(b_0, b_1, \ldots, b_m) = {} & \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{L}}(\alpha) - b_0 - b_1 \widehat{X}_i^{1\mathrm{L}}(\alpha) - \cdots - b_m \widehat{X}_i^{m\mathrm{L}}(\alpha))^2 \, \mathrm{d}\alpha \\ & + \sum_{i=1}^{k} \int_0^1 (Y_i^{\mathrm{U}}(\alpha) - b_0 - b_1 \widehat{X}_i^{1\mathrm{U}}(\alpha) - \cdots - b_m \widehat{X}_i^{m\mathrm{U}}(\alpha))^2 \, \mathrm{d}\alpha. \end{aligned} \tag{34}$$

Using expression (34) we can calculate gradients of the cost function $\widehat{H}(.)$ with respect to the regression parameters as

$$\frac{\partial \widehat{H}(b_0, \ldots, b_m)}{\partial b_0} = -2\tilde{Y} + 4kb_0 + 2b_1 \tilde{\widehat{X}}^1 + \cdots + 2b_m \tilde{\widehat{X}}^m \tag{35}$$

and

$$\begin{aligned} & \frac{\partial \widehat{H}(b_0, \ldots, b_m)}{\partial b_j} \\ & \quad = -2 \overset{\leftrightarrow}{SS}_{X^j Y} + 2b_0 \tilde{\widehat{X}}^j + 2b_1 \overset{\leftrightarrow}{SS}_{X^j X^1} + \cdots + 2b_m \overset{\leftrightarrow}{SS}_{X^j X^m}, \end{aligned} \tag{36}$$

where

$$\tilde{Y} = \int_0^1 \left(\sum_{i=1}^{k} Y_i^{\mathrm{L}}(\alpha) + \sum_{i=1}^{k} Y_i^{\mathrm{U}}(\alpha) \right) \mathrm{d}\alpha, \tag{37}$$

$$\tilde{\widehat{X}}^j = \int_0^1 \left(\sum_{i=1}^{k} \widehat{X}_i^{j\mathrm{L}}(\alpha) + \sum_{i=1}^{k} \widehat{X}_i^{j\mathrm{U}}(\alpha) \right) \mathrm{d}a, \quad j = 1, \ldots, m, \tag{38}$$

Table 1
Computational complexity of regression algorithms

Number of variables fuzziness of data	Crisp data (analytical solution)	Fuzzy data (analytical solution)	Fuzzy data (iterative solution)
1 independent and 1 dependent var. (simple regression)	(A) $\mathrm{O}(p(k))$	(C) $\mathrm{O}(2 * P(k))$ (16)–(26)	(E) $\mathrm{O}(\mathrm{iter} * Q(k))$ (27)–(29)
n independent and 1 dependent var. (multiple regression)	(B) $\mathrm{O}(m * p(k))$	(D) $\mathrm{O}(2^m * P(k))$ (NP problem: exponential number of Eqs. (16)–(17))	(F) $\mathrm{O}(\mathrm{iter} * m * Q(k))$ (41)–(43)

$$\overset{\leftrightarrow}{SS}_{X^jY} = \int_0^1 \sum_{i=1}^{k} (Y_i^{\mathrm{L}}(\alpha)\, \overset{\frown}{X}{}_i^{j\mathrm{L}}(\alpha) + Y_i^{\mathrm{U}}(\alpha)\, \overset{\frown}{X}{}_i^{j\mathrm{U}}(\alpha))\,\mathrm{d}\alpha, \quad j = 1, \ldots, m, \tag{39}$$

$$\overset{\leftrightarrow}{SS}_{X^jX^p} = \int_0^1 \sum_{i=1}^{k} (\overset{\frown}{X}{}_i^{p\mathrm{L}}(\alpha)\, \overset{\frown}{X}{}_i^{j\mathrm{L}}(\alpha) + \overset{\frown}{X}{}_i^{p\mathrm{U}}(\alpha)\, \overset{\frown}{X}{}_i^{j\mathrm{U}}(\alpha))\,\mathrm{d}\alpha, \quad j, p = 1, \ldots, m. \tag{40}$$

With the above equations (30)-(38) we can now define the gradient-descent algorithm for multiple fuzzy regression.

Gradient-descent algorithm for multiple fuzzy regression.

(a) Make an initial guess of $b_0, b_1, \ldots, b_m$: $b_0^0, b_1^0, \ldots, b_m^0$;

(b) Set the iteration counter $i = 1$;

(c) Evaluate the α-cut intervals for individual regression variables taking into account the sign of the corresponding regression variable; (32)–(33);

(d) Evaluate gradient of $\overset{\frown}{H}(.)$ with respect to regression model parameters as per Eq. (35) or (36);

(e) Calculate the value of the regression parameters update

$$\Delta b_0 = \mu_0 \frac{\partial \overset{\frown}{H}(b_0, \ldots, b_m)}{\partial b_0} \tag{41}$$

and

$$\Delta b_j = \mu_j \frac{\partial \overset{\frown}{H}(b_0, \ldots, b_m)}{\partial b_j}, \quad j = 1, \ldots, m, \tag{42}$$

where μ_j are the parameters controlling the convergence of the gradient-descent optimisation;

(f) Update parameter estimates

$$b_j^i = b_j^{i-1} - \Delta b_j, \quad j = 0, \ldots, m; \tag{43}$$

(g) If $\exists_{j=0,1,\ldots,m} \Delta b_j > \varepsilon$ then update iteration counter $i = i + 1$ and repeat from (c); otherwise stop.

Table 1 provides a context for the assessment of the computational complexity of the proposed algorithm.

Computational complexity of simple and multiple regression with crisp data (cases A and B) is of order $p(k)$ and $m * p(k)$, respectively, where k is the number of observations and m is the number of regression variables. Given that $p(k)$ is a small multiple of k, it is clear that analytical solutions to large multiple regression problems with crisp data are feasible and indeed are widely used. Computational complexity of analytical solution to simple linear regression with fuzzy data (case C) is of order $2 * P(k)$ where $P(k)$ is a multiple of k that is larger than $p(k)$ but is still a small number, implying that the problem is computationally feasible. Unfortunately, the same cannot be said about case D. The exponential increase of computational complexity $2^m * P(k)$ means that the analytical solution to fuzzy regression can only be found for small numbers of regression variables m.

The proposed iterative approach to fuzzy regression offers an advantage of a smaller computational complexity of a single iteration $Q(k) < P(k)$ but it necessitates iterative refinement of regression variables. For a simple regression (cases C and E) the amount of computations implied by iter $* Q(k)$ may well be more than $2 * P(k)$. However, when the number of regression variables increases (cases D and F) the amount of computations iter $* m * Q(k)$ is much less than that implied by $2^m * P(k)$.

4. Practical considerations

While the paper focuses firmly on the discussion of multiple regression with fuzzy data, it is important to appreciate the origins of fuzzy data. Such data arise in real life as an abstraction (granulation) of a less fuzzy or indeed crisp data. The resulting information granules have different semantics from the semantics of individual data items because they place a greater emphasis on the generalisation of the character of data rather than the precise numerical value of it. In this sense the objectives of the information granulation are closely aligned with the objectives of the regression analysis where the regression model is intended to capture the general character of the data and not to memorise the numerical values of dependent variables. Consequently, it is reasonable to expect that the regression model built on fuzzy information granules, as opposed to the model built on the original crisp data, will benefit from making use of input data that are semantically closer to the regression model.

In order to provide a firm reference point for the methodology of information granulation, without distracting the reader from the main focus of this paper, we outline the fuzzy C-means (FCM)-based granulation approach in the section below. There are several legitimate reasons behind the choice of FCM. The algorithm is well known within the community so that this helps comprehend the essence of the data granulation supported in this manner. The computational facet of the FCM is well documented so that the reader can easily quantify the effect of various design decisions in the process of information granulation.

It must be emphasised, however, that we do not make any statements about the relative advantages of FCM-based granulation compared to some other granulations. This topic has been investigated by the authors in some detail and has been reported elsewhere [4]. What we do attempt, however, is to provide a detailed evaluation of the performance of the proposed regression analysis against the alternative methods using several synthetic and real-life data sets. These are documented in Section 4.2.

4.1. Fuzzy regression data

Starting with the FCM-based data granulation, we consider a set of N data patterns $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ where each pattern is an element in the $n + 1$-dimensional space $\mathbf{R}^{n+1}$ (comprising of n independent variables and 1 dependent variable). The objective is to cluster data into "c" clusters by minimising the following objective function:

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 d_{ik}, \tag{44}$$

where $U = [u_{ik}, u_{ik} \geqslant 0, i = 1, 2, \ldots, c, k = 1, 2, \ldots, N$, is a partition matrix describing clusters in data. The distance function between the kth pattern and ith prototype is denoted as $d_{ik} = \text{dist}(\mathbf{x}_k, \mathbf{v}_i)$ where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_c$ are the prototypes characterising the clusters.

The FCM prototypes provide a good basis for the generation of triangular fuzzy sets representing the independent and dependent variables in the regression problem. Let us consider data depicted in Fig. 1. The four FCM prototypes that are calculated for this data are projected onto the two axes and result in partitioning of the domain of the two variables. By constructing triangular fuzzy sets around the projections of FCM prototypes, as illustrated in Fig. 2, we ensure that the sum of membership grades for variables that fall between projections of prototypes is always summing up to 1. As a consequence, the generated triangular fuzzy sets provide a parsimonious coverage of the pattern space.

An alternative approach to the generation of fuzzy sets describing independent and dependent variables is to consider physical constraints on the accuracy of individual data points. If these represent measurements obtained using instrumentation of a given accuracy it is well justified to represent readings as fuzzy sets representing both the value of the measurement and the accuracy of the meter. Clearly, the exact form of the fuzzy set needs to relate to the characteristics of the measuring devices but it is quite intuitive to adopt triangular fuzzy sets to convey the meaning that the values

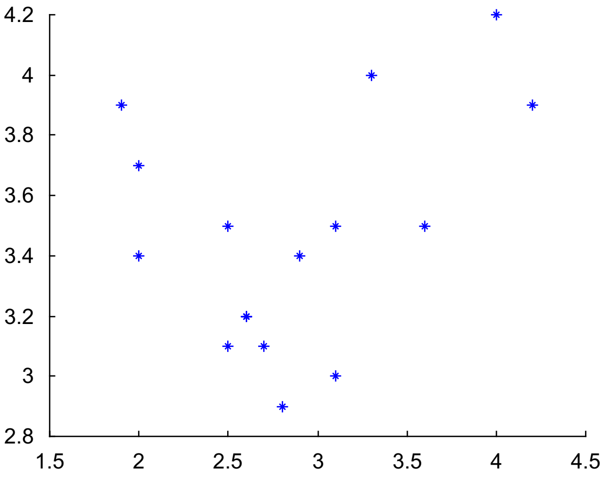

Fig. 1. Original numerical data.

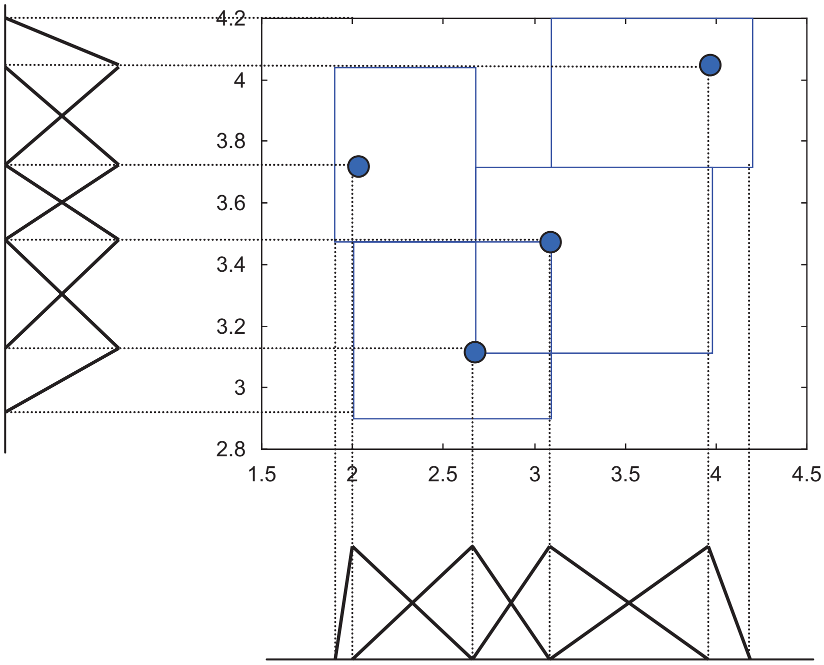

Fig. 2. Construction of triangular fuzzy sets around projections of FCM prototypes.

around the actual reading are more representative than the values at the extreme ends of the accuracy range. It must be stressed, however, that other forms of fuzzy sets can be considered with only minor implications on subsequent processing.

4.2. Numerical examples

Example 1. The simple linear regression of fuzzy data generated through FCM clustering is illustrated in Fig. 3. The effect of the fuzzy set representation of data can be appreciated by comparing the regression line obtained with fuzzy independent and dependent variables (depicted in red) to the regression line calculated for crisp prototypes (depicted in blue), Fig. 3.

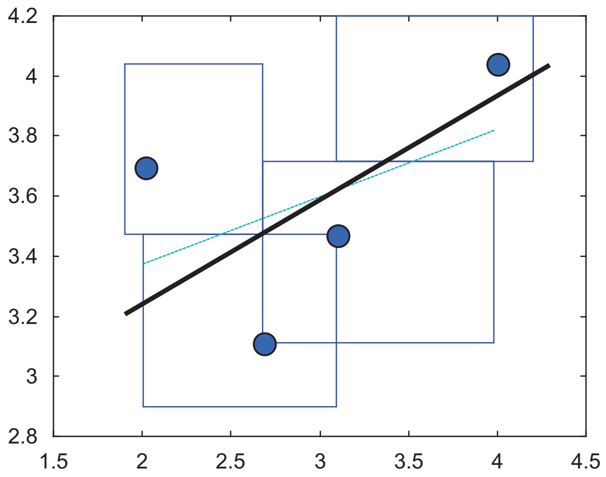

Fig. 3. Regression lines evaluated for FCM prototypes (dashed line) $b = [2.9251\ 0.2244]$ and fuzzy variables (solid line) $b = [2.5508\ 0.3479]$.

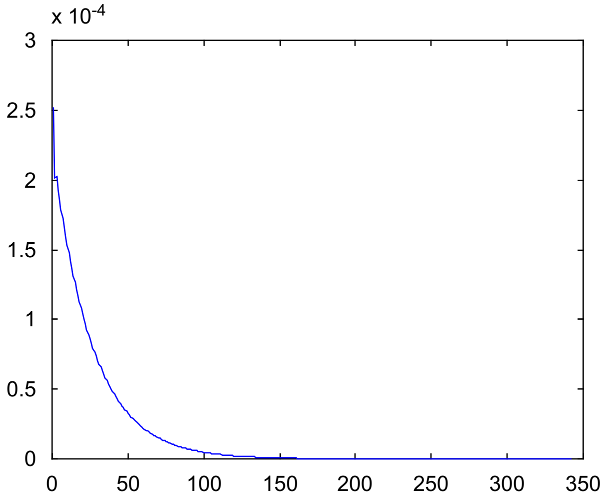

Fig. 4. Cost function; convergence of the iterative gradient-descent algorithm.

The improvement of the quality of the regression line is quantified here as a reduction of the value of the cost function H specified in (34). The value of this function over the whole iterative process is illustrated in Fig. 4.

The evolution of the regression parameters and the value of their updates in the course of the iterative gradient-descent optimisation is given in Fig. 5. It is clear that the iterative optimisation process controlled with the selected value of the learning rate, $\mu = 0.1$, and the asymptotical convergence to the optimal values of the regression parameters is quite rapid. Further improvement of the convergence rate is possible by employing some adaptive learning strategy but for all practical purposes the fixed learning rate is quite satisfactory.

Example 2. The crisp measurement values, illustrated in Fig. 6a, are augmented with the information about the measurement accuracy of independent and dependent variables giving the corresponding triangular fuzzy sets. The regression lines evaluated for the crisp (dashed line) and fuzzy (solid line) data show the effect of the additional information on the regression model. Looking at the values of the cost function H, illustrated in Fig. 7, it is clear that the adjustment of the regression line results in a significant reduction of the distance between the predicted and actual dependent variables.

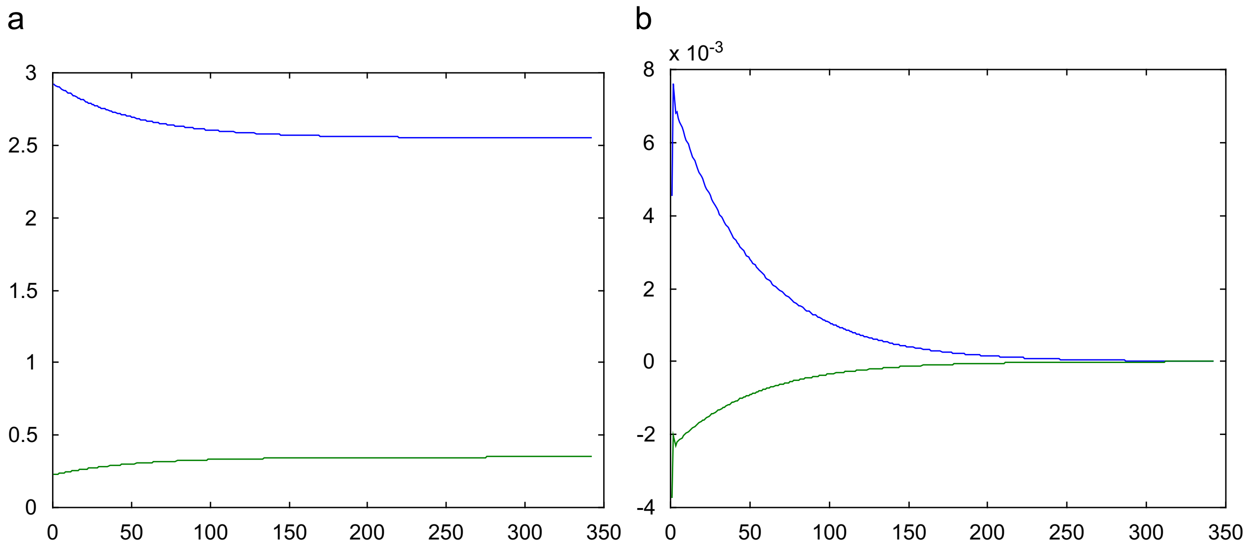

Fig. 5. (a) Values of regression parameters and (b) adjustments to regression parameters in consecutive iterations.

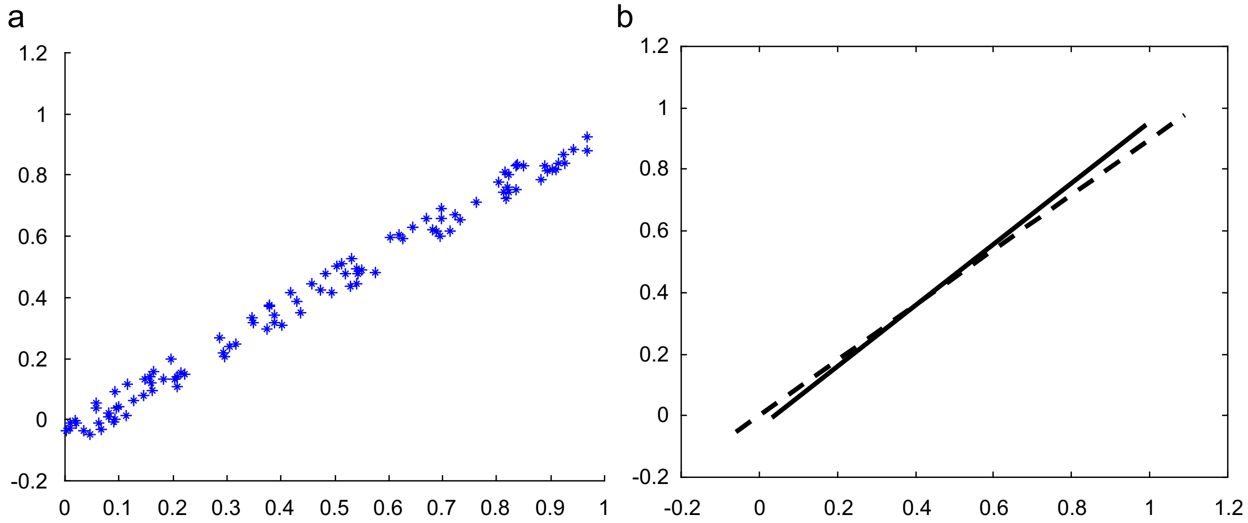

Fig. 6. (a) Crisp data and (b) regression lines evaluated for crisp data (dashed line) and using fuzzy independent and dependent variables (solid line).

Because the cost function is a quadratic form with respect to the regression variables the convergence to the final solution is, as expected, quite rapid and dependent only on the magnitude of the necessary adjustment from the initial guess of the parameters, Fig. 8.

Example 3. In order to explore the convergence properties of the iterative gradient-descent algorithm on multi-dimensional data we generated a synthetic data set with 3 independent and 1 dependent variable. The data was represented by triangular fuzzy sets reflecting the varying measurement accuracies of individual variables.

The convergence of the algorithm is illustrated by the evolution of the cost function in the iterative process, shown in Fig. 9, and the corresponding values of the regression parameters and their updates are shown in Fig. 10. It is clear that in the case of three-dimensional data the adjustments that were necessary to bring the regression parameters from their

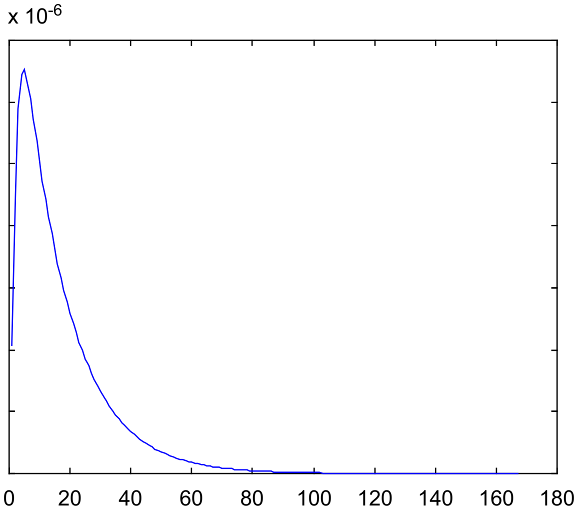

Fig. 7. Cost function; convergence of the iterative gradient-descent algorithm.

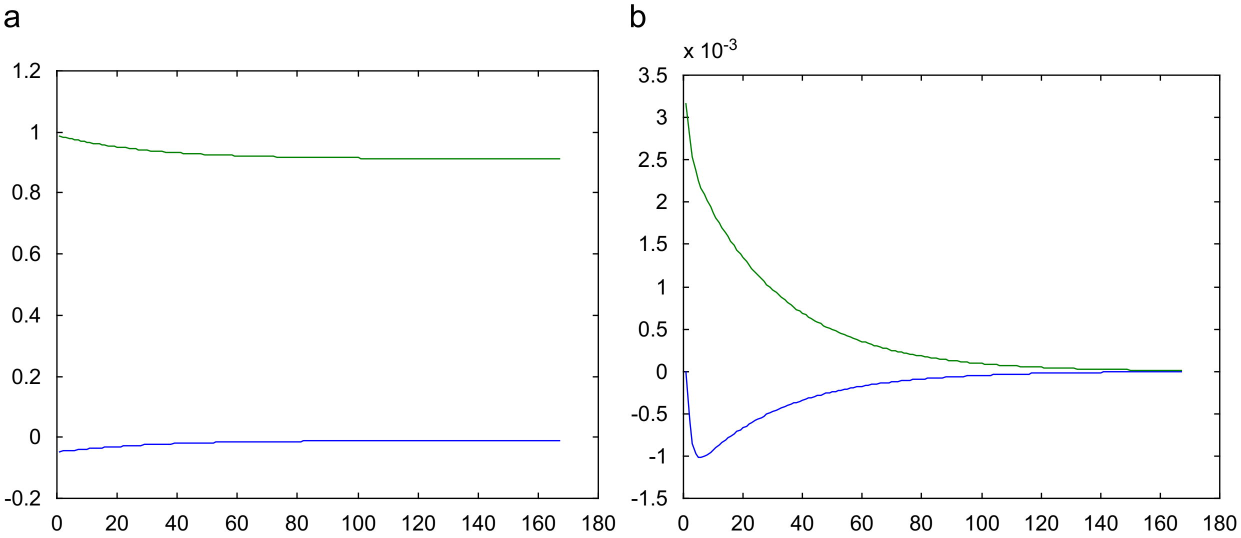

Fig. 8. (a) Values of regression parameters in consecutive iterations and (b) adjustments to regression parameters.

initial guess to their optimal value were quite extensive thus resulting in a larger number of iterations. Nevertheless, the convergence has the same asymptotic character as in the previous two examples.

Example 4. Further increase of the dimensionality of the regression problem has been achieved by introducing for every independent variable from Example 3 additional two copies of these thus resulting in a 10-dimensional pattern space with 9 independent and 1 dependent variable. In this case the regression hyperplane is nine-dimensional so that rather than displaying the projections of the hyperplane we confine ourselves to documenting the convergence characteristics of the algorithm on this data.

Figs. 11 and 12 demonstrate that the increase of the dimension of the regression problem does not affect the convergence properties of the proposed algorithm. Indeed the number of iterative steps needed in the case of nine-dimensional data is only 30% greater than the number of iterative steps needed in the case of three-dimensional data.

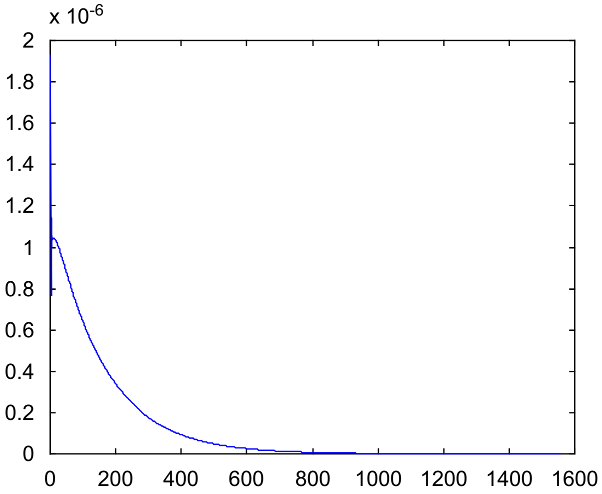

Fig. 9. Cost function; convergence of the iterative gradient-descent algorithm.

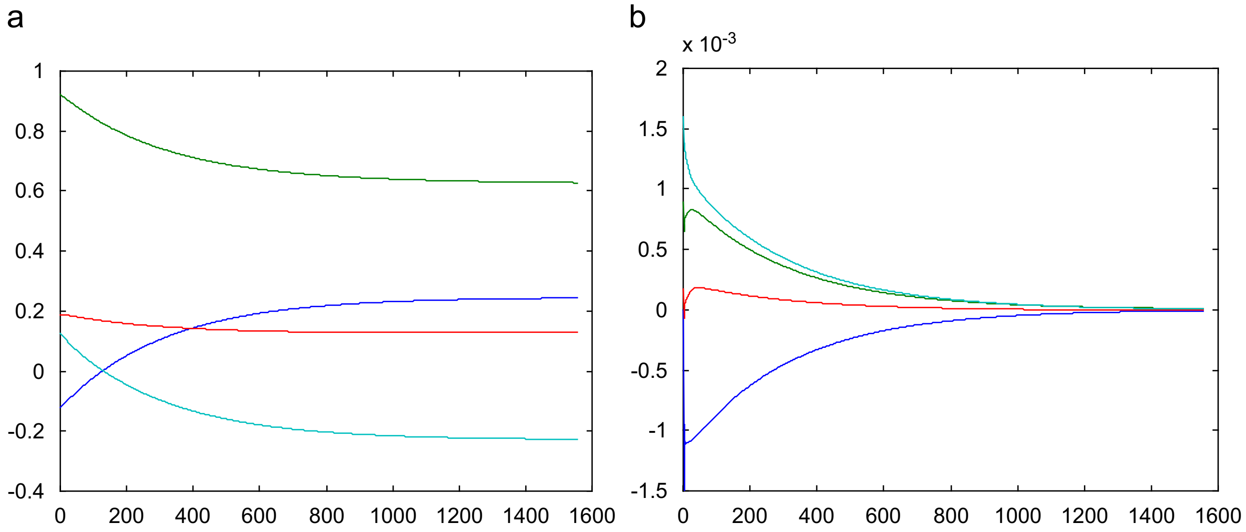

Fig. 10. Values of regression parameters in consecutive iterations and (b) adjustments to regression parameters.

The stopping criterion has been, as before, the attainment of absolute values of corrections to individual regression parameters to be less than 0.00001.

Example 5. We compare here prediction errors associated with three regression models: the model obtained with the proposed iterative gradient-descent algorithm, the model derived by Tanaka et al. [19] and the model derived by Kao and Chyu [14]. The evaluation is performed using two synthetic data sets that were used in [19,14]. Bearing in mind that Tanaka's model was derived using crisp independent and fuzzy dependent variables, while both our model and the one proposed in [14] are derived using fuzzy variables, we represent the independent variables in the first data set as degenerate fuzzy numbers as given in Table 2.

The second data set, given in Table 3, consists of eight pairs of triangular fuzzy numbers and was used to evaluate the performance of the regression model proposed in [13].

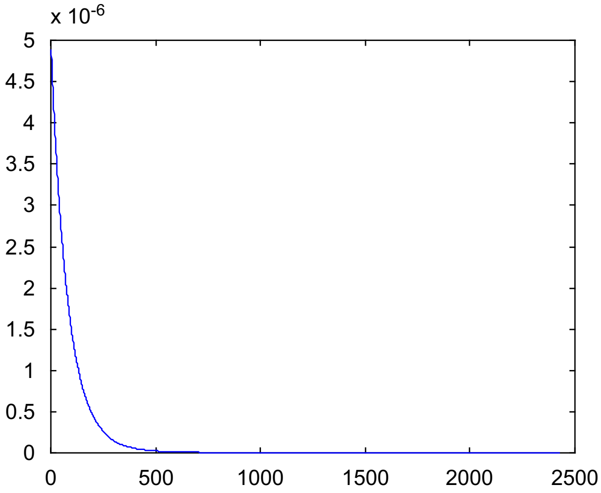

Fig. 11. Cost function; convergence of the iterative gradient-descent algorithm.

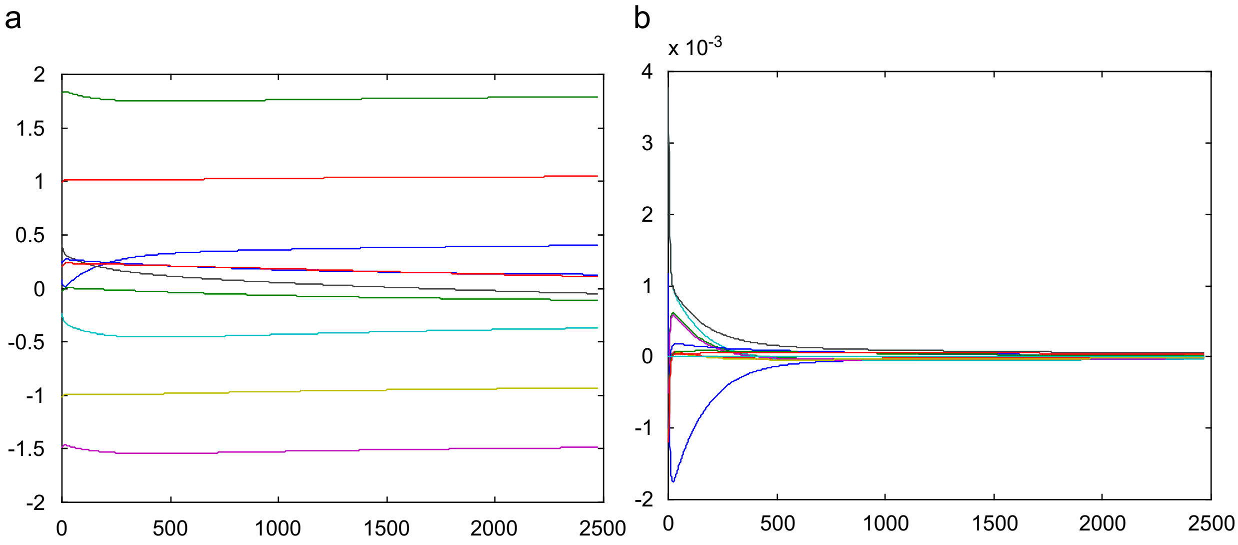

Fig. 12. (a) Values of regression parameters in consecutive iterations and (b) adjustments to regression parameters.

Table 2
Synthetic dataset used in [19]

Independent variable	Dependent variable
(1.0, 1.0, 1.0)	(6.2, 8.0, 9.8)
(2.0, 2.0, 2.0)	(4.2, 6.4, 8.6)
(3.0, 3.0, 3.0)	(6.9, 9.5, 12.1)
(4.0, 4.0, 4.0)	(10.9, 13.5, 16.1)
(5.0, 5.0, 5.0)	(10.6, 13.0, 15.4)

The parameters (b_0, b_1) of the linear regression models obtained for the data set given in Table 2 are: (4.450, 1.833), (4.808, 1.718), (4.950, 1.719) for Tanaka's, Kao's and our model, respectively. For the data set given in Table 3 the corresponding regression models are (3.201, 0.579), (3.565, 0.522), (3.4467, 0.5360). These are illustrated in Fig. 13.

Table 3
Synthetic data set used in [14]

Independent variable	Dependent variable
(1.5, 2.0, 2.5)	(3.5, 4.0, 4.5)
(3.0, 3.5, 4.0)	(5.0, 5.5, 6.0)
(4.5, 5.5, 6.5)	(6.5, 7.5, 8.5)
(6.5, 7.0, 7.5)	(6.0, 6.5, 7.0)
(8.0, 8.5, 9.0)	(8.0, 8.5, 9.0)
(9.5, 10.5, 11.5)	(7.0, 8.0, 9.0)
(10.5, 11.0, 11.5)	(10.0, 10.5, 11.0)
(12.0, 12.5, 13.0)	(9.0, 9.5, 10.0)

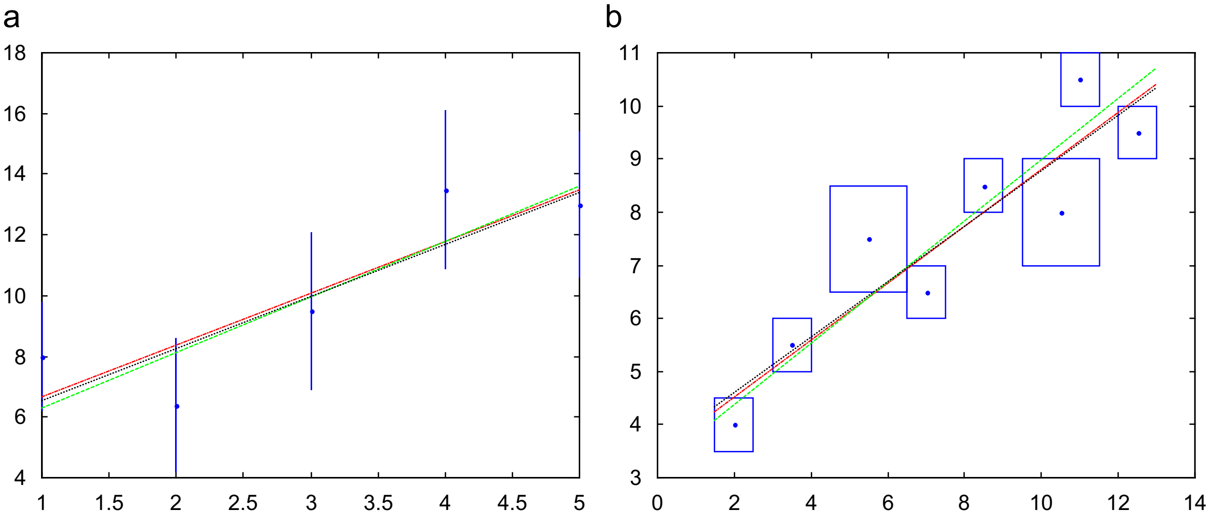

Fig. 13. Regression lines for the data set from Tables 2(a) and 3(b). Tanaka's model (highest gradient line), Kao's model (lowest gradient line) and the proposed model (intermediate gradient line).

Table 4
Root-mean-squared prediction error for the three regression models

	Regression model [19]	Regression model [14]	Proposed model
RMSE Table 2—data set	1.3580	1.3539	1.3522
RMSE Table 3—dataset	0.4981	0.4864	0.4862

The prediction accuracy of the regression models is assessed by evaluating the root-mean-squared-error (RMSE) on the discrepancy between the actual and the predicted values of the dependent variables as given in

$$\mathrm{RMSE} = \sqrt{\frac{1}{c}\sum_{i=1}^{c}(|Y_i - Y_i^*|)^2}. \tag{45}$$

For the triangular fuzzy sets, Y_i represents an average of the three values calculated for the minimum, maximum and the median value of the corresponding sets, and, for the numerical prototypes, Y_i is a single numerical value calculated from the regression model. The value of the RMSE evaluated for the three models is given in Table 4. It is clear that the proposed model offers improved performance over the other two models.

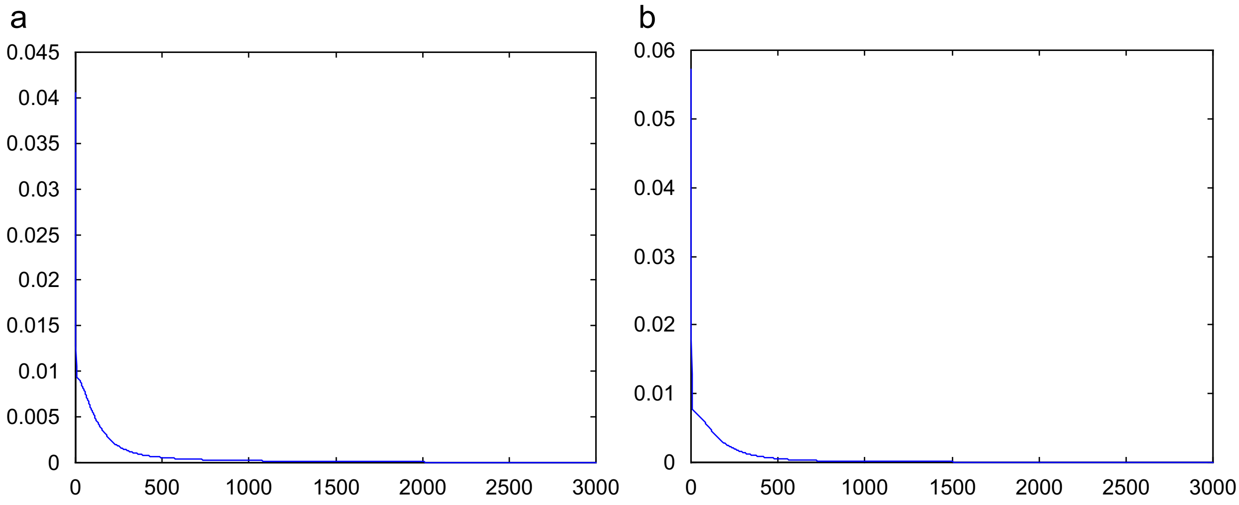

Fig. 14. A typical convergence of the cost function evaluated on the first subset of data (a) and the second subset of data (b).

Example 6. To assess the algorithm on real-life data we make use of the Boston Housing data set from the machine learning repository, http://www.ics.uci.edu/∼mlearn/MLRepository.html.

The data comprises 506 records with 13 continuous and 1 binary-valued attributes. For the purpose of our evaluation, the discrete variable (fourth attribute in the original data) was eliminated. The 13 continuous attributes represent the following: per-capita crime rate by town, proportion of residential land, proportion of non-retail business area, nitric oxides concentration, average number of rooms per dwelling, proportion of owner-occupied units, weighted distance to Boston five main employers, full-value property tax-rate, pupil–teacher ratio, proportion of blacks by town, proportion of lower status of population and median value of owner-occupied homes.

The assessment focuses on two issues: the computational performance of the algorithm and the quantification of the quality of the regression model.

Bearing in mind the vastly different numerical values of the individual attributes it is important that the data are normalised to a common range for all attributes so that the cost function is represented, as nearly as possible, by a hyper-sphere in the 13-dimensional space. Should the data be left unnormalised the cost function would be represented by some hyper-ellipsoid with some semi-axis several orders of magnitude larger than others. This would mean that in any iterative process it would be very difficult, if not impossible, to find suitable learning rates for the updates of the regression parameters.

Since the values of attributes come from finite ranges and the data are representative we are justified in performing a linear normalisation of data according to the formula

$$d_{\mathrm{norm}} = \frac{d - d_{\min}}{d_{\max} - d_{\min}}, \tag{46}$$

where d, $d_{\min}$ and $d_{\max}$ represent the specific data value and the minimum and maximum ranges for a given attribute. Of course, such a normalisation implies that should there be additional data appended to the current set that have some attributes having values outside their $[d_{\min} d_{\max}]$ range, the normalisation for the new data set would return some negative or greater than 1 values. However, this is unlikely to be a problem since the overall shape of the cost function in the 13-dimensional pattern space is not going to depart much from the hyper-sphere.

Bearing in mind the subsequent assessment of the quality of the regression model we subdivide the normalised data set into two equal subsets of 253 records each. One of the subsets is used for identifying the regression model and the other for the evaluation of the prediction ability of this regression model. In order to make sure that the conclusions that are drawn from such an assessment are not unduly affected by the way the two subsets are created, the roles of the two subsets are switched round and the assessment is repeated. Consequently, the assessment of computational performance is also run for each of the two subsets as illustrated in Fig. 14. It is clear that apart from some detailed

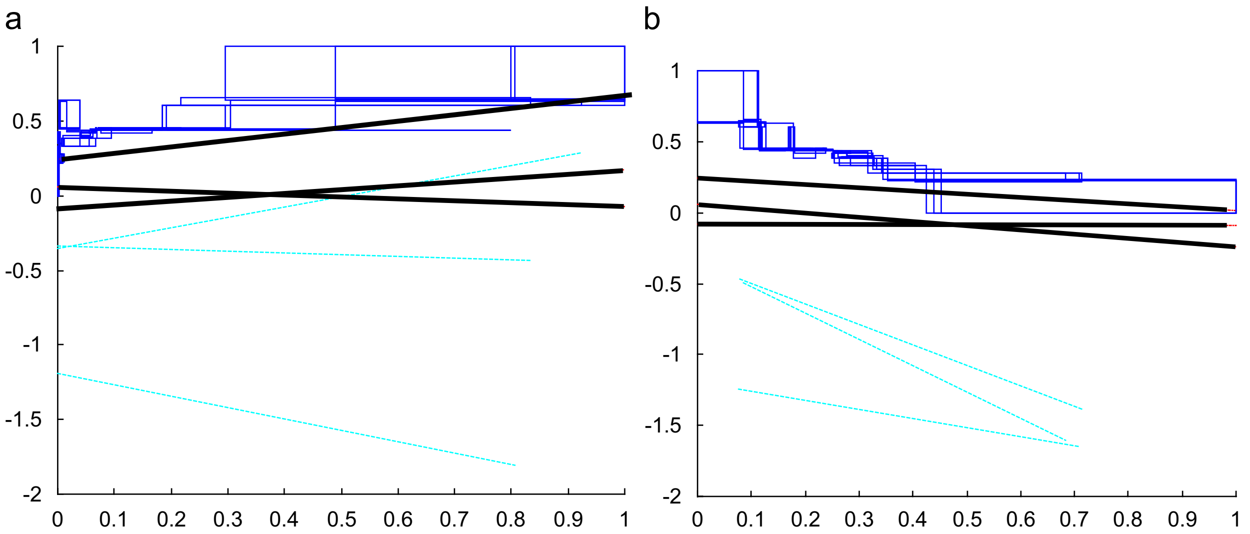

Fig. 15. Projections of the regression hyperplanes onto the planes “dependent variable-attribute 2” (a) and the “dependent variable-attribute 12” (b). Heavy solid lines represent hyperplanes evaluated for triangular fuzzy sets. Light solid lines represent hyperplanes evaluated for numerical values.

numerical values, the general convergence of the algorithm on the two data subsets is the same in nature; within the first 100 iterations the value of the cost function is reduced by 1 order of magnitude and subsequently a similar reduction is achieved approximately every 500 iterations thus resulting in the algorithm convergence in some 3000 iterations. This number of iterative steps compares favourably to the number of variants of error function ($2^{13} = 8192$) that would need to be considered if a direct analytical approach was adopted.

Since we have no additional information about the accuracy of values of individual attributes in the data set we adopt the FCM-based approach as discussed above in Example 1. With the intended assessment of the number of FCM prototypes on the quality of the regression model we select 15, 20, 25 and 30 prototypes reference numbers in the evaluation process and repeat the assessment of the computational performance of the algorithm for these values. Using these prototypes we build the corresponding triangular fuzzy sets that are utilised by our algorithm in evaluating the corresponding regression models. In order to appreciate the advantage of using the triangular fuzzy sets we also calculate the standard regression models based on numerical prototypes and compare the two sets of models.

Since it is difficult to visualise a 12-dimensional hyperplane we provide in Fig. 15a representative sample of projections of the regression hyperplane on the planes defined by the independent variable and the attributes 2 and 12, respectively. The hyperplanes evaluated for 20, 25 and 30 triangular fuzzy sets are depicted as heavy solid lines and the corresponding hyperplanes evaluated for 20, 25 and 30 numerical prototypes are depicted as light solid lines.

One immediate observation is that the regression models based on triangular fuzzy sets are more consistent with each other for the varying numbers of FCM prototypes. This is because the construction of the triangular fuzzy sets ensures that the membership function for each attribute remains constant throughout the range of each attribute regardless of the number of FCM prototypes. By contrast, the change of the distribution of numerical prototypes in the pattern space can have quite profound effect on the regression model that is based on such numerical values. Fig. 15(a) shows that the numerical regression models change from positive to negative correlation between attribute 2 and the dependent variable when there is a change of the number of FCM prototypes. A similar, but somewhat less pronounced, observation can be made with respect to attribute 12.

The prediction accuracy of the regression models is assessed by evaluating the RMSE on the discrepancy between the actual and the predicted values of the dependent variables as given in (45).

The first step of the evaluation is the assessment of the accuracy of the regression models on the training data set. Table 5 itemises the accuracies obtained by the regression models based on triangular fuzzy sets (RMSE-f-base) and the one based on numerical values of FCM prototypes (RMSE-n-base). It is clear that the numerical-based regression model offers a very good fit to the training data and, in this sense, it outperforms the regression model based on triangular fuzzy sets.

Table 5
Regression model accuracy evaluated on the training data

Prototypes	RMSE-f-base	RMSE-n-base
15	0.0703	0.0002
20	0.0245	0.0071
25	0.0315	0.0047
30	0.0239	0.0348

Table 6
Regression model accuracy evaluated on the test data

Prototypes	RMSE-f	RMSE-n
15	0.5091	16.8175
20	0.0993	24.3315
25	0.0578	5.7300
30	0.0837	3.4781

Table 7
Regression model accuracy evaluated on the training data

Prototypes	RMSE-f-base	RMSE-n-base
15	0.0340	0.0424
20	0.0356	0.0512
25	0.0367	0.0767
30	0.0399	0.1008

Table 8
Regression model accuracy evaluated on the test data

Prototypes	RMSE-f	RMSE-n
15	0.0611	0.9312
20	0.0490	0.2566
25	0.0452	0.2997
30	0.0440	0.1905

However, the results look very different when the regression models are applied to the test data set. The generalisation ability of the regression model based on triangular fuzzy sets demonstrates itself through much smaller values of the prediction error (RMSE-f) compared to the corresponding prediction error obtained for the regression model based on numerical FCM prototypes, Table 6.

In order to ensure that the conclusions stated above are not just a reflection of the way in which the training and test sets were created, the roles of the two sets were switched round and the assessment was repeated. Table 7 lists the regression model accuracy evaluated on the new training data set. It is interesting to note that in this case the regression models based on numerical prototypes are not as accurate as before; meaning that the new training data set has a wider mix of patterns resulting in a regression model that balances the influence of individual patterns and, as a consequence, accumulates larger RMSE. The regression model based on triangular fuzzy sets has a broadly constant level of RMSE, which is comparable to RMSE-f-base listed in Table 5.

Having established the baseline in Table 7, we can proceed to the evaluation of the performance of the regression models on test data. Table 8 clearly shows that the generalisation ability of the regression model built on the triangular fuzzy sets translates onto small values of RMSE. The corresponding RMSE evaluated for the regression model built on numerical prototypes RMSE-n is an order of magnitude larger than the corresponding RMSE-f.

5. Conclusions

Regression model based on fuzzy data shows a very beneficial characteristic of enhanced generalisation of data patterns compared to the regression models that are based on numerical data only. This is because the membership function associated with fuzzy sets has a significant informative value in terms of capturing either a notion of accuracy of information or a notion of proximity of patterns in the data set used for the derivation of the regression model. This is well substantiated by empirical tests using both synthetic and real-life data sets.

The new formulation of the regression with fuzzy data as a gradient-descent optimisation problem enabled a natural generalisation of the simple regression model to multiple regression in a way that is computationally feasible. The prior knowledge of the sign (and value) of individual regression parameters, in each iteration, means that it is possible to consider just one cost function at a time for the calculation of gradients. This contrasts with the analytical approach that requires consideration of all permutations of positive/negative values of regression coefficients to come up with a solution. This means that although analytical solution of multiple regression is theoretically feasible it is not practical for real-life systems with more than few regression variables. On the other hand, we have demonstrated that the proposed gradient-descent-based regression copes very well with multi-dimensional data. The optimisation approach provides also a basis for further generalisation to multiple non-linear regression with fuzzy variables.

The proposed method shows several advantages that make it better suited to various real-life situations. Firstly, the method is easily adapted to processing non-triangular fuzzy variables, something that is difficult if not impossible with many of the existing methods. Secondly, the crisp regression coefficients obtained with this method ensure that, unlike with the fuzzy regression coefficients, the spread of the dependent variable increases only as a result of the fuzziness of independent variables. Thirdly, and perhaps most importantly, from a practical standpoint, when all fuzzy observations degenerate to crisp numerical values the proposed method becomes equivalent to conventional least-squares estimation of regression coefficients and generates crisp numerical values of dependent variables while the existing fuzzy regression methods still generate fuzzy predictions.

The performance of the proposed method assessed on a number of synthetic data sets and a multi-dimensional real-life data set suggests that the proposed method has:

- robust convergence for multi-dimensional data,
- improved accuracy of model predictions,
- improved generalisation ability as assessed by the prediction error on test data.

The above characteristics make the method well suited for practical applications.

Acknowledgement

The support from the Engineering and Physical Sciences Research Council (EPSRC, UK), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Osaka Prefecture University, Japan, is gratefully acknowledged.

References

[1] J. de Andrés Sánchez, A.T. Gómez, Estimating a fuzzy term structure of interest rates using fuzzy regression techniques, European J. Oper. Res. 154 (3) (2004) 804–818.

[2] A. Bardossy, Note on fuzzy regression, Fuzzy Sets and Systems 37 (1990) 65–75.

[3] A. Bardossy, I. Bogardi, L. Duckstein, Fuzzy regression in hydrology, Water Resources Res. 26 (1990) 1497–1508.

[4] A. Bargiela, W. Pedrycz, Granular Computing, Kluwer Academic Publishers, Dordrecht, 2002.

[5] A. Bargiela, W. Pedrycz, Recursive information granulation, IEEE Trans. Systems Man Cybernet. 33 (1) (2003) 96–112.

[6] C.B. Cheng, E.S. Lee, Fuzzy regression with radial basis function network, Fuzzy Sets and Systems 119 (2) (2001) 291–301.

[7] P. Diamond, Fuzzy least squares, Inform. Sci. 46 (1988) 141–157.

[8] P. Diamond, Least squares and maximum likelihood regression for fuzzy linear models, in: J. Kacprzyk, M. Fedrizzi (Eds.), Fuzzy Regression Analysis, 1992, pp. 137–151.

[9] P. Diamond, R. Koerner, Extended fuzzy linear models and least squares estimates, Comput. Math. Appl. 33 (9) (1997) 15–32.

[10] P. Grzegorzewski, E. Mrowka, Regression analysis with fuzzy data, in: P. Grzegorzewski, M. Krawczak, S. Zadrozny (Eds.), Soft Computing: Tools, Techniques and Applications, 2004.

[11] M. Hojati, C.R. Bector, K. Smimou, A simple method for computation of fuzzy linear regression, European J. Oper. Res. 166 (1) (2005) 172–184.

[12] D.H. Hong, C. Hwang, Extended fuzzy regression models using regularization method, Inform. Sci. 164 (1–4) (2004) 31–46.

[13] J. Kacprzyk, M. Fedrizzi, Fuzzy regression analysis, Omnitech Press, Physica-Verlag, Warsaw Heildelberg, 1992.

[14] C. Kao, C.L. Chyu, Least-squares estimates in fuzzy regression analysis, European J. Oper. Res. 148 (2) (2003) 426–435.

[15] R. Koerner, W. Nather, Linear regression with random fuzzy variables, Inform. Sci. 109 (1998) 95–118.

[16] M.N. Nasrabadi, E. Nasrabadi, A mathematical-programming approach to fuzzy linear regression analysis, Appl. Math. Comput. 155 (3) (2004) 873–881.

[17] W. Pedrycz, F. Gomide, An Introduction to Fuzzy Sets: Analysis and Design, MIT Press, Cambridge, MA, 1998.

[18] D.A. Savic, W. Pedrycz, Evaluation of fuzzy linear regression models, Fuzzy Sets and Systems 39 (1991) 51–63.

[19] H. Tanaka, S. Uegima, K. Asai, Linear regression analysis with fuzzy model, IEEE Trans. Systems Man Cybernet. 12 (1982) 903–907.

[20] Y. Xue, I.S. Kim, J.S. Son, C.E. Park, H.H. Kim, B.S. Sung, I.J. Kim, H.J. Kim, B.Y. Kang, Fuzzy regression method for prediction and control the bead width in the robotic arc-welding process, J. Mat. Process. Technol. 164–165 (2005) 1134–1139.

[21] K.K. Yen, S. Ghoshray, G. Roig, A linear regression model using triangular fuzzy number coefficients, Fuzzy Sets and Systems 106 (1999) 167–177.

[22] L.A. Zadeh, Fuzzy sets, Inform. and Control 8 (1965) 338–353.

[23] L.A. Zadeh, Fuzzy sets and information granularity, in: M.M. Gupta, R.K. Ragade, R.R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, North Holland, Amsterdam, 1979, pp. 3–18.

[24] L.A. Zadeh, Fuzzy logic = computing with words, IEEE Trans. Fuzzy Systems 4 (2) (1996) 103–111.