Fuzzy Regression Analysis Based on Fuzzy Neural Networks Using Trapezoidal Data

R. Naderkhani[1] · M. H. Behzad[1] · T. Razzaghnia[2] · R. Farnoosh[3]

Abstract Fuzzy regression is a generalized regression model to represent the relationship between dependent and independent variables in a fuzzy environment. The fuzzy linear regression analysis seeks for regression models fitting well all the data based on a specific criterion. In this paper, an adaptive neuro-fuzzy inference system (ANFIS) is employed for the analysis and prediction of a nonparametric fuzzy regression function with non-fuzzy inputs and symmetric trapezoidal fuzzy outputs. To this end, two new hybrid algorithms are proposed in which the fuzzy least squares and linear programming have been used to optimize the secondary weights. The algorithms are applied to a multi-layered validation method to confirm the models' reliability. In addition, three methods of nonparametric fuzzy regression with crisp inputs and asymmetric trapezoidal fuzzy outputs, are compared. Three nonparametric techniques in statistics, namely local linear smoothing (L-L-S), K-nearest neighbor smoothing (K-NN) and kernel smoothing (K-S) with trapezoidal fuzzy data have been analyzed to obtain the best smoothing parameters. The performance of the models is illustrated through numerical examples and simulations. More specifically, the accuracy of the algorithms is confirmed by exhaustive simulations.

Abbreviations

MF_i	Membership function
$\widetilde{Y}_i$	A trapezoidal fizzy number
$o_{k.i}$	Placed as layers
$d^2\left(\widetilde{Y}_i \cdot \widehat{Y}_i\right)$	Diamond distance measures
$\left(\widehat{\widetilde{Y}_i}\right)$	A predicted of a trapezoidal fuzzy number
$\omega_j(x)$	Weight sequence at x
CV	Cross-validation
$k(.)$	Kernel smoothing

1 Introduction

Regression is a very powerful method to understand the relationship between the dependent and independent variables. The theory of fuzzy sets was introduced in 1965 by Zadeh [1], who generalized the theory of formal and conventional collections. Fuzzy sets are potentially a wider field of operation, especially in the area of information processing and classification patterns. Fuzzy theory is able to infer, control, and make decisions against uncertainties, especially, for concepts of inaccurate and vague nature. Fuzzy regression, which firstly introduced by Tanaka et al. [2], is a simple modification of regression that it demonstrates the relationship among variables in a fuzzy environment. The expression "linear regression model" points

✉ T. Razzaghnia
razaghnia@riau.ac.ir

1 Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran

2 Department of Statistics, North Tehran Branch, Islamic Azad University, Tehran, Iran

3 School of Mathematics, Iran University of Science and Technology, Narmak, 16846 Tehran, Iran

out to situations where fuzzy membership functions are used to describe behaviors of the parameters instead of the probability distribution functions. The concept became more known by other researchers afterwards, for example, fuzzy regression methods by Ishobishi [3], and the least squares method by Diamond [4], chang and Lee [5] and Yang et al. [6].

In 1996, Karacapulos [7] showed that fuzzy neural systems, i.e., the integration of neural networks and fuzzy logic, have capabilities far beyond those of either system individually. This advantage has been mirrored in studies such as Ishobishi and Tanaka [8], Cheng and Lee [9]. Arnold [10] combined the neural network, fuzzy logic and genetic algorithm. Mosleh. et al. [11] used the fuzzy Neural Networks to evaluate fuzzy linear regression, fuzzy polynomial regression, fuzzy matrix equations, simulation and evaluation of fuzzy derivative equations. Statistical nonparametric smoothing techniques have achieved significant development in recent years. Danesh. et al. [12] used ANFIS in fuzzy nonparametric regression with triangular numbers in order to predict and estimate the data based on the least squares regression method. A regression analysis of fuzzy trapezoidal numbers with both fuzzy inputs and fuzzy outputs was performed by Sun and Lu [13].

Cheng and Lee [14] have extended the K-nearest neighbor (K-NN) and kernel smoothing (K-S) methods for the context of fuzzy nonparametric regression. Razzaghnia et al. [15, 16] used hybrid methods in nonparametric regression with triangular fuzzy data. Also, Razzaghnia [17] upgraded the neural network in fuzzy linear regression to avoid the effect of outliers. Skrjanc [18] focused on evolving fuzzy rule-based models and neuro-fuzzy networks for clustering, classification and regression and system identification, in online and real-time environments. Junhong et al. [19] proposed a new fuzzy regression model based on trapezoidal fuzzy number and least absolute deviation method. Deng and Zhao [20] used The adaptive control parameters to make the relatively uniform distribution, and solved the contradiction between expanding search and finding optimal solution. Khosravi and shahabi [21] introduced three Multi-Criteria Decision-Making (MCDM) analysis techniques, along with two machine learning methods. Liu et al. [22] proposed a new algorithm and as compared with some commonly used learning algorithms, such as support vector machine, the k-nearest neighbors and other combination algorithms.

In this paper, a fuzzy regression algorithm is proposed based on ANFIS whose aim is to reduce errors compared to other algorithms. The input data are assumed to be of fuzzy nature. The output estimated parameters must be better or at least overlap with the rival algorithms. The choice of trapezoidal membership function (MF) is due to their lower ambiguities against to the other MFs.

The advantages of the present study with respect to many of previous works can be summarized as follows:

One of the advantage of this work is to use trapezoidal data. In this paper, we extended the triangular fuzzy numbers to trapezoidal fuzzy numbers for avoiding unnecessary fuzziness of the linear fuzzy model in ANFIS. When a large membership grades are given to many observations, the linear fuzzy model with triangular fuzzy coefficients tends to have large fuzziness. In some cases, such large fuzziness numbers can be avoiding by using trapezoidal fuzzy numbers as fuzzy coefficients. The other advantage is to use V-fold Cross Validation technique for training ANFIS network. The proposed method and the smoothing methods are compared with the CV scale. The conducted simulation of experiments are shown that the performance of the proposed method is better than that of the smoothing methods, which reduces the CV. In the proposed approach when the observation numbers are increased, compared to the existing smoothing methods, the accuracy is increased. Generally, the proposed method (V-fold cross validation technique with trapezoidal data) reduces the fuzziness of the system and it has faster adaptation.

The rest of the paper is as follows. Section 2 presents the preliminary math of nonparametric fuzzy regression and trapezoidal MF needed for the sections afterwards. The validation method, smoothing methods for trapezoidal fuzzy numbers, ANFIS and the projected hybrid algorithm will be presented, respectively in Sects. 3 and 4 and 5. Section 6 gives some numerical examples receptively. Finally, Sect. 7 concludes the paper.

2 Definitions

In this section, the basic notations used in the paper are defined. We start by trapezoidal fuzzy number.

A fuzzy number $\tilde{A}$ is a convex normalized fuzzy subset of the real line **R** with an upper semi-continuous membership function of bounded support [23].

Definition 1 A symmetric fuzzy number $\tilde{A}$, denoted by $\tilde{A} = (\alpha.c)_L$ is defined as.

$$\tilde{A}(x) = L\left(\frac{(x - \alpha)}{c}\right) c > 0.$$

where α and c are the center and spread of $\tilde{A}$ and $L(x)$, a shape function of fuzzy numbers such that:

  i. $L(x) = L(-x)$;
  ii. $L(0) = 1. L(1) = 0$;
  iii L is strictly decreasing on $[0, \infty)$;
  iv. L is invertible on $[0, 1]$.

The set of all symmetric fuzzy numbers is denoted by $F_L(R)$. Specifically, If $L(x) = 1 - |x|$ the fuzzy number is a symmetric triangular fuzzy number [23].

Definition 2 An asymmetric trapezoidal fuzzy number $\tilde{A}$, denoted by $\tilde{A} = (a^{(1)}, a^{(2)}, a^{(3)}, a^{(4)})$, is defined as [24]:

$$
\tilde{A}(x) = \begin{cases} L\left(\dfrac{a^{(2)} - x}{a^{(2)} - a^{(1)}}\right); & x < a^{(2)}. \\ 1; & a^{(2)} \leq x \leq a^{(3)}. \\ R\left(\dfrac{x - a^{(3)}}{a^{(4)} - a^{(3)}}\right); & x > a^{(3)}. \end{cases} \tag{1}
$$

where $a^{(1)}, a^{(2)} a^{(3)}, a^{(4)}$ are four parameters of the asymmetric trapezoidal fuzzy number. (See Fig. 1).

Definition 3 For two trapezoidal fuzzy numbers $\tilde{A} = (a^{(1)}.a^{(2)}.a^{(3)}.a^{(4)})$ and $\tilde{A} = (b^{(1)}.b^{(2)}.b^{(3)}.b^{(4)})$, the Diamond distance [4] between $\tilde{A}$ and $\tilde{B}$ can be expressed as:

$$
d^2(\tilde{A}.\tilde{B}) = \left(a^{(1)} - b^{(1)}\right)^2 + \left(a^{(2)} - b^{(2)}\right)^2 + \left(a^{(3)} - b^{(3)}\right)^2 + \left(a^{(4)} - b^{(4)}\right)^2 \tag{2}
$$

The Diamond distance measures the closeness of the MFs of the two trapezoidal fuzzy numbers. The definition three satisfies the distance properties, i.e.

- $d^2(\tilde{A}.\tilde{B}) \geq 0$;
- $d^2(\tilde{A}.\tilde{B}) = 0 \iff \tilde{A} = \tilde{B}$. (i.e. $a^{(1)} = b^{(1)}.a^{(2)} = b^{(2)}.$ $a^{(3)} = b^{(3)}.a^{(4)} = b^{(4)}$));
- $d^2(\tilde{A}.\tilde{B}) \leq d^2(\tilde{A}.\tilde{C}) + d^2(\tilde{C}.\tilde{B})$.

Definition 4 Suppose that $F = \{Y : Y = (Y^{(1)}.Y^{(2)}.Y^{(3)}.Y^{(4)})\}$ be a set of all trapezoidal fuzzy numbers. The following univariate fuzzy regression model is considered by,

$$
\tilde{Y} = F(x)\{+\}\varepsilon = \left(Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x).Y^{(4)}(x)\right)\{+\}\varepsilon. \tag{3}
$$

In which, $\tilde{Y}$ is the dependent trapezoidal fuzzy number (output), variable X is an independent variable (input), ε is the error and $\{+\}$ is an operator whose definition depends on the applied fuzzy ranking method. X's domain is assumed to be D. Thus, $F(x)$ is a mapping $D \to F$.

Definition 5 The CV is a measure for the difference between a fuzzy regression function $\widehat{Y}_i = \left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}\right)$ and its estimation $\widehat{\widehat{Y}}_i = \left(\hat{Y}_i^{(1)}.\hat{Y}_i^{(2)}.\hat{Y}_i^{(3)}.\hat{Y}_i^{(4)}\right)$. Based on the Diamond distance, the CV is defined as,

$$
CV = \frac{1}{n}\sum_{i=1}^{n} d^2(\tilde{Y}_i.\widehat{\tilde{Y}}_i)
$$
$$
= \frac{1}{n}\sum_{i=1}^{n}\left(\left(Y_i^{(1)} - \hat{Y}_i^{(1)}\right)^2 + \left(Y_i^{(2)} - \hat{Y}_i^{(2)}\right)^2 \right.
$$
$$
\left. + \left(Y_i^{(3)} - \hat{Y}_i^{(3)}\right)^2 + \left(Y_i^{(4)} - \hat{Y}_i^{(4)}\right)^2\right). \tag{4}
$$

Remark 1 In k-fold cross-validation (CV), the original samples are randomly partitioned into k equal size subsamples. Of these, one subsample is retained as the validation data for testing the model, and the remaining $k - 1$ subsamples are used as training data. The CV process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation. In total, k fit models and k validation statistics are derived. The model giving the best validation statistic is chosen as the final model. This method is specifically useful for small data sets, because it makes efficient use of the limited data samples. In this paper we assume $k = 5$.

3 Smoothing Methods for Trapezoidal Fuzzy Numbers

The basic idea of smoothing is that if a function f is fairly smooth, then the observations made at and near x should contain information about value of x. Thus, it should be possible to use local averaging of the data x to construct an estimator for $F(x)$ which is called the smoother. There are several smoothing techniques. We proposed K-NNK-S and (L-L-S) methods for trapezoidal variable in this section.

In the following discussion, asymmetric trapezoidal fuzzy numbers are applied as asymmetric trapezoidal membership functions for deriving nonparametric regression model based on the smoothing parameters.

These models are considered as univariate fuzzy nonparametric regression model:

$$
\tilde{Y} = F(x)\{+\}\varepsilon = \left(Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x).Y^{(4)}(x)\right)\{+\}\varepsilon. \tag{5.}
$$

where Y is a trapezoidal fuzzy dependent variable as output and x is a crisp independent variable as the input whose domain and range are assumed to be D and $\mathbb{R}$. Also $F(x)$ is the mapping $D \to F$. The definition of the three smoothing methods for trapezoidal fuzzy variables is as follows:

3.1 Local Linear Smoothing Method (L-L-S)

Razzaghnia et al. [25] proposed the first linear regression analysis with trapezoidal coefficients. The asymmetric trapezoidal fuzzy numbers are applied as asymmetric trapezoidal membership functions for deriving bivariate regression model. A univariate regression model can be expressed as

$$\widehat{\widetilde{Y}}_i = \tilde{A}_0 + \tilde{A}_1 X_i$$
$$= \left(a_0^{(1)}.a_0^{(2)}.a_0^{(3)}.a_0^{(4)}\right) + \left(a_1^{(1)}.a_1^{(2)}.a_1^{(3)}.a_1^{(4)}\right)X_i. \quad (6)$$

This model can be rewritten as

$$\widehat{\widetilde{Y}}_i = \left(a_0^{(1)} + a_1^{(1)}X_i.a_0^{(2)} + a_1^{(2)}X_i.a_0^{(3)} + a_1^{(3)}X_i.a_0^{(4)} + a_1^{(4)}X_i\right). \quad (7)$$

where $i = 1.2.\ldots.n$, is the sample size and $\widetilde{Y}_i = \left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}\right)$ is the observed value for $i = 1.2.\ldots.n$. So $\widehat{\widetilde{Y}}_{i.L}$ and $\widehat{\widetilde{Y}}_{i.R}$ are the left bound and right bound of the predicted $\widehat{\widetilde{Y}}_i$ at membership level h. Also $\widetilde{Y}_{i.L}$ and $\widetilde{Y}_{i.R}$ are left bound and right bounds of observed $\widetilde{Y}_i$ at membership level h. Thereupon,

$$\widehat{\widetilde{Y}}_{i.L} = ha_0^{(2)} + ha_1^{(2)}X_i + (1-h)a_0^{(1)} + (1-h)a_1^{(1)}X_i. \quad (8)$$

$$\widehat{\widetilde{Y}}_{i.R} = ha_0^{(3)} + ha_1^{(3)}X_i + (1-h)a_0^{(4)} + (1-h)a_1^{(4)}X_i. \quad (9)$$

$$\widehat{\widetilde{Y}}_{i.L} = hY_i^{(2)} + (1-h)Y_i^{(1)}. \quad (10)$$

$$\widehat{\widetilde{Y}}_{i.R} = hY_i^{(3)} + (1-h)Y_i^{(4)}. \quad (11)$$

Let $\left(X_i.\widehat{\widetilde{Y}}_i\right)$ be a sample of the observed crisp inputs and trapezoidal fuzzy outputs with underlying fuzzy regression function of model (6).

$F(x)$ is estimated at any $x \in D$ based on $(x_i.\tilde{Y}_i)$ for $i = 1.2.\ldots.n$. When the local linear smoothing technique is used, we shall estimate $Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x)$ and $Y^{(4)}(x)$ for each $x \in D$ by using the distance proposed by Diamond [4] as a measure of the fit. This distance is used to fit the fuzzy nonparametric model (5).

Let $Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x)$ and $Y^{(4)}(x)$ have continuous derivatives in the domain $x \in D$. Then for a given $x_0 \in D$, The Taylor expansions of $Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x)$ and $Y^{(4)}(x)$ can be locally approximated in neighborhood of x_0, respectively by the following linear functions:

$$Y^{(k)}(x) \sim \hat{Y}^{(k)}(x) = Y^{(k)}(x_0) + Y^{'(k)}(x_0)(x-x_0).k$$
$$= 1, 2, 3, 4. \quad (12)$$

where $Y^{'(1)}(x_0).Y^{'(2)}(x_0).Y^{'(3)}(x_0)$ and $Y^{'(4)}(x_0)$ are respectively, the derivatives of $Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x)$ and

$Y^{(4)}(x)$ based on Diamond distance. The local linear smoothing method estimated at x_0 is:

$$F(x_0) = \left(Y^{(1)}(x_0).Y^{(2)}(x_0).Y^{(3)}(x_0).Y^{(4)}(x_0)\right) \quad (13)$$

by minimizing,

$$\sum_{i=1}^{n} d^2\left(\widetilde{Y}_i.\widehat{\widetilde{Y}}_i\right)$$
$$= \sum_{i=1}^{n} d^2\left(\left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}\right).\left(\widehat{Y}_i^{(1)}.\widehat{Y}_i^{(2)}.\widehat{Y}_i^{(3)}.\widehat{Y}_i^{(4)}\right)\right)$$
$$K_h(|x_i - x_0|) \quad (14)$$

with respect to $Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}$ and $\widehat{Y}_i^{(1)}.\widehat{Y}_i^{(2)}.\widehat{Y}_i^{(3)}.\widehat{Y}_i^{(4)}$ for the given kernel $k(\cdot)$ and the smoothing parameter h, where $K_h(|x_i - x_0|) = k\left(\frac{|x_i-x_0|}{h}\right)$ for $i = 1.2.\ldots.n$, are a sequence of weights at x_0. The two commonly use.

d kernel functions are parabolic shape functions:

$$k_1(x) = \begin{cases} 0.75(1-x^2); & if |x| \le 1. \\ 0; & Otherwise. \end{cases} \quad (15)$$

and Gaussian function:

$$k_2(x) = (2\pi)^{-\frac{1}{2}}\exp\left(\frac{-x^2}{2}\right). \quad (16)$$

Also, by substituting (12) into (14), the following can be obtained

$$\sum_{i=1}^{n} d^2\left(\widetilde{Y}_i.\widehat{\widetilde{Y}}_i\right)$$
$$= \sum_{i=1}^{n} d^2\left(\left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}\right).\left(\widehat{Y}_i^{(1)}.\widehat{Y}_i^{(2)}.\widehat{Y}_i^{(3)}.\widehat{Y}_i^{(4)}\right)\right)K_h$$
$$(|x_i - x_0|) \quad (17)$$

By solving this weighted least-squares problem, the following can be obtained for $Y^{(1)}(x), Y^{(2)}(x), Y^{(3)}(x), Y^{(4)}(x)$ and $Y^{'(1)}(x), Y^{'(2)}(x), Y^{'(3)}(x), Y^{'(4)}(x)$ at x_0. So the estimation $F(x)$ at x_0 is: $\widehat{\widetilde{Y}}(x_0) = (\widehat{\widetilde{Y}}^{(1)}(x_0).\widehat{\widetilde{Y}}^{(2)}(x_0).\widehat{\widetilde{Y}}^{(3)}(x_0).\widehat{Y}^{(4)}(x_0))$. Equation (17) has eight unknown parameters $Y^{(1)}(x).Y^{(2)}(x).Y^{(3)}(x).Y^{(4)}(x)$ and $Y^{'(1)}(x_0).Y^{'(2)}(x_0).Y^{'(3)}(x_0).Y^{'(4)}(x_0)$. In order to derive a formula for the unknown parameters of nonparametric regression based on minimizing this distance, the derivatives of Eq. (8) with respect to the eight unknown parameters need to be derived, set to zero and solve for the eight unknown parameters.

According to the principle of the weighted least-squares (and utilizing matrix notations), we can obtain

$$\left(\widehat{Y}^{(k)}(x).\widehat{Y}^{(k)}(x)\right)^T$$
$$= \left(X^T(x_0)W(x_0;h)X(x_0)\right)^{-1}X^T(x_0)W(x_0;h)\underset{Y}{\sim}^{(k)},$$
$$k = 1,2,3,4.$$
$$(18)$$

where $\quad X(x_0) = \begin{pmatrix} 1 & x_1 - x_0 \\ 1 & x_2 - x_0 \\ \vdots & \vdots \\ 1 & x_n - x_0 \end{pmatrix}. \quad \widetilde{Y}^{(k)} = \begin{pmatrix} Y_1^{(k)} \\ Y_2^{(k)} \\ \vdots \\ Y_n^{(k)} \end{pmatrix}.$

$k=1, 2, 3, 4$, and $\quad W(x_0;h) = \mathrm{Diag}(K_h(|x_1 - x_0|).K_h(|x_2 - x_0|). \dots .K_h(|x_n - x_0|)).$ is a $n \times n$ diagonal matrix with its diagonal elements being $K_h(|x_i - x_0|)$ for $i = 1.2. \dots .n$. If we suppose $e_1 = (1.0)^T$ and $H(x_0;h) = (X^T(x_0)W(x_0;h)X(x_0))^{-1}X^T(x_0)W(x_0;h)$, then the estimate of $F(x)$ at x_0 would be:

$$\widehat{\widetilde{Y}}(x_0) = \left(\hat{Y}^{(1)}(x_0) \cdot \hat{Y}^{(2)}(x_0) \cdot \hat{Y}^{(3)}(x_0) \cdot \hat{Y}^{(4)}(x_0)\right)$$
$$= \left(e_1^T H(x_0;h)\tilde{Y}^{(1)} \cdot e_1^T H(x_0;h)\tilde{Y}^{(2)} \cdot e_1^T H(x_0;h)\tilde{Y}^{(3)} \cdot e_1^T H(x_0;h)\tilde{Y}^{(4)}\right)$$
$$(19)$$

3.2 K-Nearest Neighbor Smoothing (K-NN)

The K-NN weight sequence was introduced by Loftsgaarden and Quesenberry [26] in the related field of density estimation and has been used by Cover and Hart [27] for classification purposes. The K-NN smother is defined as:

$$\underset{Y}{\sim}_i = \sum_{j=1}^n \omega_j(x)Y_j. \tag{20}$$

where $\omega_j(x)\, for\, j = 1.2. \dots .n$, is a the weight sequence at x and is defined as :

$$\omega_j(x) = \begin{cases} \frac{1}{k}; j \in J(x), j = 1.2. \dots .n. \\ 0; \qquad Otherwise. \end{cases} \tag{21}$$

where J_x is one of the K-nearest observations to x and $\widehat{\widetilde{Y}}_i$ is the estimate of the observations $\widetilde{Y}_i$ for $i = 1.2. \dots .n$ and $\widetilde{Y}_i = (Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)})$, asymmetric trapezoidal fuzzy numbers. Based on (20) and (21) we have

$$\widehat{\widetilde{Y}}_i = \left(\sum_{j=1}^n \omega_j Y_j^{(1)} \cdot \sum_{j=1}^n \omega_j Y_j^{(2)} \cdot \sum_{j=1}^n \omega_j Y_j^{(3)} \cdot \sum_{j=1}^n \omega_j Y_j^{(4)}\right)$$
$$= (\frac{1}{k}\sum_{j \in J(x)} Y_j^{(1)} \cdot \frac{1}{k}\sum_{j \in J(x)} Y_j^{(2)} \cdot \frac{1}{k}\sum_{j \in J(x)} Y_j^{(3)} \cdot \frac{1}{k}\sum_{j \in J(x)} Y_j^{(4)}).$$
$$(22)$$

The K-NN smoothing parameter is the neighborhood size k. So if a relatively small neighborhood size is used, this will increase the regression noise; on the other hand, a relatively large neighborhood size will increase the regression error. So Remark 1 describes leave-one-out cross-validation for finding k optimal value which can be obtained by minimizing cross-validation criterion.

3.3 Kernel Smoothing (K-S)

K-NN smoothing is a weighted averaging neighborhood in which weights in neighborhood are treated equally. In kernel smoothing $S(x)$ is defined by a fixed neighborhood around x. It is determined by kernel function and band of h. The fuzzy regression equations for kernel smoothing and K-NN smoothing are the same and represented by Eq. (20). In kernel smoothing method, $\omega_j(x_o)$ for $j = 1 \dots .n$, is defined as

$$\omega_j(x_o) = \frac{K_h(|x_j - x_o|)}{\sum_{i=1}^n K_h(|x_i - x_o|)} = \frac{K(\frac{|x_j - x_o|}{h})}{\sum_{i=1}^n K(\frac{|x_j - x_o|}{h})}. \tag{23}$$

$$\widehat{\underset{Y}{\sim}}_i = \left(\frac{\sum_{i=1}^n K_h(|x_j - x_o|Y_j^{(1)}}{\sum_{i=1}^n K_h(|x_i - x_o|)} \cdot \frac{\sum_{i=1}^n K_h(|x_j - x_o|Y_j^{(2)}}{\sum_{i=1}^n K_h(|x_i - x_o|)} \cdot \right.$$
$$\left. \frac{\sum_{i=1}^n K_h(|x_j - x_o|Y_j^{(3)}}{\sum_{i=1}^n K_h(|x_i - x_o|)} \cdot \frac{\sum_{i=1}^n K_h(|x_j - x_o|Y_j^{(4)}}{\sum_{i=1}^n K_h(|x_i - x_o|)}\right).$$
$$(24)$$

in which weight sequence is defined by $K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right)$, which is the kernel with scale factor. So the kernel smoothing parameter is band h and weight depends on smoothing parameter h.

3.4 Selection of Smoothing Parameters

The most important aspect for averaging techniques and local linear smoothing method is selecting the size of neighborhood to average k and parameter h. There are different methods for selecting parameter h such as the cross-validation method, and generalized cross validation which are used to obtain parameter h. Let $\widehat{\underset{Y}{\sim}}(x_i.h) = \left(\widehat{Y}^{(1)}(x_i.h).\widehat{Y}^{(2)}(x_i.h).\widehat{Y}^{(3)}(x_i.h).\widehat{Y}^{(4)}(x_i.h)\right).$

The fuzzified cross-validation procedure (CV) for selecting parameter h in local linear smoothing method based on Diamond distance is defined as:

$$CV(h) = \frac{1}{n}\sum_{i=1}^n d^2\left(\underset{Y}{\sim}_i . \widehat{\underset{Y}{\sim}}_i\right)$$
$$= \frac{1}{n}\sum_{i=1}^n ((Y_i^{(1)} - \widehat{\underset{Y}{\sim}}_i^{(1)})^2 + (Y_i^{(2)} - \widehat{\underset{Y}{\sim}}_i^{(2)})^2$$
$$+ (Y_i^{(3)} - \widehat{\underset{Y}{\sim}}_i^{(3)})^2 + (Y_i^{(4)} - \widehat{\underset{Y}{\sim}}_i^{(4)})^2). \tag{25}$$

where minimization gives the optimal value h, $CV(h_0) = \min_{h > o}CV(h)$.

Indeed, h can be obtained by minimization of $CV(h)$. The selected optimal value of h by the $CV(h)$ depends on the degree of smoothness of Y_{iL} and Y_{iR}. Large value of h leads to lack-of-fit and small value of h to an over-fit.

Also the cross validation leave-one-out technique is used for selecting values k and h in K-NN and KS methods that are obtained by minimization of the cross-validation criterion. According to Stone [28], the CV criterion is defined as

$$CV(b) = \frac{1}{n}\sum_{i=1}^{n} L[\underset{Y_i}{\sim} . \underset{Y_b}{\widehat{\sim}}(x_i.o_i)] = \frac{1}{n}\sum_{i=1}^{n}(D_i(b) + C_i(b)).$$
(26)

where $\underset{Y_b}{\widehat{\sim}}(x_i.o_i) = \sum_{j=1j \neq i}^{n} \omega_j(x_i)\widetilde{Y}_j$ and $D_i(b)$ is the measure of difference and $C_i(b)$ is the measure of inclusion as in:

$$C_i(b) = P\left(\left[\underset{Y_b}{\widehat{\sim}}(x_i.o_i)\right]_\alpha^L - \left[\underset{Y_i}{\sim}\right]_\alpha^L\right)$$
$$+ Q\left(\left[\underset{Y_b}{\widehat{\sim}}(x_i.o_i)\right]_\alpha^R - \left[\underset{Y_i}{\sim}\right]_\alpha^R\right).$$
(27)

where P and Q are penalty terms defined as:

$$P = \begin{cases} 1; & if \left[\widetilde{Y}_i\right]_\alpha^L \leq \left[\underset{Y_b}{\widehat{\sim}}(x_i.o_i)\right]_\alpha^L. \\ 0; & Otherwise. \end{cases} \quad \text{and}$$

$$Q = \begin{cases} 1; & if \left[\widetilde{Y}_i\right]_\alpha^R \leq \left[\underset{Y_b}{\widehat{\sim}}(x_i.o_i)\right]_\alpha^R. \\ 0; & Otherwise. \end{cases}$$

To obtain $D_i(b)$, we will use difference measure of a trapezoidal fuzzy number A, by using the method of Chang and Lee [5], which is defined as: $OM(A) = \int_0^1 \underset{\omega}{\sim}[\chi_1(v) A_L(v) + \chi_2(v)A_R(v)]dv$. Thus $D_i(b)$ is calculated using Eq. (26). For the calculation of parameter b, we minimize the CV criterion Eq. (25) such that $b^* = argmin\{CV(b)\}$. So b^* is the neighborhood size k in K-NN smoothing method and band h in kernel smoothing method.

4 Extension of the Proposed Method to Multivariate Input

It is straightforward to extent the proposed methods to the case of multivariate input. In fact, let $X = (x_1.x_2.\ldots.x_p)$ be a p-dimensional crisp input and Y be a trapezoidal fuzzy output. The fuzzy nonparametric regression model can be written as follows:

$$Y = Y^{(1)}(x_1.x_2.\ldots.x_p) + Y^{(2)}(x_1.x_2.\ldots.x_p)$$
$$+ Y^{(3)}(x_1.x_2.\ldots.x_p) + Y^{(4)}(x_1.x_2.\ldots.x_p)$$
$$= F(x)\{+\}\varepsilon.$$
(28)

4.1 K-Nearest Neighbor Smoothing (K-NN)

The K-NN smother is defined as:

$$\tilde{Y}_i = \sum_{j=1}^{n} \omega_j(x)Y_j.$$
(29)

where $\omega_j(x)$ for $j = 1.2.\ldots.n$, is a the weight sequence at x and is defined as :

$$\omega_j(x) = \begin{cases} \frac{1}{k}; & j \in J(x), j = 1.2.\ldots.n. \\ 0; & Otherwise. \end{cases}$$
(30)

where J_x is one of the K-nearest observations to x and $\widehat{\widetilde{Y}}_i$, the estimate of the observations $\widetilde{Y}_i$ for $i = 1.2.\ldots.n$ and $\widetilde{Y}_i = \left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)}\right)$ be asymmetric trapezoidal fuzzy numbers. Based on (29) and (30) we have

$$\widehat{\widetilde{Y}}_i = \left(\sum_{j=1}^{n} \omega_j Y_j^{(1)}.\sum_{j=1}^{n} \omega_j Y_j^{(2)}.\sum_{j=1}^{n} \omega_j Y_j^{(3)}.\sum_{j=1}^{n} \omega_j Y_j^{(4)}\right)$$
$$= \left(\frac{1}{k}\sum_{j \in J(x)} Y_j^{(1)}.\frac{1}{k}\sum_{j \in J(x)} Y_j^{(2)}.\frac{1}{k}\sum_{j \in J(x)} Y_j^{(3)}.\frac{1}{k}\sum_{j \in J(x)} Y_j^{(4)}\right)$$
(31)

The K-NN smoothing parameter is the neighborhood size k. $\widehat{Y}_i(l = 1.2.\ldots.p)$ are computed for $X = (x_1.x_2.\ldots.x_p)$; then $\widehat{\widetilde{Y}}_i = \frac{1}{n}\widehat{\widetilde{Y}}_{il}$.

4.2 Kernel Smoothing Method

This is the same as K-Nearest neighbor smoothing method in which $\omega_j(X_o)$ for $j = 1.2.\ldots.n$, at X_0 are defined as:

$$\omega_j(x_0) = \frac{K_h(\|x_j - x_0\|)}{\sum_{i=1}^{n} K_h(\|x_i - x_0\|)} = \frac{K(\frac{\|x_j - x_0\|}{h})}{\sum_{i=1}^{n} K(\frac{\|x_i - x_0\|}{h})}.j$$
$$= 1.2.\ldots.n.$$
(32)

4.3 Local Linear Smoothing

Suppose that $Y^{(1)}(X).Y^{(2)}(X).Y^{(3)}(X)$ and $Y^{(4)}(X)$ have continuous derivatives in the domain $x \in D$. Then for a given $x_0 \in D$ and Taylor's expansion, $Y^{(1)}(X).Y^{(2)}(X).Y^{(3)}(X)$ and $Y^{(4)}(X)$ can be locally approximated in neighborhood of x_0, respectively by the following linear functions:

$$Y^{(k)}(x) \simeq \hat{Y}^{(k)}(x)$$
$$= Y^{(k)}(x_0) + Y^{(k)(x_1)}(x_0)(x_1 - x_{01}) + \cdots$$
$$+ Y^{(1k)}(x_p - x_{0p}).k$$
$$= 1.2.3.4. \tag{33}$$

where $Y^{(1)(x_j)}(x_0).Y^{(2)(x_j)}(x_0).Y^{(3)(x_j)}(x_0)$ and $Y^{(4)(x_j)}(x_0)$ are, respectively, the derivatives of $Y^{(1)}(X).Y^{(2)}(X).Y^{(3)}(X)$ and $Y^{(4)}(X)$ with respect to (x_j) based on Diamond distance and the local linear smoothing method is estimated at x_0,

$$F(x_0) = \left(Y^{(1)}(x_0).Y^{(2)}(x_0).Y^{(3)}(x_0).Y^{(4)}(x_0) \right). \tag{34}$$

by minimizing:

$$\sum_{i=1}^{n} d^2 \left(\underset{\sim}{Y_i} \cdot \widehat{\underset{\sim}{Y_i}} \right)$$
$$= \sum_{i=1}^{n} d^2 \left(\left(Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.Y_i^{(4)} \right) . \left(\hat{Y}_i^{(1)}.\hat{Y}_i^{(2)}.\hat{Y}_i^{(3)}.\hat{Y}_i^{(4)} \right) \right)$$
$$K_h(\|X_i - X_0\|). \tag{35}$$

with respect to $Y_i^{(1)}.Y_i^{(2)}.Y_i^{(3)}.,Y_i^{(4)}$ and $\hat{Y}_i^{(1)}.\hat{Y}_i^{(2)}.\hat{Y}_i^{(3)}.\hat{Y}_i^{(4)}$ for the given kernel $k(.)$ and smoothing parameter h, where $K_h(\|X_i - X_0\|) = k\left(\frac{\|X_i - X_0\|}{h} \right)$ for $i = 1.2.....n$ are a sequence of weights at X_0. Also, by substituting (33) into (35), the following can be obtained :

$$\sum_{i=1}^{n} d^2 \left(\tilde{Y}_i \cdot \widehat{\tilde{Y}}_i \right) = \sum_{i=1}^{n} d^2 \left(\left(Y_i^{(1)} \cdot Y_i^{(2)} \cdot Y_i^{(3)} \cdot Y_i^{(4)} \right) \right.$$
$$\left. \cdot \left(\hat{Y}_i^{(1)} \cdot \hat{Y}_i^{(2)} \cdot \hat{Y}_i^{(3)} \cdot \hat{Y}_i^{(4)} \right) \right) K_h(\|X_i - X_0\|)$$
$$= \sum_{(k=1)}^{4} \sum_{i=1}^{n} \left(Y_i^{(k)} - Y^{(k)}(x_0) - \sum_{j=1}^{p} Y^{(k)(x)_j}(X_0)(x_{ij} - x_{0j}) \right)^2 K_h(\|X_i - X_0\|), \tag{36}$$

where $\|X_i - X_0\|$ is Euclidean distance between X_i and X_0 and

$$\widehat{\tilde{Y}}(x) = \left(\hat{Y}^{(1)}(x_0) \cdot \hat{Y}^{(2)}(x_0) \cdot \hat{Y}^{(3)}(x_0) \cdot \hat{Y}^{(4)}(x_0) \right)$$
$$= \left(e_1^T H(X_0;h)\tilde{Y}^{(1)} \cdot e_1^T H(X_0;h)\tilde{Y}^{(2)} \cdot e_1^T H(X_0;h)\tilde{Y}^{(3)} \cdot e_1^T H(X_0;h)\tilde{Y}^{(4)} \right), \tag{37}$$

where $X(x_0) = \begin{pmatrix} 1 & x_{11} - x_{01}...x_{1p} - x_{0p} \\ 1 & x_{21} - x_{01}...x_{2p} - x_{0p} \\ \vdots & \vdots \\ 1 & x_{n1} - x_{01}...x_{np} - x_{0p} \end{pmatrix}$. $\tilde{Y}^{(k)} = \begin{pmatrix} Y_1^{(k)} \\ Y_2^{(k)} \\ \vdots \\ Y_n^{(k)} \end{pmatrix}$. k=1, 2, 3, 4, and $W(X_0;h) = $

$\text{Diag}(K_h(\|X_1 - X_0\|).K_h(\|X_2 - X_0\|).....K_h(\|X_p - X_0\|))$
is a $n \times n$ diagonal matrix with its diagonal elements being $K_h(\|x_i - x_o\|)$ for $i = 1.2.....n$. If we suppose $e_1 = (1.0)^T$ and

$$H(x_0;h) = (X^T(X_0)W(X_0;h)X(X_0))^{-1}X^T(X_0)W(X_0;h).$$

then the estimate of $F(x)$ at x_0 is

$$\widehat{\hat{Y}}(x) = \left(\hat{Y}^{(1)}(x_0) \cdot \hat{Y}^{(2)}(x_0) \cdot \hat{Y}^{(3)}(x_0) \cdot \hat{Y}^{(4)}(x_0) \right)$$
$$= \left(e_1^T H(x_0;h)\tilde{Y}^{(1)} \cdot e_1^T H(x_0;h)\tilde{Y}^{(2)} \cdot e_1^T H(x_0;h)\tilde{Y}^{(3)} \right.$$
$$\left. \cdot e_1^T H(x_0;h)\tilde{Y}^{(4)} \right). \tag{38}$$

5 Adaptive Neural Fuzzy Inference System (ANFIS)

ANFIS borrows its name from fuzzy deduction system which are based on open adaptive networks or adaptive neural fuzzy detection systems. It is a kind of artificial neural network based on Takagi–Sugeno fuzzy inference system. The network is a famous hybrid method, in which combinations of nodes $o_{k,i}$, are placed as layers for performing specific functions (Fig. 1).

Referring to Fig. 2, the first and second layers are called, the input and fuzzification layers. The third layer expresses the rules and number of Sugeno fuzzy logic deduction system. The output of each rule node $\mu_{\underset{\sim}{A_i}}(x_j)$ is the multiplication of membership degrees coming from the second layer. The outputs of layer three, then drives the layer forth whose task is calculation of the normalized ignition level for each rule. The fifth layer is called the debugging layer, which produces the weighted outputs from the previous layers. The output of rules is calculated for each node and the last layer takes all the outputs then gives an algebraic summation. Table 1 presents layers of ANFIS.
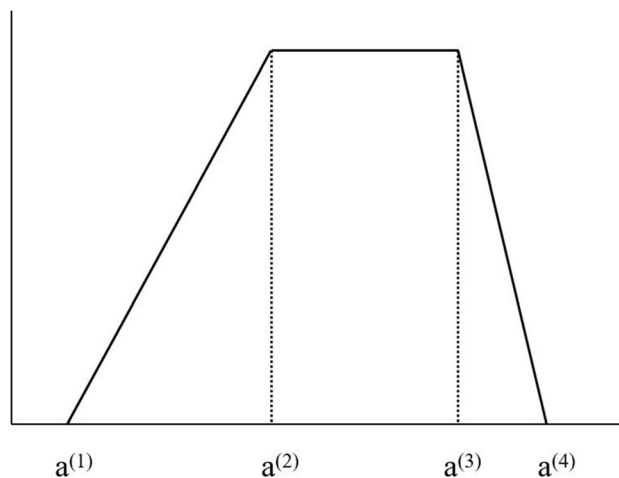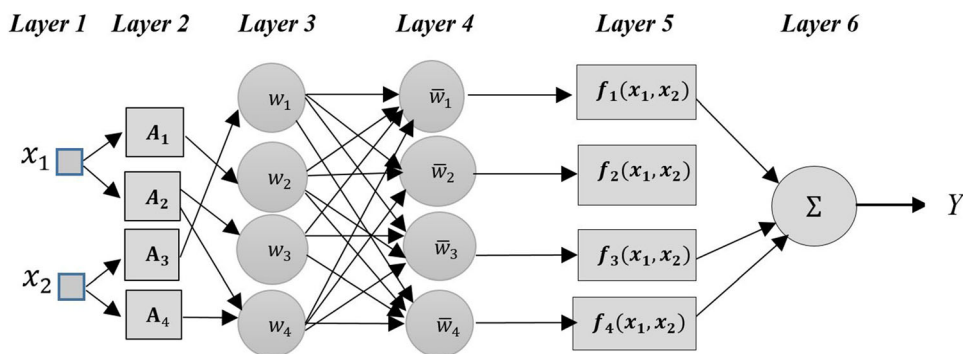

Fig. 1 A trapezoidal fuzzy number

Fig. 2 Configuration of an ANFIS

Table 1 The layers of ANFIS

First layer	$x_j = (x_{j1}, x_{j2}, \ldots, x_{jp})$, $j = 1, 2, \ldots, n$ and Rules are
	If x_1 is fuzzy set $\tilde{A}_1$ and x_2 is fuzzy set $\tilde{A}_3$ then $f_1 = q_0^1 + q_1^1 x_1 + q_2^1 x_2$;
	If x_1 is fuzzy set $\tilde{A}_1$ and x_2 is fuzzy set $\tilde{A}_4$ then $f_2 = q_0^2 + q_1^2 x_1 + q_2^2 x_2$;
	If x_1 is fuzzy set $\tilde{A}_2$ and x_2 is fuzzy set $\tilde{A}_3$ then.$f_3 = q_0^3 + q_1^3 x_1 + q_2^3 x_2$;
	If x_1 is fuzzy set $\tilde{A}_2$ and x_2 is fuzzy set $\tilde{A}_4$ then.$f_4 = q_0^4 + q_1^4 x_1 + q_2^4 x_2$;
Second layer	$o_{2,i} = \mu_{\tilde{A}_i}(x_1) = \exp\left[-\left(\frac{x_1 - \tau_i}{2\sigma_i}\right)^2\right]$, $i = 1, 2$ and $o_{2,i} = \mu_{\tilde{A}_i}(x_2) = \exp\left[-\left(\frac{x_2 - \tau_i}{2\sigma_i}\right)^2\right]$,
Third layer	$i = 3, 4; o_{3,i} = w_i = \mu_{\tilde{A}_i}(x_1) . \mu_{\tilde{A}_i}(x_2) i = 1, 2, 3, 4$;
Fourth layer	$o_{4,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + w_4} i = 1, 2, 3, 4$;
Fifth layer	$o_{5,i} = \bar{w}_i f_i = \bar{w}_i (q_0^i + q_1^i x_1 + q_2^i x_2) i = 1, 2, 3, 4$;
Sixth layer	$\widehat{\tilde{Y}}_j = o_{6,i} = \sum_{i=1}^{4} \bar{w}_i f_i = \frac{w_1}{w_1 + w_2 + w_3 + w_4} f_1 + \frac{w_2}{w_1 + w_2 + w_3 + w_4} f_2 + \frac{w_3}{w_1 + w_2 + w_3 + w_4} f_3 + \frac{w_4}{w_1 + w_2 + w_3 + w_4} f_4$
	$= \bar{w}_1 f_1 + \bar{w}_2 f_2 + \bar{w}_3 f_3 + \bar{w}_4 f_4$

5.1 The Proposed Hybrid Algorithm

The proposed hybrid algorithm benefits from both post-propagation and propagation techniques. First, the fuzzy least squares method is used to optimize the secondary parameters. In the second stage, when secondary parameters are obtained, the initial parameters are optimized using the slope slider algorithm. Here, we have used the MATLAB software package to implement the algorithms. Figure 3 shows the proposed hybrid algorithm.

Report network with the lowest error in the data output of the test along with the corresponding results.

6 Examples

In this section, a simulation examples and a numerical one are presented, in which the input is a crisp number and the output, a trapezoidal fuzzy number. We estimate the values of the parameters using ANFIS methods; then these results are compared with each other based on *CV* and the charts.

Example 1 The grinding is a process in which using an abrasive material, along with a grinding wheel. The aim is providing surfaces with precise tolerances, create optimal surface, creating accurate surface form and machining of hard and brittle materials (see Fig. 4).

The trapezoidal fuzzy numbers have been constructed using Table 3 in Ref. [29]. The $\widetilde{Y}_i$ is the roughness of a workpiece surface and input x is the feed speed of a grinding wheel. The output $\widetilde{Y}_i$ is measured by symmetric trapezoidal fuzzy numbers as $\widetilde{Y}_i = \left(Y_i^{(1)} . Y_i^{(2)} . Y_i^{(3)} . Y_i^{(4)}\right)$. The estimation is performed using ANFIS method. Thus, regarding this dataset, the nonparametric regression model is an appropriate one. The proposed method is used to fit the regression model while as before, *CV* is computed for
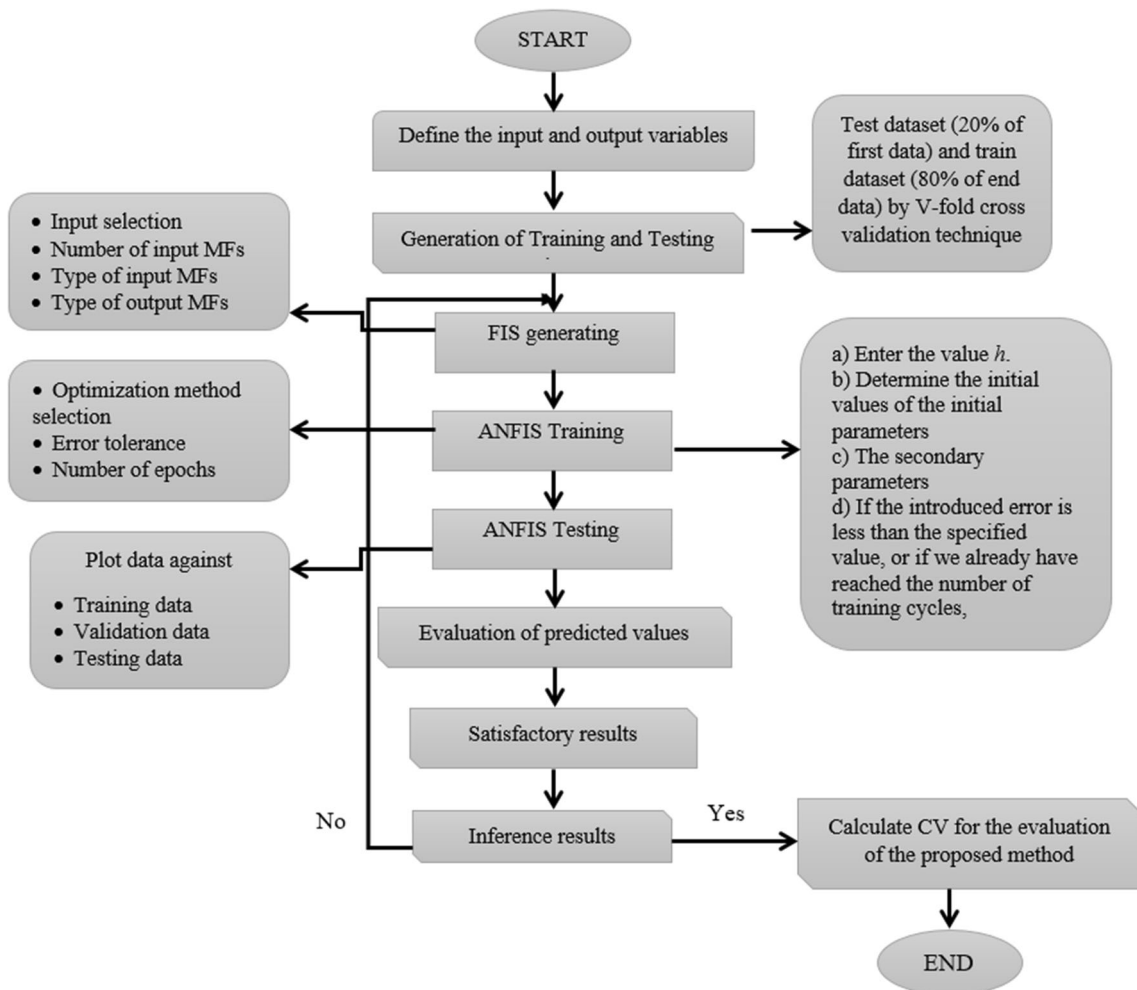
Fig. 3 The proposed hybrid algorithm

validation purposes. In the proposed method, we use 5-layer validation technique. The results of the parameters estimation and those of the corresponding errors are listed in Tables 2 and 3, respectively. In this example, the network was convergent and, as a justification, the error value reaches a constant value with acceptable level after the training. The observed values and the predicted values of proposed method is shown in Fig. 5. The center lines, the upper and lower boundary lines of the real function and their estimates are plotted in the figure. The observed and estimated values of different smoothing methods (K- S, L-L-S and K-NN) are shown in Fig. 6, 7 and 8, respectively. In these figures, the center lines, the upper and lower boundary lines of the real function and their estimates are plotted. Moreover, the low *CV* value, as shown in Table 2, indicates that the proposed method tends to produce predicted values close to the observed values and in turn, has a

remarkable accuracy in prediction. For different fitting methods, the different smoothing methods are applied such as K-NN, K- S and L-L-S technique. The value of parameter smoothing method in K-NN and K- S method is determined by cross validation method. Table 3 shows the values of the primary and secondary parameters of proposed method.

Example 2 In this example the function is considered to be $f(x) = \frac{x^2}{5} + 2e^{\frac{x}{10}}$. Suppose that x_is are uniformly random variables, within the interval [0,1] and $i = 1, 2, \ldots, 100$,

$$\underset{\sim}{Y_i} = \left(Y_i^{(1)} . Y_i^{(2)} . Y_i^{(3)} . Y_i^{(4)} \right)$$

$$= (y_i - e_i . y_i + \frac{1}{3} e_i . y_i + \frac{2}{3} e_i . y_i + e_i) \qquad (39)$$

So that $y_i = f(X_i) + rand[-0.5.0.5]$ and $e_i = 1/4 f(X_i) + rand[0.1]$.
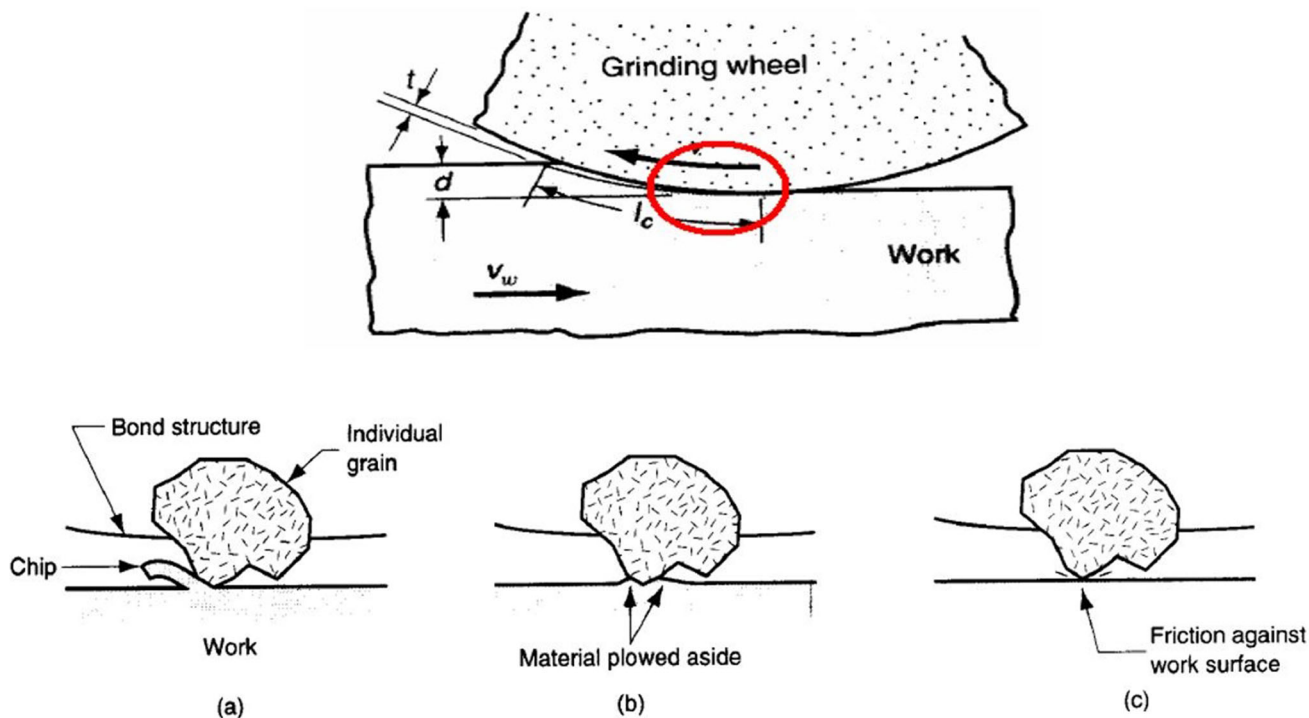
Fig. 4 Grinding method

Table 2 Achieved results of proposed method (ANFIS)

CV(Proposed method)	0.0055
CV (K-S)	0.0058
CV (L-L-S)	0.0057
CV (K-NN)	0.0059
Membership function type for ANFIS	Gaussian
Number of training data pairs for ANFIS	15
Number of testing data pairs for ANFIS	4
Number of fuzzy rules for ANFIS	2

In this example we estimate the values by using the proposed method (ANFIS). So in this dataset, the non-parametric regression model is a convenient model. The proposed method is used to fit the regression model and, like the example above, we compute the amount of CV to
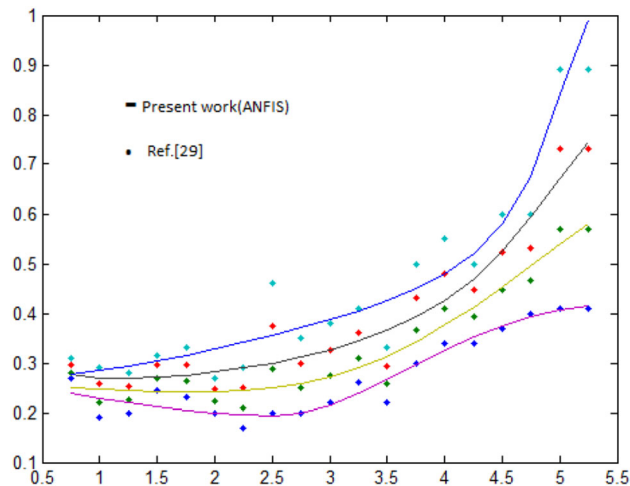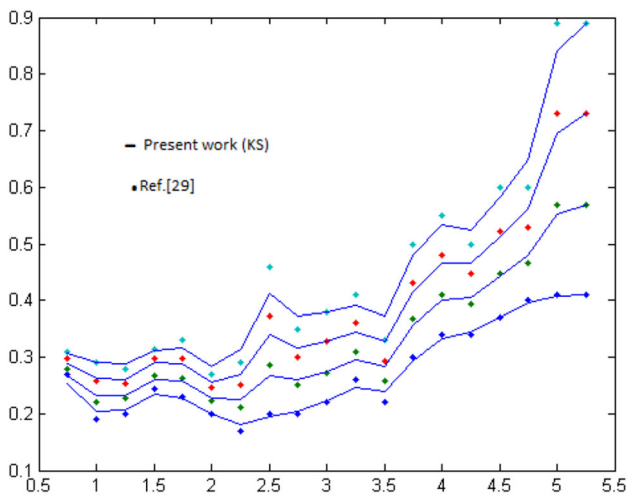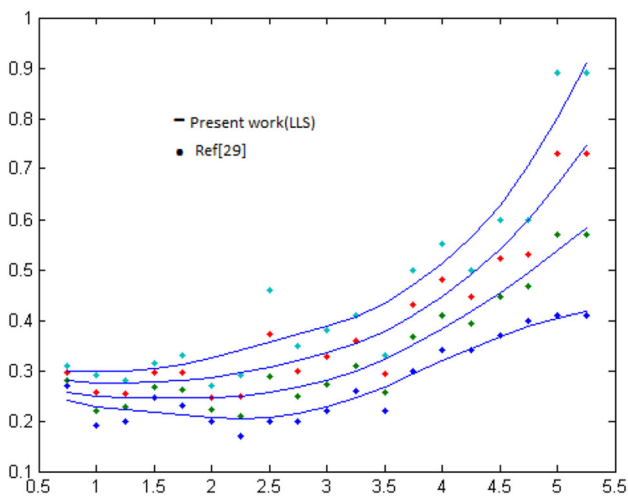

Fig. 5 Achieved results by ANFIS

**Table 3** The values of the primary and secondary parameters obtained in the proposed method when $k = 5$

i	$(\tau_1.\sigma_1)$	$(\tau_2.\sigma_2)$	$(b_0^i.\alpha_0^i)$	$(b_1^i.\alpha_1^i)$
1	(2.64, 1.10)			
	(2.00, 5.21)	(− 0.01, 0.29)	(0.16, − 0.20)	
2	(0.988, 0.666)	(0.929, 5.43)	(− 0.005, 0.208)	(0.073, 0.035)
3	(1.50, 2.08)	(2.25, 6.48)	(− 0.02, 0.27)	(0.06, 0.28)
4	(2.14, 1.26)	(1.49, 6.102)	(− 0.06, 0.30)	(− 0.12, 1.14)
5	(2.33, 1.42)	(1.48, 6.38)	(− 0.032, 0.262)	(− 0.116, 1.137)

Fig. 6 Achieved results by K-S method with Gausian kernel for h = 0.65

Fig. 7 Achieved results by L-L-S method with Gausian kernel for h = 0.69
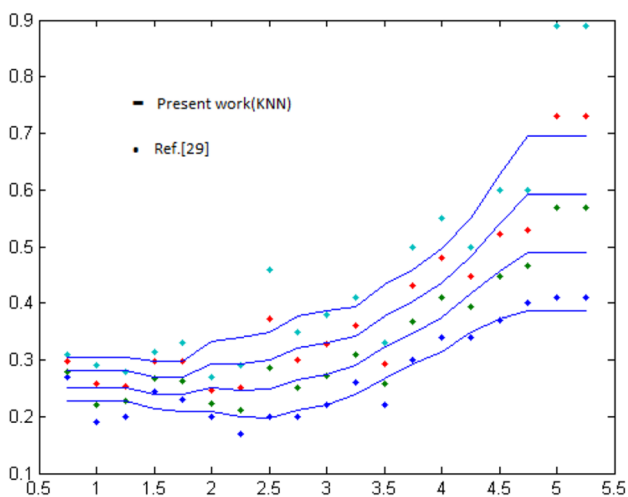
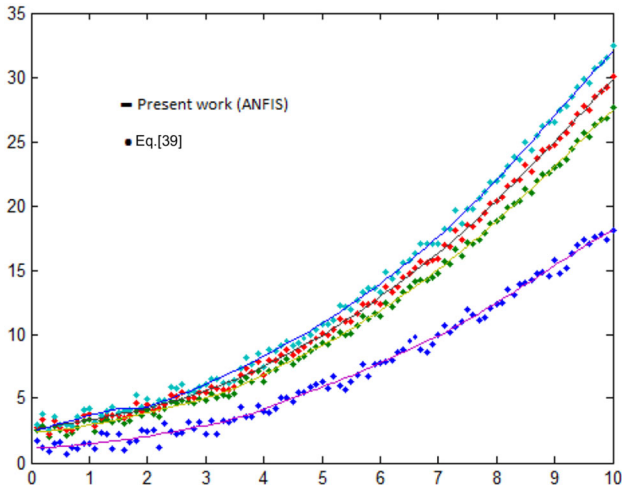**Fig. 8** Achieved results by KNN method with K = 5

evaluate the model. In the proposed method, we use 5-layer Validation technique. The low *CV* value, as shown in Table 4, indicates that the proposed method tends to produce predicted values close to the observed values and in turn, has a remarkable accuracy in prediction. The behavior of the proposed method is shown in Fig. 9, 10, 11 and 12 (like the above example). The values of the parameters obtained in the proposed algorithm are shown in Table 5, respectively. In this example, this network is convergent and, as can be seen, after the training, the error value reaches a constant value. Linear smoothing method, K-NN and kernel smoothing are applied to the fitting model. Gauss and Parabolic shape kernels are used to produce the weight sequence for local linear smoothing and kernel smoothing methods.

7 Conclusions

The method of adaptive fuzzy inference system for fitting non-parametric fuzzy regression model with crisp inputs and fuzzy outputs was investigated. Then two hybrid fuzzy regression algorithms have been proposed. Within the algorithm, a reduction slope method was used to optimize the initial parameters (fuzzy weights). The secondary parameters have been estimated by least squares Diamond method. It has been considered non-fuzzy inputs and trapezoidal fuzzy outputs. Simulation and numerical examples have been used to the evaluate performance of method. Using the results, the conducted simulation experiments are shown that the performance of the proposed method is better than that of the smoothing methods, which reduces the CV. In the proposed approach, when the observations numbers are increased, the accuracy is increased, in comparison with the existing smoothing methods. These advantages would make our algorithm an acceptable one to generate nonparametric regression functions. Thus current proposed method reduced the fuzziness of the system and it has faster adaptation.

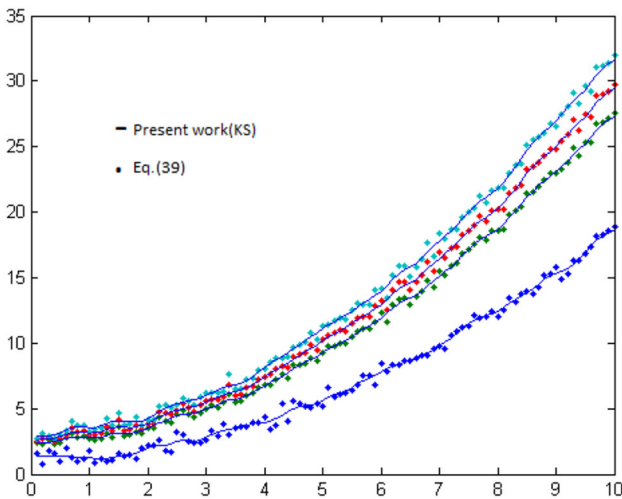**Table 4** Achieved results of proposed method (ANFIS)

|				
CV Proposed method (ANFIS)	0.05197	Membership function type for ANFIS		Gaussian
CV (K-S)	0.4959	Number of training data pairs for ANFIS		15
CV (L-L-S)	0.3378	Number of testing data pairs for ANFIS		4
CV (K-NN)	0.442	Number of fuzzy rules for ANFIS		2

Fig. 9 Achieved results by ANFIS method
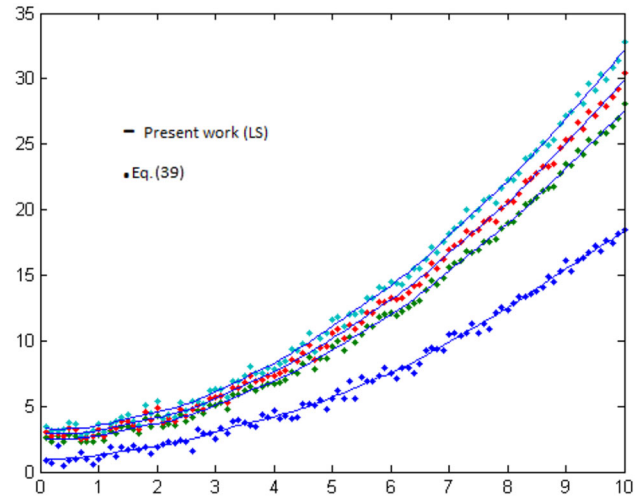
Fig. 11 Achieved results by L-L-S method with Gausian kernel for $h = 0.43$

Fig. 10 Achieved results by K-S method with Gausian kernel for $h = 0.12$
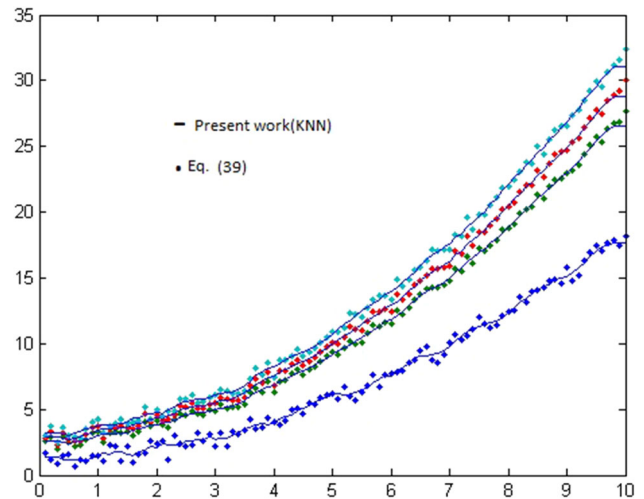
Fig. 12 Achieved results by KNN method with $k = 19$

Table 5 The values of the primary and secondary parameters obtained using the proposed method when k = 5

i	$(\tau_1.\sigma_1)$	$(\tau_2.\sigma_2)$	$(\tau_3.\sigma_3)$	$(b_0^i.\alpha_0^i)$	$(b_1^i.\alpha_1^i)$	$(b_2^i.\alpha_2^i)$
1	(1.59, 2.18)	(1.69, 3.24)	(0.41, 8.82)	(0.44, 6.12)	(3.23, − 6.8)	(4.24, − 15)
2	(0.847, 2.32)	(3.24, 6.74)	(3.31, 9.35)	(0.90, 1.94)	(0.53, 0.0)	(2.91, 18.8)
3	(0.46, 2.28)	(1.93, 8.73)	(1.90, 9.37)	(1.03, 1.82)	(− 1.1, 0.14)	(4.05, 19.8)
4	(2.28, 1.91)	(2.47, 3.89)	(0.68, 7.69)	(0.82, 9.2)	(4.38, − 13.35)	(4.28, − 14.0
5	(2.21, 1.38)	(1.61, 3.04)	(0.51, 8.18)	(2.95, 4.18)	(2.57, − 7.58)	(4.6, − 16.55)

References

1. Zadeh, A.: Fuzzy sets. Inform Control. **8**, 338–353 (1965)
2. Tanaka, H., Uejima, S., Asia, K.: Linear regression analysis with fuzzy model. IEEE Trans. Syst. Man Cybernet. **12**, 903–907 (1982)
3. Ishibuchi, H., Tanaka, H.: Fuzzy regression analysis using neural networks. Fuzzy Sets Syst. **50**, 257–265 (1992)
4. Diamond, P.: Fuzzy least squares. Inf. Sci. **46**, 141–157 (1988)
5. Chang, P.T., Lee, E.S.: A generalized fuzzy weighted least-squares regression. Fuzzy Sets Syst. **82**, 289–298 (1996)
6. Yang, M.S., Lin, T.S.: Fuzzy least-squares linear regression analysis for fuzzy input-output data. Fuzzy Sets Syst. **126**, 389–399 (2002)
7. Kartalopous, S.: Underestanding neural networks and fuzzy logic. IEEE Press, New York (1996)
8. Tanaka, H., Hayashi, I., Watada, J.: Possibilistic linear regression analysis for fuzzy data. Eur. J. Oper. Res. **40**, 389–396 (1989)
9. Cheng, C.B., Lee, E.S.: Fuzzy regression with radial basis function networks. Fuzzy Sets Syst. **119**, 291–301 (2001)
10. Arnold, S.: The merging of neural networks, fuzzy logic, and genetic algorithms. Insur. Math. Econ. **31**(1), 115–131 (2002)
11. Mosleh, M., Otadi, M., Abbasbandy, S.: Evaluation of fuzzy regression model by fuzzy neural networks. J. Comput. Appl. Math. **234**, 825–834 (2010)
12. Danesh, S., Farnoosh, R., Razzaghnia, T.: Fuzzy nonparametric regression based on an adaptive neuro-fuzzy inference system. Neuro comput. **173**, 1450–1460 (2016)
13. Sun, J., Lu, Q.: Regression analysis of a kind of trapezoidal fuzzy numbers based on a shape preserving operator. Data Anal. Inf. Process. **5**, 96–114 (2017)
14. Cheng, C.B., Lee, E.S.: Nonparametric fuzzy regression k-NN and kernel smoothing techniques. Comput. Math. Appl. **38**, 239–251 (1999)
15. Razzaghnia, T., Danesh, S., Maleki, A.: Hybrid fuzzy regression with trapezoidal fuzzy data. In: Proceedings of the SPIE. 8349, 834921–1–6 (2011).
16. Razzaghnia, T., Danesh, S.: Nonparametric regression with trapezoidal fuzzy data. Int. J. Recent Innov. Trends Comput. Commun. (IJRITCC) **3**, 3826–3831 (2015)
17. Razzaghnia, T.: Regression parameters prediction in data set with outliers using neural network. Hacettepe J. Math. Stat. **48**(4), 1170–1184 (2019)
18. Škrjanc, I., Antonio Iglesias, J., Sanchis, A., Leite, D., Lughofer, E., Gomide, F.: Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: a survey. Inf. Sci. **490**, 344–369 (2019)
19. Junhong, L., Zeng, W.X., Yin, Q.: A new fuzzy regression model based on least absolute deviation. Eng. Appl. Artif. Intell. **52**, 54–64 (2016)
20. Deng, W., Zhao, A.: A novel collaborative optimization algorithm in solving complex optimization problems. Soft Comput. **21**(15), 4387–4398 (2017)
21. Khosravia, K., Shahabib, H.: A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J. Hydrol. **573**, 311–323 (2019)
22. Liu, T., Zhang, W., Mc Lean, P.: Electronic nose-based odor classification using genetic algorithms and fuzzy support vector machines. Int. J. Fuzzy Syst. **20**, 1309–1320 (2018)
23. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Application. Academic Press, New York (1980)
24. Ishibuchi, H., Nii, M.: Fuzzy regression using asymmetric fuzzy coefficients and fuzzified neural networks. Fuzzy Sets Syst. **119**, 273–290 (2001)
25. Razzaghnia, T., Pasha, E., Khorram, E., Razzaghnia, A.: Fuzzy linear regression analysis with trapezoidal coefficients. First Joint Congress On Fuzzy and Intelligent Systems. Aug. 29–31 (2007).
26. Loftsgaarden, D.O., Quesenberry, G.P.: A nonparametric estimate of a multivariate density function. Ann. Math. Stat. **36**, 1049–1051 (1965)
27. Hart, J.D.: Nonparametric Smoothing and Lack-of-Fit Tests. Springer, New York (1997)
28. Stone, M.: Cross validation choice and assessment of statistical predictions. J. Roy. Stat. Soc. **36**, 111–147 (1974)
29. Lee, H., Tanaka, H.: Fuzzy approximations with non-symmetric fuzzy parameters in fuzzy regression analysis. J. Oper. Res. Soc. Jpn. **42**, 98–112 (1999)

R. Naderkhani was born in 1977 in Iran. She received the B.S degree in statistics from Shahid Bahonar University of Iran in 2000, and M.S degree in pure statistics from Shahid Beheshti University of Iran in 2003, and is studying Ph.D degree in applied statistics in Azad University of Iran now, respectively. She is currently faculty member

in department of Mathematics and Statistics, Aliabad katul Branch, Islamic Azad university, Golestan, Iran. She is researching in fuzzy regression and neural networking.

M. H. Behzadi was born in 1979 in Iran. He received the B.S degree in statistics from Razi University of Iran in 2002, and M.S degree in pure statistics from Amirkabir University of Technology of Iran in 2004, Ph.D degree in applied statistics from Azad University of Iran in 2009, respectively. He is currently an Associate Professor in department of statistics, Science and Research Branch, Islamic Azad university, Tehran, Iran. He is researching in fuzzy regression, Random optimization, Statistical inference. He has published some paper in this erea.

T. Razzaghnia was born in 1978 in Iran. She received the B.S degree in statistics from Shahid Bahonar University of Iran in 2000, and M.S degree in pure statistics from Kharazmi University of Iran in 2002,

and Ph.D degree in applied statistics from Azad University of Iran in 2007, respectively. She is currently an assistant Professor in department of statistics, Roudehen Branch, Islamic Azad university, Roudehen, Iran. She is researching in fuzzy regression, Nonparametrics, entropy and neural networking. She has published some paper in this erea.

R. Farnoosh was born in 1960 in Iran. He received the B.S degree in applied mathematics from Sharif University of Technology ,Iran in 1986, and M.S degree in Mathematical statistics from Shiraz University Iran in 1988and Ph.D degree in statistics from Leeds University of England in 2001, respectively. He is currently a Professor in department of mathematics, Iran university of science and technology, Tehran, Iran. He is researching in fuzzy regression, Advanced simulation, Image processing and Time series modeling. He has published numerous paper in this erea.