



Optimization-Based Fuzzy Regression in Full Compliance with the Extension Principle

B. Stanojević, M. Stanojević

Bogdana Stanojević*

Mathematical Institute of the Serbian Academy of Sciences and Arts
Kneza Mihaila 36, 11000 Belgrade, Serbia

*Corresponding author: bgdnpop@mi.sanu.ac.rs

Milan Stanojević

Faculty of Organizational Sciences, University of Belgrade
Jove Ilića 154, 11000 Belgrade, Serbia
milan.stanojevic@fon.bg.ac.rs

Abstract

Business Analytics – which unites Descriptive, Predictive and Prescriptive Analytics – represents an important component in the framework of Big Data. It aims to transform data into information, enabling improvements in making decisions. Within Big Data, optimization is mostly related to the prescriptive analysis, but in this paper, we present one of its applications to a predictive analysis based on regression in fuzzy environment.

The tools offered by a regression analysis can be used either to identify the correlation of a dependency between the observed inputs and outputs; or to provide a convenient approximation to the output data set, thus enabling its simplified manipulation. In this paper we introduce a new approach to predict the outputs of a fuzzy in – fuzzy out system through a fuzzy regression analysis developed in full accordance to the extension principle. Within our approach, a couple of mathematical optimization problems are solve for each desired α -level. The optimization models derive the left and right endpoints of the α -cut of the predicted fuzzy output, as minimum and maximum of all crisp values that can be obtained as predicted outputs to at least one regression problem with observed crisp data in the α -cut ranges of the corresponding fuzzy observed data. Relevant examples from the literature are recalled and used to illustrate the theoretical findings.

Keywords: fuzzy regression, extension principle, optimization

1 Introduction

Nowadays, the importance of handling Big Data is continuously increasing. Managing Big Data by providing better data insights is a focal point to majority of the research fields. A fuzzy representation of information is one way to handle the uncertainty.

The fuzzy set theory introduced in [20] is widely applied in many research fields, from industry to management and education. Fuzzy logic was developed to generalize the classic logic, and found

a wide applicability in Decision Support (see [18] for more details on its methods, applications and future trends) and Data Processing (see for instance [1] that presents particular fuzzy methods used in Social Sciences Researches). A general view on how fuzzy logic evolved from the classic logic, and where it meets the quantum logic can be found in [10]. The concept of linguistic variable was proposed in [21] emphasizing its effective applications to approximate reasoning. The extension principle was introduced to formalize the arithmetic on fuzzy quantities.

The fuzzy linear regression model was firstly proposed by Tanaka et al. [15], and extended to possibilistic linear regression in [16]. Since then many variants of fuzzy regression models were proposed in the literature. Kahraman et al. [8] surveyed the relevant fuzzy regression approaches, and reported many of their practical applications as: forecasting models for predicting sales of computers and peripheral equipment; the models revealing the relationship of cumulative trauma disorders risk factors, predicting the injuries, and evaluating the risk levels of individuals; the models forecasting the production value of mechanical industry. A systematic review and a huge amount of bibliographic references to fuzzy regression analysis can be found in [5]. The key open questions in the field and important research directions were also indicated in [5].

In the recent literature, Chachi et al. [3] discussed the fuzzy regression based on M-estimates, providing robust estimators of the parameters to avoid undesired effects; Bas [2] proposed a robust fuzzy regression functions approach whose forecasting performance is not affected by the presence of outliers; Wang et al. [17] introduced a fuzzy regression model that uses approximate Bayesian computation instead of usual optimization techniques; Hose and Hanss [7] presented a fuzzy regression approach that took into consideration the worst case variation of the parameters aiming to encode the whole relevant observed information in their membership functions. A quadratic least squared regression analysis was carried out in [12]. The results showed that procedures which deviate from a full compliance with the extension principle can derive misleading solutions.

Kao and Chyu [9] discussed the fuzzy linear regression based on least squares estimates and extension principle. They adapted the extension principle to the problem they solved, and proposed an approximated solution to surmount the complexity of the formulation. Our approach is based on the same idea, but we succeeded to derive the exact values of the endpoints of any α -cut of the regression's coefficients and predicted outputs. Kao and Chyu [9] used constraint programming to optimize the index for ranking fuzzy numbers. That index was formulated using several α -cuts of the membership function of the regression error estimation, and a triangular fuzzy number for each coefficient of the regression was derived. Our approach will use optimization with polynomial objective functions and constraints to derive the supports of the regression's fuzzy number coefficients.

Hojati et al. [6] proposed a more simple method for computing the coefficients of the fuzzy regression, but provided solutions that were far away from those based on the extension principle. They compared their results with those reported in [15] and [11], emphasizing that their approach provides reasonably narrow fuzzy bands. However, a narrow fuzzy band, although desirable, might be misleading, when the observations do not properly fit inside.

Wu [19] proved that the α -cuts of fuzzy-valued functions can be computed with the help of the α -cuts of its coefficients and arguments. Their approach to fuzzy regression followed the extension principle until they introduced a simplified computational procedure by replacing the variables ranging within the α -cut intervals of the observed data with their corresponding endpoints.

Chen and Nien [4] simplified the approaches based on the extension principle by reducing the amount of information extracted from the fuzzy observations and included in their regression model. As usual, using less accurate models simplifies the solving procedure but increases the possibility to obtain misleading results.

The rest of the paper is organized as follows. Section 2 provides details on the needed notation and terminology related to the crisp regression and fuzzy concepts. In Section 3, we formally describe the problem we aim to solve. We describe our novel approach in Section 4, providing detailed description of the optimization models involved in a fuzzy regression analysis with a polynomial predictor function. Section 5 is devoted to experiments: we report our numerical results and their comparative illustrations with relevant results from the literature. Final conclusion and directions for further researches are provided in Section 6.

2 Preliminaries

In this section, we provide the specific notation and terminology related to crisp regression analysis, and fuzzy numbers that we will need in the sequel.

The tools offered by a regression analysis can be used either to identify the strength of a dependency between the observed inputs and outputs; or to provide a convenient approximation to the data set, thus enabling its simplified manipulation. To describe the input-output relationship one uses a predictor function whose unknown parameters must be determined from the observed data. The wide variety of predictor functions defines a wide variety of regression models whose classification is generally related to the norms involved in the criterion used to derive the best values of the parameters. For instance, the least square regression model uses the l_2 norm, the RIDGE regression uses a penalized l_2 norm, the LASSO regression uses a penalized l_1 norm, all for evaluating the distance between the observed and predicted outputs.

For a crisp linear regression analysis with multiple explanatory variables, there are given n observed input k -tuples $(x_{i1}, x_{i2}, \dots, x_{ik})_{i=1, \dots, n}$ and n output scalars $(y_i)_{i=1, \dots, n}$. The coefficients $A = (a_0, a_1, \dots, a_k)^T$ of the predictor function $f_A(x_1, x_2, \dots, x_k)$ are computed such that $f_A(x_{i1}, x_{i2}, \dots, x_{ik})$ provides a good approximation to y_i , for each $i = 1, \dots, n$.

The parametric expression of the linear predictor function is then

$$f_A(X) = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k. \tag{1}$$

In the least square linear regression the coefficients a_0, a_1, \dots, a_k are determined such that the sum of the squared Euclidean distances between y_i and $f_A(x_{i1}, x_{i2}, \dots, x_{ik})$ is minimal, i.e. solving the following optimization problem

$$\min_A \sum_{i=1}^k (y_i - f_A(x_i))^2. \tag{2}$$

Let us denote

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}. \tag{3}$$

The well known formula for computing the coefficients A by minimizing the least squared distances between the observed and predicted outputs is then $A = (X^T X)^{-1} X^T Y$. We will use this formula written in its equivalent form

$$(X^T X) A = X^T Y. \tag{4}$$

In the next section we adapt the crisp linear regression to a regression in fuzzy environment, namely we model the dependencies between inputs and outputs expressed by fuzzy numbers. The fuzzy sets theory introduced by Zadeh [20] provides an efficient tool for modeling the uncertainty. The main concepts of the fuzzy sets theory that are of interest to our study will be briefly presented below.

A fuzzy set \tilde{A} over the universe X is defined as a collection of pairs $(x, \mu_{\tilde{A}}(x))$, where $x \in X$, and $\mu_{\tilde{A}}(x) \in [0, 1]$. The function $\mu_{\tilde{A}} : X \rightarrow [0, 1]$ is the membership function of \tilde{A} , and $\mu_{\tilde{A}}(x)$ is called the membership degree of the crisp value x in the fuzzy set \tilde{A} . Let us denote by $\text{Supp}(\tilde{A})$ the support of the fuzzy set \tilde{A} defined as the set of values with non-zero membership degree.

Let $[\tilde{A}]_\alpha$ denote the α -cut of the fuzzy set \tilde{A} , that is the set of those values x whose membership degrees are greater or equal to α , $\alpha > 0$, i.e.

$$[\tilde{A}]_\alpha = \{x \in X | \mu_{\tilde{A}}(x) \geq \alpha\}. \tag{5}$$

A fuzzy set \tilde{A} of the universe R of real numbers is called fuzzy number (FN) if and only if: (i) it is a fuzzy normal and fuzzy convex set; (ii) its membership function $\mu_{\tilde{A}}$ is upper semi-continuous; and (iii) its support is bounded.

Through the paper, we use triangular fuzzy numbers \tilde{A} that are expressed by triples of real numbers (a^L, a^C, a^U) , $a^L \leq a^C \leq a^U$. The interval (a^L, a^U) is the support of \tilde{A} ; and the interval

$$[\tilde{A}]_\alpha = \left[(1 - \alpha) a^L + \alpha a^C, \alpha a^C + (1 - \alpha) a^U \right], \tag{6}$$

is its corresponding α -cut.

The extension principle, detailed in [21] in the context of linguistic variables, is widely used to complete the fuzzy arithmetic on fuzzy numbers. We will use it in a wider context, namely to define the predicted fuzzy outputs with respect to the crisp observed data that belongs to the α -cut intervals of the original fuzzy observed data. The formal definition of the extension principle given in (7), i.e.

$$\mu_{\tilde{B}}(y) = \begin{cases} \sup_{(x_1, \dots, x_r) \in f^{-1}(y)} \left(\min \{ \mu_{\tilde{A}_1}(x_1), \dots, \mu_{\tilde{A}_r}(x_r) \} \right), & f^{-1}(y) \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

provides the membership degree of y in the fuzzy set \tilde{B} of the universe Y , where \tilde{B} is the result of evaluating the function f at the fuzzy sets $\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_r$ over their corresponding universes X_1, X_2, \dots, X_r . In other words, (7) generalizes the crisp evaluation $y = f(x_1, x_2, \dots, x_r)$ to the fuzzy evaluation $\tilde{B} = f(\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_r)$.

3 Problem formulation

We focus on linear regression applied to observed fuzzy data with multiple explanatory variables, whose parameters will be derived using the l_2 norm, and whose predictions will be in full accordance to the extension principle. Both observed inputs and outputs are triangular fuzzy numbers, while the predicted outputs do not have an a priori imposed shape type.

By analogy to the crisp case, function $f_{\tilde{A}}(\tilde{x}) = \tilde{a}_0 + \tilde{a}_1\tilde{x}_1 + \tilde{a}_2\tilde{x}_2 + \dots + \tilde{a}_k\tilde{x}_k$ predicts the fuzzy output $\tilde{y} = f_{\tilde{A}}(\tilde{x})$ with respect to the parameters $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$ at fuzzy input \tilde{x} . Applying the fuzzy regression in a complete accordance to the extension principle means finding all sets of crisp values a_0, a_1, \dots, a_k that are parameters of at least one crisp regression derived from at least one set of crisp input/output observations $x_{ij} \in [\tilde{x}_{ij}]_\alpha$, $y_i \in [\tilde{y}_i]_\alpha$, $i = 1, \dots, n$ and $j = 1, \dots, k$. The membership degree of such set of values a_0, a_1, \dots, a_k in the fuzzy sets $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$ respectively, must be

$$\underbrace{\max}_{x, y \text{ that yields } a} \left(\min \left(\left\{ \mu_{\tilde{x}_{ij}}(x_{ij}) \mid i = \overline{1, n}, j = \overline{1, k} \right\} \cup \left\{ \mu_{\tilde{y}_i}(y_i) \mid i = \overline{1, n} \right\} \right) \right), \tag{8}$$

where a stands for $\{a_0, a_1, \dots, a_k\}$, x stands for $\{x_{ij} \mid i = \overline{1, n}, j = \overline{1, k}\}$, and y stands for $\{y_i \mid i = \overline{1, n}\}$.

Within this study we aim to determine the accurate fuzzy-valued coefficients a of the regression and predicted outputs \tilde{y} . In the next section we reach the goal by formulating and solving certain pairs of optimization models.

4 Our approach

The involvement of the extension principle in defining the fuzzy regression is not new in the literature (see for instance the recent survey [5]). However, so far, the complexity of (8) determined the authors to search for simplified approaches that introduced certain deviations from a complete compliance with the extension principle.

Building around (8), we adapt a methodology already used within fuzzy mathematical programming problems ([14]), and focus on how to find the values a_0, a_1, \dots, a_k with a membership degree at least α , α arbitrary fixed in the interval $[0, 1]$. Wu [19] proved that the α -cuts of fuzzy-valued functions can be computed with the help of the α -cuts of its coefficients and arguments. In accordance

to Proposition 3.3 [19], we impose the following box constraints (9) on variables x and y to keep them varying within the ranges of their corresponding α -cuts.

$$\begin{aligned} (1 - \alpha) (\tilde{x}_{ij})^L + \alpha (\tilde{x}_{ij})^C &\leq x_{ij} \leq \alpha (\tilde{x}_{ij})^C + (1 - \alpha) (\tilde{x}_{ij})^U, & i = 1, \dots, n, \quad j = 1, \dots, k \\ (1 - \alpha) (\tilde{y}_i)^L + \alpha (\tilde{y}_i)^C &\leq y_i \leq \alpha (\tilde{y}_i)^C + (1 - \alpha) (\tilde{y}_i)^U, & i = 1, \dots, n. \end{aligned} \tag{9}$$

We use the upper indexes L, C, U to identify the lower, center and upper values of each involved triangular fuzzy number.

Further on, we use the relation between the crisp x, y and a provided by the crisp linear regression theory (4) as additional constraint; and propose the pair of models

$$\begin{aligned} \min (\max) \quad & a_q \\ \text{s.t.} \quad & X^T X A = X^T Y, \\ & (1 - \alpha) (\tilde{x}_{ij})^L + \alpha (\tilde{x}_{ij})^C \leq x_{ij} \leq \alpha (\tilde{x}_{ij})^C + (1 - \alpha) (\tilde{x}_{ij})^U, & i = 1, \dots, n, \\ & & j = 1, \dots, k, \\ & (1 - \alpha) (\tilde{y}_i)^L + \alpha (\tilde{y}_i)^C \leq y_i \leq \alpha (\tilde{y}_i)^C + (1 - \alpha) (\tilde{y}_i)^U, & i = 1, \dots, n, \\ & a_j \text{ free variable,} & j = 0, \dots, k, \end{aligned} \tag{10}$$

able to derive the α -cuts of the fuzzy parameter $\tilde{a}_q, q = 0, \dots, k$ of the fuzzy regression. Therefore, models (10) can be used to construct the fuzzy sets $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$; and based on them, we propose Algorithm 1 (EPBRC) that derives numerically the left and right sides of the membership functions of the fuzzy linear regression's coefficients in full accordance with the extension principle. The input parameters for EPBRC are as follows: the sequence $\alpha_1, \alpha_2, \dots, \alpha_p$ of values from $[0, 1]$ representing the desired membership levels; the fuzzy numbers representing the observed inputs \tilde{x}_{ij} and outputs $\tilde{y}_i, i = 1, \dots, n, j = 1, \dots, k$. The outputs of EPBRC are the matrices that describe the α -cut intervals of the membership functions of the coefficients \tilde{A} .

Algorithm 1 Extension-Principle-Based algorithm for **R**egression **C**oefficients (**EPBRC**)

Input: $\alpha_1, \alpha_2, \dots, \alpha_p \in [0, 1]; \tilde{x}_{ij}, \tilde{y}_i, i = 1, \dots, n, j = 1, \dots, k$.

- 1: Define the matrices X and Y using (3).
- 2: **for** $s = \overline{1, p}$ **do**
- 3: **for** $q = \overline{0, k}$ **do**
- 4: Solve min problem (10) and denote by a_{qs}^L the minimal value of the objective function.
- 5: Solve max problem (10) and denote by a_{qs}^U the maximal value of the objective function.
- 6: **end for**
- 7: **end for**

Output: Matrices $a^L = \left(a_{qs}^L \right)_{q=\overline{0, k}}^{s=\overline{1, p}}, a^U = \left(a_{qs}^U \right)_{q=\overline{0, k}}^{s=\overline{1, p}}$.

However, the fuzzy sets $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_k$ defined by Models (10) and derived by EPBRC cannot be used directly (i.e. through formal arithmetic) to derive the predicted outputs at a given input \tilde{v} . Another pair of models, namely

$$\begin{aligned} \min (\max) \quad & a_0 + a_1 v_1 + a_2 v_2 + \dots + a_k v_k \\ \text{s.t.} \quad & X^T X A = X^T Y, \\ & (1 - \alpha) (\tilde{x}_{ij})^L + \alpha (\tilde{x}_{ij})^C \leq x_{ij} \leq \alpha (\tilde{x}_{ij})^C + (1 - \alpha) (\tilde{x}_{ij})^U, & i = 1, \dots, n, \\ & & j = 1, \dots, k, \\ & (1 - \alpha) (\tilde{v}_j)^L + \alpha (\tilde{v}_j)^C \leq v_j \leq \alpha (\tilde{v}_j)^C + (1 - \alpha) (\tilde{v}_j)^U, & j = 1, \dots, k, \\ & (1 - \alpha) (\tilde{y}_i)^L + \alpha (\tilde{y}_i)^C \leq y_i \leq \alpha (\tilde{y}_i)^C + (1 - \alpha) (\tilde{y}_i)^U, & i = 1, \dots, n, \\ & a_j \text{ free variable,} & j = 0, \dots, k, \end{aligned} \tag{11}$$

must be utilized for this purpose. For a fixed α , Models (11) determine the left and right endpoints of the α -cut of the predicted fuzzy output at the fuzzy input \tilde{v} . When the given fuzzy input is one

of the observed inputs, for instance \tilde{x}_h , then the variables $v_j, j = 1, \dots, k$ and their box constraints can be ignored in Models (11); and the variables $x_{hj}, j = 1, \dots, k$, are used instead of $v_j, j = 1, \dots, k$ in the objective function.

Algorithm 2 (EPBRO) uses Models (11), and derives the estimated fuzzy outputs in full accordance with the extension principle. The values of $\tilde{v}_j, j = 1, \dots, k$ are inputs to Algorithm 2 together with all inputs of Algorithm 1. The outputs of Algorithm 2 are the vectors $\hat{y}^L = (\hat{y}_s^L)_{s=1, \dots, p}, \hat{y}^U = (\hat{y}_s^U)_{s=1, \dots, p}$ representing the left and right sides of the membership function of the fuzzy set that is the evaluation of the predictor function at \tilde{v} .

Algorithm 2 Extension-Principle-Based algorithm for Regression Outputs (EPBRO)

Input: $\alpha_1, \alpha_2, \dots, \alpha_p \in [0, 1]; \tilde{x}_{ij}, \tilde{y}_i, \tilde{v}_j, i = 1, \dots, n, j = 1, \dots, k$.

- 1: Define the matrices X and Y using (3).
- 2: **for** $s = \overline{1, p}$ **do**
- 3: **for** $i = \overline{1, n}$ **do**
- 4: Solve min problem (11) and denote by \hat{y}_s^L the minimal value of the objective function.
- 5: Solve max problem (11) and denote by \hat{y}_s^U the maximal value of the objective function.
- 6: **end for**
- 7: **end for**

Output: Vectors $\hat{y}^L = (\hat{y}_s^L)_{s=1, \dots, p}, \hat{y}^U = (\hat{y}_s^U)_{s=1, \dots, p}$.

Neither the coefficients, nor the predicted outputs derived by running Algorithms 1 and 2, respectively are triangular fuzzy numbers. However, their accurate shapes can be approximated by the triangular fuzzy numbers $\bar{a}_q = (a_{q1}^L, a_{qp}^L, a_{q1}^U), q = \overline{0, k}$, and $\bar{y} = (y_1^L, y_p^L, y_1^U), i = \overline{1, n}$.

4.1 Generalizations to fuzzy polynomial regression analysis

Models (10) and (11) can be generalized to derive the coefficients and the predicted outputs of an fuzzy polynomial regression. Models fully complying to the extension principle and needed in the quadratic regression analysis of a crisp in – fuzzy out system with a single observed explanatory variable were provided in [12]. In this section we generalize the crisp formula provided in [12] for a quadratic regression to the formula needed to model the fuzzy polynomial regression analysis of degree p , i.e. we propose the constraint system

$$\begin{pmatrix} \sum_{i=1}^n x_i^{2p} & \sum_{i=1}^n x_i^{2p-1} & \dots & \sum_{i=1}^n x_i^{p+1} & \sum_{i=1}^n x_i^p \\ \sum_{i=1}^n x_i^{2p-1} & \sum_{i=1}^n x_i^{2p-2} & \dots & \sum_{i=1}^n x_i^p & \sum_{i=1}^n x_i^{p-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^n x_i^{p+1} & \sum_{i=1}^n x_i^p & \dots & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i^p & \sum_{i=1}^n x_i^{p-1} & \dots & \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} a_p \\ a_{p-1} \\ \vdots \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i^p y_i \\ \sum_{i=1}^n x_i^{p-1} y_i \\ \vdots \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}, \quad (12)$$

that we will further use within our optimization models. Observing (12) from the point of view of a crisp regression analysis, the scalars a_0, a_1, \dots, a_p are the coefficients of the predictor function

$$f(x, a_0, a_1, \dots, a_p) = a_0 + a_1x + \dots + a_px^p,$$

that is a polynomial of degree p in x ; and the scalars x_i and $y_i, i = 1, \dots, n$ used in (12) are the crisp observed data. Observing the same matrix equality (12) from the point of view of a fuzzy polynomial regression, all scalar quantities $a_q, q = 0, \dots, p, x_i$ and $y_i, i = 1, \dots, n$ are seen as variables that range within their corresponding α -cuts.

Table 1: Observed and predicted fuzzy data based on regression analysis for the first example recalled from [11]. The EPBRO predictions are reported as triangular fuzzy numbers approximations

Index	Observed inputs	Observed outputs	HBS2 predictions	EPBRO predictions
1	(1.5, 2.0, 2.5)	(3.5, 4.0, 4.5)	(3.75, 4.20, 4.72, 5.18)	(3.113, 4.611, 5.935)
2	(3.0, 3.5, 4.0)	(5.0, 5.5, 6.0)	(4.50, 4.96, 5.50, 6.00)	(4.062, 5.390, 6.496)
3	(4.5, 5.5, 6.5)	(6.5, 7.5, 8.5)	(5.25, 5.76, 6.81, 7.36)	(5.160, 6.429, 7.618)
4	(6.5, 7.0, 7.5)	(6.0, 6.5, 7.0)	(6.25, 6.81, 7.33, 7.91)	(6.087, 7.208, 8.218)
5	(8.0, 8.5, 9.0)	(8.0, 8.5, 9.0)	(7.00, 7.59, 8.12, 8.73)	(6.844, 7.987, 8.957)
6	(9.5, 10.5, 11.5)	(7.0, 8.0, 9.0)	(7.75, 8.38, 9.43, 10.10)	(7.750, 9.026, 10.272)
7	(10.5, 11.0, 11.5)	(10.0, 10.5, 11.0)	(8.25, 8.90, 9.43, 10.10)	(8.072, 9.285, 10.547)
8	(12.0, 12.5, 13.0)	(9.0, 9.5, 10.0)	(9.00, 9.68, 10.20, 10.90)	(8.735, 10.064, 11.406)

Denoting by \tilde{x}_i and \tilde{y}_i , $i = 1, \dots, n$ the observed triangular fuzzy numbered inputs and outputs, respectively, we formulate the box constraints system (13)

$$\begin{aligned}
 (1 - \alpha) (\tilde{x}_i)^L + \alpha (\tilde{x}_i)^C &\leq x_i \leq \alpha (\tilde{x}_i)^C + (1 - \alpha) (\tilde{x}_i)^U, & i = 1, \dots, n, \\
 (1 - \alpha) (\tilde{y}_i)^L + \alpha (\tilde{y}_i)^C &\leq y_i \leq \alpha (\tilde{y}_i)^C + (1 - \alpha) (\tilde{y}_i)^U, & i = 1, \dots, n,
 \end{aligned}
 \tag{13}$$

that corresponds to their α -cut intervals, $\alpha \in [0, 1]$.

The optimization models that derive the endpoints of the α -cut intervals of the regression's coefficient a_q , $q = 0, \dots, p$ minimize and maximize, respectively the objective function a_q over the feasible set defined by (12) and (13), with $x_i, y_i, i = 1, \dots, n$ bounded, and a_0, a_1, \dots, a_p free variables.

In the same manner, the optimization models that derive the endpoints of the α -cut intervals of the predicted output \hat{y}_h , $h = 1, \dots, n$ minimize and maximize, respectively the objective function $f(x_h, a_0, a_1, \dots, a_p)$ over the feasible set defined by (12) and (13). Moreover, a predicted output at a given fuzzy input \tilde{v} , other than the previously observed ones, can be determined by minimizing and maximizing, respectively the objective function $f(v, a_0, a_1, \dots, a_p)$ over the feasible set defined by (12) and (13) and the additional box constraint on v , i.e.

$$(1 - \alpha) \tilde{v}^L + \alpha \tilde{v}^C \leq v \leq \alpha \tilde{v}^C + (1 - \alpha) \tilde{v}^U.$$

5 Computation results

In this section we report the numerical results aiming to illustrate the theoretical statements and provide a comparison with the results found in the literature. Both considered examples are recalled from the literature and use triangular fuzzy numbers to describe the observed data. We derive the predicted fuzzy outputs numerically, for eleven equidistant α -cuts, $\alpha \in \{0, 0.1, 0.2, \dots, 0.9, 1\}$.

Our experiments include one instance with a single fuzzy explanatory variable; and one instance with two fuzzy explanatory variables. The output estimations are derived using a linear predictor function in both cases.

5.1 One explanatory fuzzy variable

The first example containing one explanatory fuzzy variable is taken from [11]. Hojati et al. [6] solved the same example, and we use their results for comparison. The fuzzy observed data is reported in Table 1 together with the estimated fuzzy outputs which are also graphed in Figure 1.

Table 1 reports the triangular fuzzy number approximations of the predicted fuzzy outputs derived by our approach. HBS2 is the approach proposed by Hojati et al. [6]. HBS2 derived two intervals that together define the trapezoidal fuzzy predicted outputs included in the comparison.

Figure 1 also illustrates how the application of the Monte Carlo method in the simulation of the extension principle [13] proves that the results of Hojati et al. [6] do not fully comply to the extension principle, thus EPBRO improves their results. More precisely, analyzing the predicted outputs y_1, y_2, y_7 and y_8 one may notice the simulated values that are out of the membership functions of the predicted outputs reported in [6].

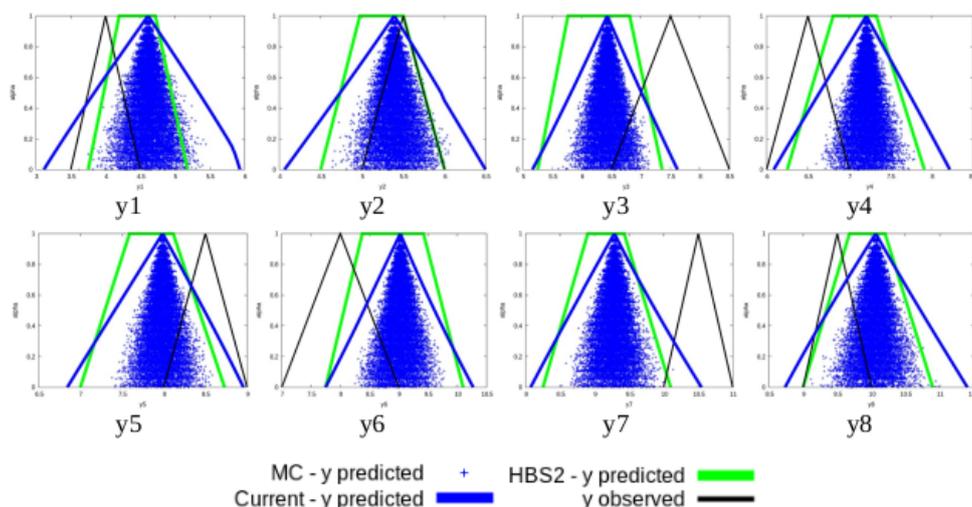


Figure 1: Graphic illustration of the predicted fuzzy outputs provided by Hojati et al. [6], the Monte Carlo simulation [13], and our algorithm EPBRO

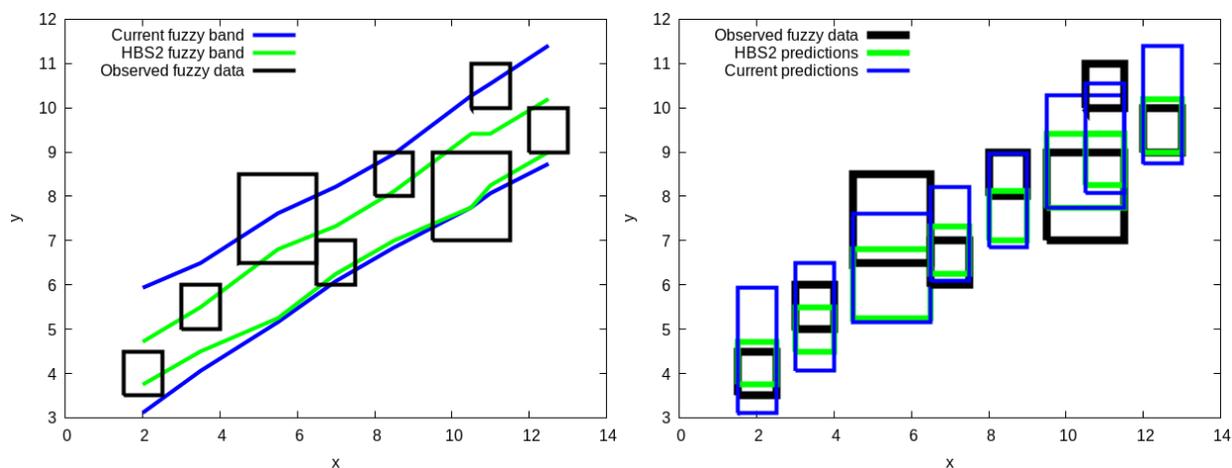


Figure 2: Hojati et al.’s approach [6] versus EPBRO through predicted fuzzy bands and squared representations for the observed data reported in Table 1

Another way to visually compare the derived predictions is to graph the fuzzy bands covered by the regression’s outputs. Figure 2 compares the fuzzy bands obtained in [6] and by our approach (on the left), and the predicted outputs drawn as squares whose horizontal edges represent the supports of input data, while the vertical edges represent the supports of the predicted data (on the right). A wider predicted fuzzy bands assure a better cover of the observed outputs.

Figure 3 is similar to Figure 2 but reports our results compared to the results of Kao and Chyu [9]. The results are quite similar, showing that Kao and Chyu’s methodology [9] respects in a high extent the extension principle. In fact, Kao and Chyu [9] provided approximate predictions that can be improved by increasing the number of analyzed α -cuts. The advantage of our approach is in providing the exact support of the predicted outputs, by solving two optimization problems that uses polynomials of degree three in defining the feasible set, and polynomials of degree two in defining the objective functions.

The regression’s coefficients derived by HBS2 [6] and our approach EPBRC are shown in Figure 4. EPBRC derives fuzzy sets \tilde{a}_0, \tilde{a}_1 with wider supports than HBS2. However, HBS2 used their coefficients \tilde{a}_0, \tilde{a}_1 in arithmetic expressions; and obtained output predictions with wider supports than those derived by our approach through the EPBRO algorithm. This experiment shows once again the importance of involving the extension principle in all steps of the analysis. The coefficients \tilde{a}_0, \tilde{a}_1 derived by EPBRC provide information about all possible crisp values of the regression’s coefficients and their membership levels, thus one can use them to study the whole system.

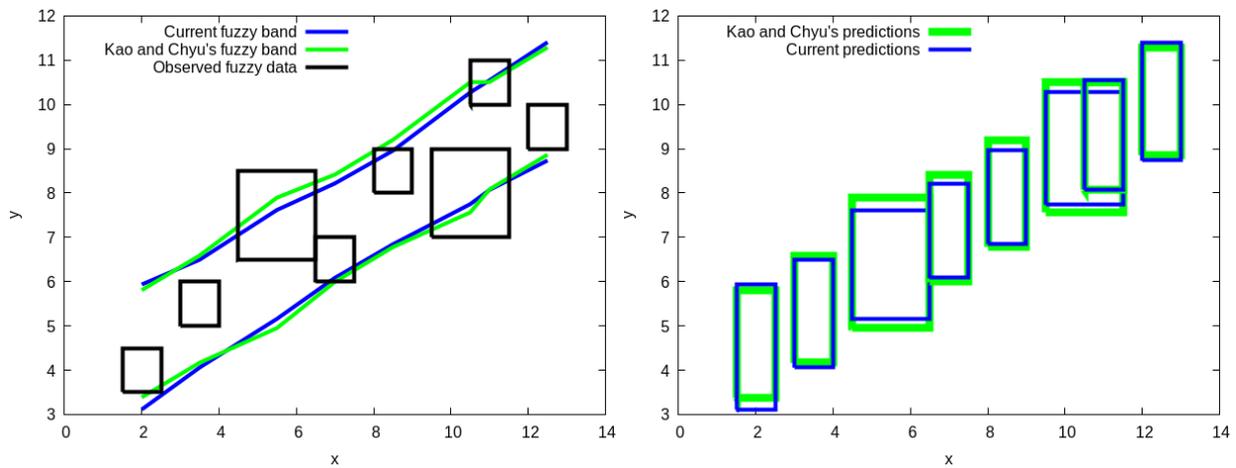


Figure 3: Kao and Chyu’s predictions [9] versus EPBRO predictions for the observed data reported in Table 1. In other words, approximate versus full compliance with the extension principle

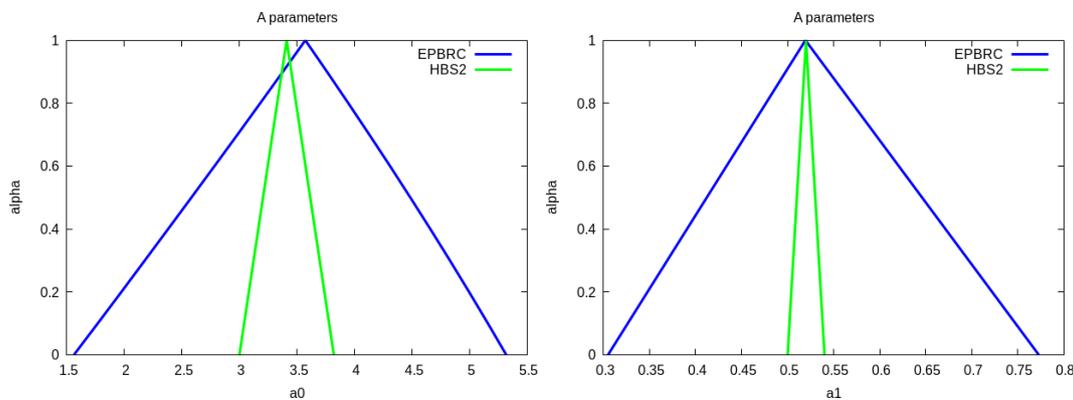


Figure 4: Visual comparison of the regression’s coefficients for data given in Table 1

5.2 Two explanatory fuzzy variables

The second example has two explanatory fuzzy variables and is recalled from [19]. The observed fuzzy data is reported in Table 2.

The estimated fuzzy outputs obtained by our approach (EPBRO), Wu’s approach [19], and Chen and Nien’s approach [4] are included in Table 3, and graphed in Figure 5. Wu’s approach [19] was used for comparison since their methodology partly followed the extension principle deviating from it to reduce the computation complexity. On the other side, we also involve in the comparison Chen and Nien’s approach [4], to emphasize their higher deviation from the accurate results. Chen and Nien’s approach is also relevant for the comparison since it was recently published.

In Figure 5 we additionally report the results obtained by the Monte Carlo simulation procedure [13] that clearly illustrate the difference between the results obtained in [19] and [4]: only a few combinations of the parameters yielded estimated values of y_{10} out of the membership function provided in [19], but many combination of the same parameters derived many predicted values for all outputs $y_1 - y_{10}$ out of the membership functions reported in [4].

6 Conclusions and further researches

In this paper we addressed the problem of predicting the fuzzy outputs of a fuzzy in – fuzzy out system through a fuzzy linear regression analysis. We developed our solution approach in full accordance to the extension principle. We proposed pairs of optimization models to be solved for each desired α -level degree of each estimated fuzzy output. Both the left and right endpoints of each α -cut interval are derived as minimum and maximum, respectively of all crisp values that can

Table 2: Observed fuzzy inputs and outputs for the numerical example recalled form [19]

Index	Observed inputs		Observed outputs
1	(151, 274, 322)	(1432, 2450, 3461)	(111, 162, 194)
2	(101, 180, 291)	(2448, 3254, 4463)	(88, 120, 161)
3	(221, 375, 539)	(2592, 3802, 5116)	(161, 223, 288)
4	(128, 205, 313)	(1414, 2838, 3252)	(83, 131, 194)
5	(62, 86, 112)	(1024, 2347, 3766)	(51, 67, 83)
6	(132, 265, 362)	(2163, 3782, 5091)	(124, 169, 213)
7	(66, 98, 152)	(1687, 3008, 4325)	(62, 81, 102)
8	(151, 330, 463)	(1524, 2450, 3864)	(138, 192, 241)
9	(115, 195, 291)	(1216, 2137, 3161)	(82, 116, 159)
10	(35, 53, 71)	(1432, 2560, 3782)	(41, 55, 71)
11	(307, 430, 584)	(2592, 4020, 5562)	(168, 252, 367)
12	(284, 372, 498)	(2792, 4427, 6163)	(178, 232, 346)
13	(121, 236, 370)	(1734, 2660, 4094)	(111, 144, 198)
14	(103, 157, 211)	(1426, 2088, 3312)	(78, 103, 148)
15	(216, 370, 516)	(1785, 2605, 4042)	(167, 212, 267)

Table 3: The predictions provided by our approach EPBRO and the approaches presented in [19] and [4]. The predicted outputs are reported as triangular fuzzy number approximations

Index	Wu's predictions (2003)	Chen - Nien predictions (2020)	EPBRO predictions
1	(40.731, 161.897, 254.675)	(104.24, 161.53, 177.82)	(-63.676, 161.53, 375.403)
2	(22.831, 122.669, 258.107)	(88.73, 122.62, 173.13)	(-72.171, 122.62, 366.204)
3	(80.253, 224.431, 397.099)	(155.81, 223.10, 288.00)	(-4.637, 223.10, 425.660)
4	(29.779, 131.242, 246.073)	(91.49, 131.17, 172.00)	(-82.896, 131.17, 374.616)
5	(-3.544, 67.701, 153.452)	(51.00, 68.46, 88.21)	(-99.875, 68.46, 266.596)
6	(35.861, 169.687, 306.367)	(102.02, 169.00, 209.95)	(-53.927, 169.00, 389.486)
7	(2.056, 79.734, 184.581)	(60.85, 80.24, 110.79)	(-85.174, 80.24, 295.228)
8	(41.247, 189.673, 334.309)	(105.30, 188.97, 243.41)	(-55.396, 188.97, 402.216)
9	(22.537, 119.833, 233.108)	(82.11, 120.01, 161.51)	(-83.392, 120.01, 356.569)
10	(-13.997, 53.293, 132.854)	(41.00, 54.18, 70.33)	(-98.220, 54.18, 247.551)
11	(120.828, 253.717, 428.6082)	(202.69, 252.00, 311.77)	(34.408, 252.00, 465.188)
12	(111.097, 228.693, 396.296)	(192.47, 227.20, 279.32)	(37.340, 227.20, 459.329)
13	(28.2686, 144.9806, 291.304)	(91.37, 144.77, 204.57)	(-60.669, 144.77, 383.842)
14	(18.0514, 100.5342, 195.2155)	(78.00, 100.95, 127.69)	(-77.172, 100.95, 334.607)
15	(73.3752, 210.9386, 364.751)	(143.75, 209.96, 268.30)	(-25.345, 209.96, 415.196)

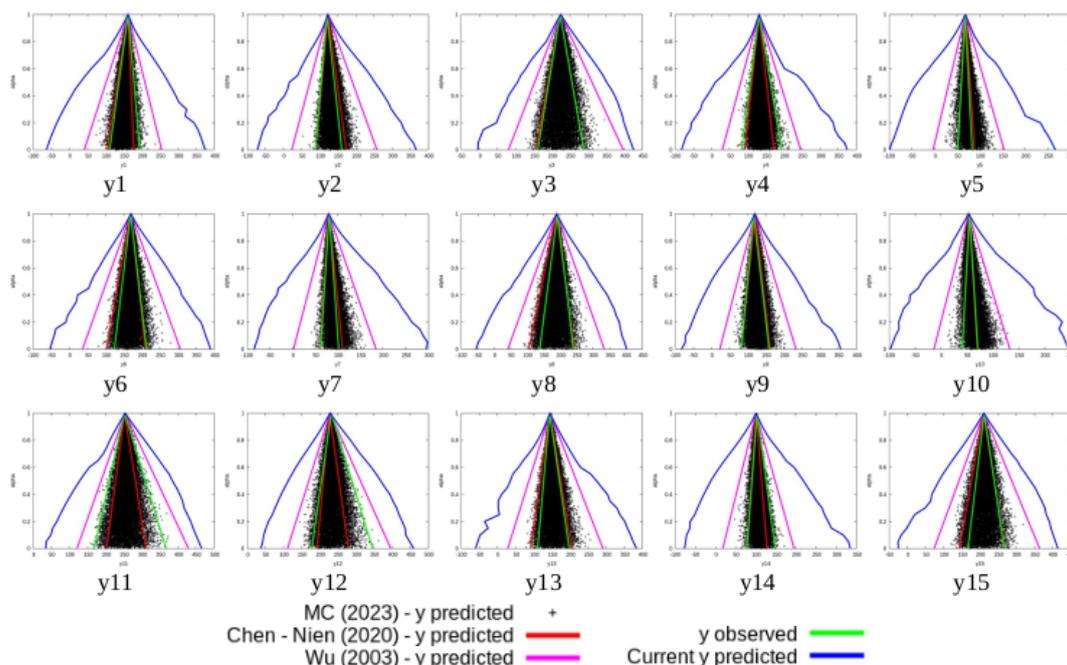


Figure 5: Visual comparison of the results obtained by our algorithm EPBRO and the results from the literature

be obtained as predicted outputs to at least one regression problem with observed crisp data in the α -cut ranges of the corresponding fuzzy observed data. Relevant examples from the literature were solved, and the carried out experiments were used to illustrate the theoretical findings.

The main advantage of the new proposed methodology is in the fact that the obtained results are realistic, being derived in full accordance to the extension principle, and better than the approximated results that can be found in the literature. The computation complexity of the approach is related to the optimization solvers that have to handle mathematical programming problems with quadratic objective functions and polynomial constraints of degree three.

We plan to continue the research, and study how to apply the crisp non linear regression models to estimate the behavior of fuzzy in – fuzzy out systems. Aiming to solve real life problems, we will pay attention on selecting the norm that best suits each specific problem, and incorporate within the approach a procedure able to remove the outliers that might be identified between the fuzzy observed data.

Acknowledgments

This work was partly supported by the Serbian Ministry of Science, Technological Development and Innovation through Mathematical Institute of the Serbian Academy of Sciences and Arts, and Faculty of Organizational Sciences, University of Belgrade.

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] O. Ban, L. Droj, Tuş D., G. Droj, and N. Bugnar. Data processing by fuzzy methods in social sciences researches. example in hospitality industry. *International Journal of Computers, Communications, Control*, 17(2):4741, 2022.
- [2] Eren Bas. Robust fuzzy regression functions approaches. *Information Sciences*, 613:419–434, 2022.
- [3] Jalal Chachi, S. Mahmoud Taheri, and Pierpaolo D’Urso. Fuzzy regression analysis based on m-estimates. *Expert Systems with Applications*, 187:115891, 2022.
- [4] Liang Hsuan Chen and Sheng Hsing Nien. A new approach to formulate fuzzy regression models. *Applied Soft Computing*, 86, January 2020.
- [5] Nataliya Chukhrova and Arne Johannssen. Fuzzy regression analysis: Systematic review and bibliography. *Applied Soft Computing*, 84:105708, 2019.
- [6] Mehran Hojati, C.R. Bector, and Kamal Smimou. A simple method for computation of fuzzy linear regression. *European Journal of Operational Research*, 166(1):172–184, 2005. Metaheuristics and Worst-Case Guarantee Algorithms: Relations, Provable Properties and Applications.
- [7] Dominik Hose and Michael Hanss. Fuzzy linear least squares for the identification of possibilistic regression models. *Fuzzy Sets and Systems*, 367:82–95, 2019. Theme: Uncertainty Management.
- [8] Cengiz Kahraman, Ahmet Beşkese, and F. Tunç Bozbura. *Fuzzy Regression Approaches and Applications*, pages 589–615. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [9] Chiang Kao and Chin-Lu Chyu. Least-squares estimates in fuzzy regression analysis. *European Journal of Operational Research*, 148(2):426–435, 2003. Sport and Computers.
- [10] Sorin Nădăban. From classical logic to fuzzy logic and quantum logic: a general view. *International Journal of Computers, Communications, Control*, 16(1):4125, 2021.
- [11] Masatoshi Sakawa and Hitoshi Yano. Multiobjective fuzzy linear regression analysis for fuzzy input-output data. *Fuzzy Sets and Systems*, 47(2):173–181, 1992.
- [12] Bogdana Stanojević and Milan Stanojević. Quadratic least square regression in fuzzy environment. *Procedia Computer Science*, 214:391–396, 2022. 9th International Conference on Information Technology and Quantitative Management.
- [13] Bogdana Stanojević and Milan Stanojević. Extension-principle-based approach to least square fuzzy linear regression. In Simona Dzitac, Domnica Dzitac, Florin Gheorghe Filip, Janusz Kacprzyk, Misu-Jan Manolescu, and Horea Oros, editors, *Intelligent Methods Systems and Applications in Computing, Communications and Control*, pages 219–228, Cham, 2023. Springer International Publishing.
- [14] Bogdana Stanojević, Milan Stanojević, and Sorin Nădăban. Reinstatement of the extension principle in approaching mathematical programming with fuzzy numbers. *Mathematics*, 9(11), 2021.
- [15] H. Tanaka, S. Uejima, and K. Asai. Linear regression analysis with fuzzy model. *IEEE Transactions on Systems Man and Cybernetics*, 12:903–907, 1982.
- [16] Hideo Tanaka, Isao Hayashi, and Junzo Watada. Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research*, 40(3):389–396, 1989.
- [17] Ning Wang, Marek Reformat, Wen Yao, Yong Zhao, and Xiaoqian Chen. Fuzzy linear regression based on approximate bayesian computation. *Applied Soft Computing*, 97:106763, 2020.

- [18] H. Wu and Z.S. Xu. Fuzzy logic in decision support: Methods, applications and future trends. *International Journal of Computers, Communications, Control*, 16(1):4044, 2021.
- [19] Hsien-Chung Wu. Linear regression analysis for fuzzy input and output data using the extension principle. *Computers & Mathematics with Applications*, 45(12):1849–1859, 2003.
- [20] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.
- [21] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning i. *Information Sciences*, 8(3):199 – 249, 1975.



Copyright ©2023 by the authors. Licensee Agora University, Oradea, Romania.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

Journal's webpage: <http://univagora.ro/jour/index.php/ijccc/>



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).

<https://publicationethics.org/members/international-journal-computers-communications-and-control>

Cite this paper as:

Stanojević, B.; Stanojević, M. (2023). Optimization-Based Fuzzy Regression in Full Compliance with the Extension Principle, *International Journal of Computers Communications & Control*, 18(2), 5320, 2023.

<https://doi.org/10.15837/ijccc.2023.2.5320>