ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

COMPARISON OF FUZZY REGRESSION MODELS BASED ON LAD USING THE WEAKEST Tw-NORM BASED OPERATIONS

B. Pushpa[1], D. Iranian[2], A. Thiruppathi[3]
Department of Mathematics, Panimalar Institute of Technology, Chennai, India

DOI: 10.5281/zenodo.1252882

ABSTRACT
Fuzzy regression models have been applied to operational research applications such as forecasting. The weakest t-norm (T_w) based fuzzy regression model is developed using the concept of least absolute deviation defined between the observed and the estimated dependent values. This study compares the efficiency of each estimated with the different measures based on LAD. The fuzzy regression model using LAD based on Hausdorff distance is evaluated with a case study of investigating the relationship between Diesel Fuel properties and emissions from engines.
KEYWORDS: Weakest t-norm, Fuzzy regression, Fuzzy input- fuzzy output, Similarity

I. INTRODUCTION

Fuzzy set theory provides a significant alternative to the probabilistic approach to find the various arithmetic operations for evaluating the performance of the system. In recent times, the use of fuzzy sets has been gaining popularity and is playing an important role in the areas of engineering and management disciplines. As compared to other research domains, the fuzzy arithmetic gained great interest in scientific areas such as decision problem, reliability analysis, optimization etc. Most of the data collected from various resources are generally imprecise, vague and uncertain. Therefore to handle these types of data, fuzzy set theory has been used and then analyzed. Recently, for example, fuzzy regression models have been applied to insurance [1], housing [2], thermal comfort forecasting [3], productively and consumer satisfaction [4], product life cycle prediction [5], reservoir operations [6], actuarial analysis [7] and business cycle analysis [8]. Tanaka et al. [9] initially applied their fuzzy linear regression procedure to non-fuzzy experimental data. They formulated the fuzzy linear regression problem as a linear programming model to determine the regression coefficients as fuzzy numbers, where the objective is to minimize the total spread of the fuzzy regression coefficients subject to the constraint that the support of the estimated values is needed to cover the support of their associated observed values for a certain pre-specified level.

The main drawback of Tanaka's approach is the scale dependent. Although this approach was later improved by Tanaka and Watada [10], Tanaka et al. [11] still suffered the problem of being extremely sensitive to outliers as pointed out by Redden and Woodall [12]. The main purpose of fuzzy regression models is to find the best model with the least error. Based on this, the methods are classified as follows:

(i) **Possibilistic approach**, which tries to minimize the fuzziness of the model by minimizing the total spread of its fuzzy coefficients, subject to include the data points of each sample within a specified feasible data interval.

(ii) **Least square approach**, which minimizes the total spread of errors in the estimated value, based on their specification. This approach is an extension of ordinary least squares which obtains the best fitting to the data, based on the distance measure under fuzzy consideration.

Regression analysis based on the method of least -absolute deviation has been used as a robust method. When outlier exists in the response variable, the least absolute deviation is more robust than the least square deviations estimators. Some recent works on this topic are as follows: Chang and Lee [13] studied the fuzzy least absolute deviation regression based on the ranking method for fuzzy numbers. Kim et al. [14] proposed a two stage method to construct the fuzzy linear regression models, using a least absolutes deviations method. Torabi and Behboodian [15] investigated the usage of ordinary least absolute deviation method to estimate the fuzzy coefficients in a linear regression model with fuzzy input – fuzzy output observations. Considering a certain fuzzy regression model, Chen and Hsueh [16] developed a mathematical programming method to determine the crisp coefficients. Fuzzy regression model, based on L1 norm (absolute norm) criteria. Choi and

Buckley [17] suggested two methods to obtain the least absolute deviation estimators for common fuzzy linear regression models using TM based arithmetic operations. Taheri and Kelkinnama [18] introduced some least absolute regression models, based on crisp input- fuzzy output and fuzzy input-fuzzy output data respectively.

There are models [11, 15, 19] in fuzzy regression in which two objectives are considered: minimizing the width of the regression coefficients and bringing the h-cut of the predicted values as close as possible to the h-cut of the observed values. In the proposed method first objective is to minimize the distance between the observed and estimated using different LAD and the second objective is to minimize the dissimilarity between these two of fuzzy numbers using GMIR.

In a regression model, multiplication of fuzzy numbers are done by arithmetic operations such as α-levels of multiplication of fuzzy numbers and the approximate formula for multiplication of fuzzy numbers. The α-cut arithmetic cannot effectively preserve the original shape of a membership function, but by using the weakest T – norm (T_w), the shape of fuzzy numbers in multiplication will be preserved. The T_w- norm based operations reduce the width of the estimated responses which will give exact prediction. In this regard, Hong et al. [20] presented a method to evaluate fuzzy linear regression models based on a possibilistic approach, using T_w - norm based arithmetic operations.

In this paper, section II focuses on some important preliminary definitions and basics on fuzzy arithmetic operations based on the weakest T-norm and the different LAD distances between the responses are discussed using the shape preserving operations on fuzzy numbers and in section III, different LAD using The weakest t- norm operations are investigated with fuzzy input/ fuzzy output and in section IV, the Least absolute deviation using Hausdorff distance is studied with all possible data types, also a case study with crisp multiple inputs investigating the relation between the fuel properties and emission of engines.

II. THE WEAKEST T-NORM ARITHMETIC OPERATIONS

Since this study concerns fuzzy arithmetic based on the weakest norm, this section will briefly discusses T_w norm (the weakest t-norm) based arithmetic operations. In Zadeh's extension principle [21], by using a norm T that replaces the original min and the binary T-norm on the interval [0, 1] is a triangular norm (t-norm).

Definition: A triangular norm (t-norm) T is an increasing associative and commutative $[0,1]^2 \to [0,1]$ mapping that satisfies the boundary condition for every $x \in [0,1]$, $T(x,1) = x$.

Therefore $\left(\tilde{A} \otimes \tilde{B}\right)(z) = \sup\limits_{x,y,x.y=z} \min\left[\tilde{A}(x), \tilde{B}(y)\right]$ can be written as $\left(\tilde{A} \otimes \tilde{B}\right)(z) = \sup\limits_{x,y,x.y=z} T\left[\tilde{A}(x), \tilde{B}(y)\right]$.

Some well-known continuous T- norm are the minimum operator T_M, the algebraic product T_p, and Lukasiewicx t-norm T_L defined by $T_L(x,y) = \max(x+y-1,0)$. The minimum operator T_M is the strongest (greatest) t-norm. the Weakest t-norm T_w is defined by

$$T_w(x,y) = \begin{cases} \min(x,y), \text{if } \max(x,y) = 1 \\ 0\ , \text{ elsewhere} \end{cases}$$

Addition of fuzzy intervals based on the weakest t-norm:

In this section, the addition based on the weakest t-norm which preserves the shape of fuzzy intervals is discussed. Consider 'n' L-R fuzzy intervals $\tilde{A}_i = (l_i, r_i, \alpha_i, \beta_i)_{LR}$, $i = 1,2,...,n$, then the T_w- sum $A = \underset{i=1}{\overset{n}{\oplus_{Tw}}} A_i$, is given by, $A = \underset{i=1}{\overset{n}{\oplus_{Tw}}} A_i = \left(\sum_{i=1}^{n} l_i, \sum_{i=1}^{n} r_i, \max_{i=1}^{n} \alpha_i, \max_{i=1}^{n} \beta_i\right)$. In the addition based on the minimum operator, the resulting spreads are the sum of the incoming spreads, while for the addition based on weakest t-norm resulting spreads are the greatest of the incoming spreads. Moreover, each t-norm may be shown to satisfy the following inequalities, $T_w(a,b) \le T(x,y) \le T_M(a,b) = \min(a,b)$ where $T_w(a,b) = \begin{cases} a\ , if\ b = 1 \\ b\ , if\ a = 1 \\ 0\ , otherwise \end{cases}$

Two characteristics can be observed in the previous research. First, the addition/ subtraction of fuzzy numbers by T_M and T_w preserve the original shape of the fuzzy numbers. With the T_M in the multiplication/division, the shapes of the original FNs may not be preserved. However, for given shapes, in multiplication, the T_w preserved the original FN's shape. Second the weakest t-norm operations can elicit more exacting performance, meaning smaller fuzzy spreads within uncertain environments. This exact performance may successfully reduce accumulating phenomena of fuzziness in complex systems. The addition, subtraction, multiplication and division of T_w fuzzy arithmetic can be seen in the following:

Let $\tilde{A} = (a, \alpha_A, \beta_A)_{LR}, \tilde{B} = (b, \alpha_B, \beta_B)_{LR}$ be two L-R fuzzy numbers. The fuzzy operations of T_w can be shown as follows:

(1) Addition:

$$\tilde{A} \oplus_W \tilde{B} = (a+b, \max(\alpha_A, \alpha_B), \max(\beta_A, \beta_B))_{LR}$$

(2) Subtraction

$$\tilde{A} (-)_W \tilde{B} = (a-b, \max(\alpha_A, \alpha_B), \max(\beta_A, \beta_B))_{LR}$$

(3) Multiplication:

$$\tilde{A} \otimes_w \tilde{B} = \begin{cases} (ab, \max(\alpha_A b, \alpha_B a), \max(\beta_A b, \beta_B a))_{LR} & for\ a, b > 0 \\ (ab, \max(\beta_A |b|, \beta_B |a|), \max(\alpha_A |b|, \beta_B |a|))_{RL}, & for\ a, b < 0 \\ (ab, \max(\alpha_A b, \beta_B |a|), \max(\beta_A b, \alpha_B |a|))_{LL}, & for\ a < 0, b > 0 \\ (0, \alpha_A b, \beta_A b)_{LR}, & for\ a = 0, b > 0 \\ (0, -\beta_A b, -\alpha_A b)_{RL}, & for\ a = 0, b < 0 \\ (0,0,0)_{LR}, & for\ a = 0, b = 0 \end{cases}$$

The $\alpha - cut$ arithmetic if repeatedly performed in an equation will accumulate the fuzziness of all fuzzy numbers involved. This property can be observed in complex systems when performed for each fuzzy interval. Therefore in order to reduce fuzzy accumulation the fuzzy arithmetic operations adopt the weakest t-norm arithmetic operations.

III. LEAST ABSOLUTE DEVIATION (LAD)

A. LAD BASED ON CONJUNCTION PROBLEM (Pushpa et al.[22])

To reduce the overall error of the model outputs, a new LAD function based on conjunction problem defined by Tanaka [11] is used. Let the h-level inequality possibility measure of two fuzzy numbers $\tilde{A}_1 = (a_1, \alpha_1)$ *and* $\tilde{A}_2 = (a_2, \alpha_2)$ is defined as inequality of two fuzzy numbers $\tilde{A}_1$ *and* $\tilde{A}_2$

$$= \left\{ \left| a_1 + L^{-1}(h)\alpha_1 - a_2 - L^{-1}(h)\alpha_2 \right| + \left| a_1 - L^{-1}(h)\alpha_1 - a_2 + L^{-1}(h)\alpha_2 \right| \right.$$

where a_i is the center and α_i is the spread. In order to minimize the distance between the two fuzzy numbers based on the above measure of inequality

B. LAD BASED ON SYMMETRIC DIFFERENCE (Pushpa et.al [23])

A method of estimating the regression coefficients using a length of the symmetric difference between the observed and the predicted response were considered. If $\tilde{A}_i = (a_i, \alpha_i)$ denotes a fuzzy number, then the addition and subtraction for two fuzzy numbers $\tilde{A}_i$ *and* $\tilde{A}_j$ in that form are $\tilde{A}_i \pm \tilde{A}_j = \left(a_i \pm a_j, \alpha_i + \alpha_j\right)$.
Generally regression analysis uses a method of minimizing the distance between the observed and predicted data to construct the regression model. However, the method of minimizing the sum of the difference between the actual and predicted outputs cannot be used. On the other hand, Assuming the length of the fuzzy number as the difference between the two end points. Thus the length of the fuzzy number $\tilde{A}_i$ is $2\alpha_i$ and denoted by $m(\tilde{A}_i) = 2\alpha_j$. Therefore, the length of symmetric difference between two fuzzy numbers $\tilde{A}_i$ *and* $\tilde{A}_j$ is denoted as,

$$m(\tilde{A}_i \Delta \tilde{A}_j) = \begin{cases} |(a_i - \alpha_i) - (a_j - \alpha_j)| + |(a_i + \alpha_i) - (a_j + \alpha_j)|, \\ \qquad if \;\; \tilde{A}_i \cap \tilde{A}_j \neq \phi \\ 2(\alpha_i + \alpha_j) \qquad if \;\; \tilde{A}_i \cap \tilde{A}_j = \phi \end{cases}$$

The Least absolute deviation based on the symmetric difference between the observed and the estimated response variable is defined as follows.

$LAD = \left|\tilde{Y}_{iL} - \hat{\tilde{Y}}_{iL}\right| + \left|\tilde{Y}_{iU} - \hat{\tilde{Y}}_{iU}\right|$ where $\tilde{Y}_i$ is the observed response variable and $\hat{\tilde{Y}}_i$ is the estimated response variable. Using the T_w – norm based operations, the above distance can be defined as

$$LAD = \left| \left(a^T x_i + \sum_1^n \max_{1 \le j \le p} \left(|a_j| \gamma_{ij},\; |x_{ij}| \alpha_j \right) \right) - (y_i + e_i) \right| + \left| \left(a^T x_i - \sum_1^n \max_{1 \le j \le p} \left(|a_j| \gamma_{ij},\; |x_{ij}| \alpha_j \right) \right) - (y_i - e_i) \right|$$

C. LAD BASED ON HAUSDORFF DISTANCE

Based on the distance between the centers and spreads, a least absolute deviation method is discussed for fuzzy linear regression using T_w norm based arithmetic operations. Several metrics are defined on the family of all fuzzy numbers. The generalized Hausdorff metric fulfill many good properties and easy to compute in terms of generalized mid and spread functions If $d_H(A_\alpha, B_\alpha)$ is the Hausdorff metric between crisp sets A_α *and* B_α given by, $d_H(A_\alpha, B_\alpha) = \max\left\{ \sup_{b \in B_\alpha} \inf_{a \in A\alpha} |a-b| \;,\; \sup_{a \in A_\alpha} \inf_{b \in B_\alpha} |a-b| \right\}$

If $I_1 = [a_1, a_2]$ *and* $I_2 = [b_1, b_2]$ are two intervals, then

$$d_H(I_1, I_2) = \max\{|a_1 - b_1|, |a_2 - b_2|\} = |mid\; I_1 - mid\; I_2| + |spr\; I_1 - spr\; I_2| \qquad (1)$$

where $mid\; I_1 = \frac{a_1 + a_2}{2}$, $sprI_1 = \frac{a_2 - a_1}{2}$ (Trutsching et al. 2009)

The generalized Hausdorff metric between symmetric triangular fuzzy numbers $\tilde{A} = (a, \alpha)_T$, $\tilde{B} = (b, \beta)_T$ using(1), is then, $D_H(\tilde{A}, \tilde{B}) = |a - b| + L^{-1}(\alpha)|\alpha - \beta|$

$$D_1(\tilde{A}, \tilde{B}) = |a - b| + L_1 |\alpha - \beta| \quad , \quad L_1 = \int_0^1 L^{-1}(\alpha)\, d\alpha \;, L_1 = 0.5$$

$$D_\infty(\tilde{A}, \tilde{B}) = |a - b| + L_\infty |\alpha - \beta| \,, \quad L_\infty = \sup_{\alpha \in [0,1]} L^{-1}(\alpha),\; L_\infty = 1$$

The least absolute deviation based on the Hausdorff distance between the observed and the estimated response variable is defined as follows. $\left|Y_i - a^T x_i\right| + 0.5\left|e_i - \alpha^T x_i\right|$ using equation (1) where $\tilde{Y}_i$ is the center of the observed response and $a^T x_i$ is the center of the estimated response variable. e_i is the spread of the observed response and $\alpha^T x_i$ is the spread of the observed response. Using the T_w – norm based operations, the above distance can be defined as $LAD = \left|y_i - a^T x_i\right| + 0.5\left|e_i - \max_{1 \le j \le n}\left(|a_j| \gamma_{ij}, |x_{ij}| \alpha_j \right)\right|$,

IV FUZZY LINEAR REGRESSION USING LEAST ABSOLUTE DISTANCE AS OBJECTIVE FUNCTION

A linear regression model of fuzzy numbers that has a relationship between a response variable $\tilde{Y}_i$ (a symmetric triangular fuzzy numbers) and explanatory variables $\tilde{X}_i$ (with crisp / fuzzy numbers) are considered. The aim is to fit a fuzzy regression model with fuzzy coefficients to the aforementioned data set as follows:

$$\hat{\tilde{Y}}_i = \tilde{A}_0 \oplus_w (\tilde{A}_1 \otimes_w \tilde{X}_{i1}) \oplus_w \ldots.. \oplus_w (\tilde{A}_p \otimes_w \tilde{X}_{ip}) = \tilde{A} \otimes_w \tilde{X}_i, \;\; \boldsymbol{i = 1, \ldots. n} \qquad (2)$$

where $\tilde{A}_j = (a_j, \alpha_j),\ j = 1,\ldots p$ are symmetric fuzzy numbers and the arithmetic operations are based on the weakest t- norm. Fuzzy linear regression analysis can be seen as an optimization problem where the aim is to derive a model which fits the given dataset. The parameters are optimized in such a way that the difference between the observed outputs $\tilde{Y}_i$ and the estimated ones $\hat{\tilde{Y}}_i$ are made as small as possible.

To show the fitness (performance) of the fuzzy linear regression model, we compare the fuzzy estimated response of the model $\hat{\tilde{Y}}_i$ with the observed one $\tilde{Y}_i$ where 'i' is the index of the data. There are different measures to determine the similarity between two fuzzy numbers. Here the similarity measure based graded mean integration representation [24] is used, which gives more accurate results. The similarity measure is just for comparison purposes, and any other type of objective function can be designed and then applied. $S(\tilde{Y}_i, \hat{\tilde{Y}}_i) = \dfrac{1}{1 + ABS(a^T x_i - y_i)}$. The similarity measure varies between 0 and 1, so the closer the value to 1, the better the model. However for the sake of conformity, the dissimilarity measure [1- $S(\tilde{Y}_i, \hat{\tilde{Y}}_i)$] is considered and the value closer to 0, the better the model. Based on the LAD defined in the above section, the objective function in the optimization problem using the T_w norm operations is defined as follows:

$$\min \sum_{i=1}^{n} \left(LAD\ between\ observed\ and\ estimated\ responses \right) \tag{3}$$

subject to

$$\sum_{j=0}^{k} a_j x_{ij} + L^{-1}(h) \max_{1 \le j \le k} \left(|a_j| \gamma_{ij},\ |x_{ij}| \alpha_j \right) \ge y_i - L^{-1}(h) e_i$$

$$\sum_{j=0}^{k} a_j x_{ij} - L^{-1}(h) \max_{1 \le j \le k} \left(|a_j| \gamma_{ij},\ |x_{ij}| \alpha_j \right) \le y_i + L^{-1}(h) e_i, \quad 0 < h < 1,$$

$$\max_{1 \le j \le k} \left(|a_j| \gamma_{ij},\ |x_{ij}| \alpha_j \right) \ge 0, \quad \forall i = 1, 2, \ldots n$$

where the variables are x_{ij} : value of the j^{th} independent variable for the i^{th} observation, y_i : value of the dependent variable for the i^{th} observation, and the parameters are e_i : spread of the dependent variable for the i^{th} observation, h : target degree of belief, a_j : midpoint of the j^{th} regression coefficient, α_j : spread of the j^{th} regression coefficient, k : number of independent variables, n : number of observations.

V Example: SYMMETRIC FUZZY INPUT AND OUTPUT DATA SET

Consider the data set in Table 1 in which the observations of the input and output variables are symmetric triangular fuzzy numbers with their center and spread. Sakawa and Yano [25] introduced this fuzzy input – fuzzy output dataset which is given in Table 2.1. Many approaches present in the literature (Nasrabedi and Nasrabedi [26], Diamond [27], Kao and Chyu [28], Chen and Dang 2008, Sakawa and Yano [25], Hojati et al [29] have used this data set for validation purposes. The proposed method is compared with some methods such as Hong et al [20], Nasrabedi and Nasrabedi [26] and Diamond [27], they applied least square method. The result of fitting model to the FIFO data set is given in Table 1.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

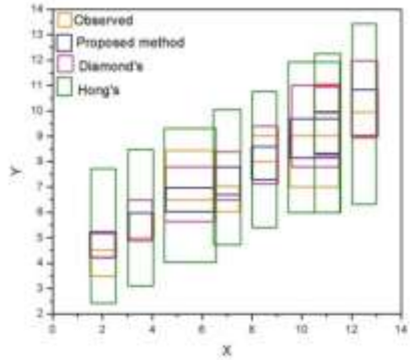

Fig.1 Comparison of Fuzzy regression using Conjunction problem with Diamond's method, Hong's method

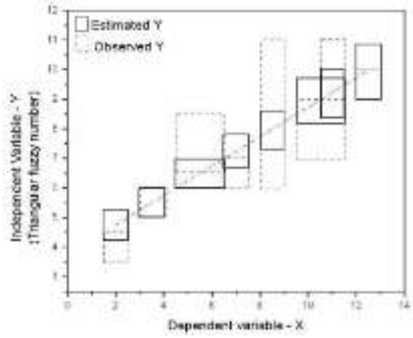

Fig. 2 Estimated fuzzy linear regression model using the LAD distance with conjunction problem with Tw – norm based operations

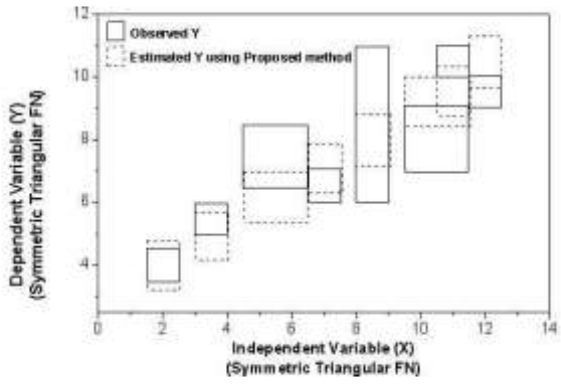

Fig. 3 Estimated fuzzy linear regression model using the LAD based on the symmetric difference with Tw – norm based operations

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

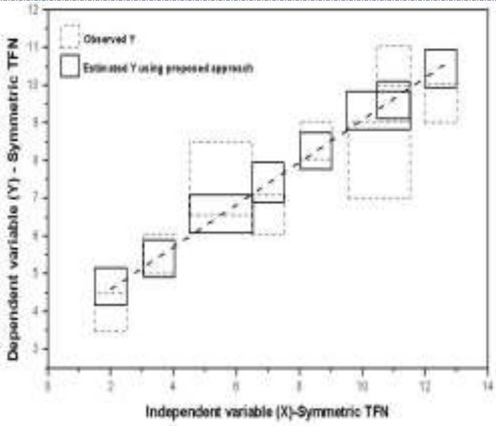

Fig.4 Estimated fuzzy linear regression model using the LAD based on the Hausdorff difference between the estimated and observed responses with Tw – norm based operations

Table 1 Comparison between the similarity measures between different LAD models for fuzzy input and fuzzy output dataset

Explanatory variable	Response variable	Similarity measure					
		Conjunction problem	**Symmetric difference**	**Hausdorff distance**	**Hong et al method**	**Nasrabedi & Nasrabedi Method**	**Diamond Method**
(2.0, 0.5)	(4.0, 0.5)	0.5682	0.9833	0.6452	0.5682	0.9833	0.6452
(3.5, 0.5)	(5.5, 0.5)	0.9950	0.6400	0.9049	0.9950	0.6400	0.9049
(5.5, 1.0)	(7.5, 1.0)	0.4938	0.4282	0.5048	0.4938	0.4282	0.5048
(7.0, 0.5)	(6.5, 0.5)	0.5848	0.6309	0.5371	0.5848	0.6309	0.5370
(8.5, 0.5)	(8.5, 0.5)	0.6431	0.6693	0.7722	0.6431	0.6693	0.7722
(10.5, 1.0)	(8.0, 1.0)	0.5195	0.4478	0.4294	0.5195	0.4478	0.4294
(11.0, 0.5)	(10.5, 0.5)	0.4292	0.5101	0.5290	0.4292	0.5101	0.5290
(12.5, 0.5)	(9.5, 0.5)	0.7117	0.5102	0.5121	0.7117	0.5102	0.5120
	Average	**0.6182**	**0.6025**	**0.6043**	0.5893	0.6131	0.5929

V A CASE STUDY USING LAD BASED ON HAUSDORFF DISTANCE

As an example with multiple inputs given in Table2, the fuzzy linear regression model obtained by the LAD with Hausdorff distance approach is $\tilde{Y} = (0,1.2346) \oplus_w (0,0.7213) \otimes_w X_1 \oplus_w (0,0.9505) \otimes_w X_2 \oplus_w (0.3414,0.2537) \otimes_w X_3$ with optimum value h= 0.2464 and the graph is given in Fig5 In the above table 2, third data is an abnormal data, using the proposed approach which is insensitive to the outlier yielded a better result. The table 4 explains that the mean similarity measure for the proposed model is 0.314 which has effective performance with other existing methods in the literature.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

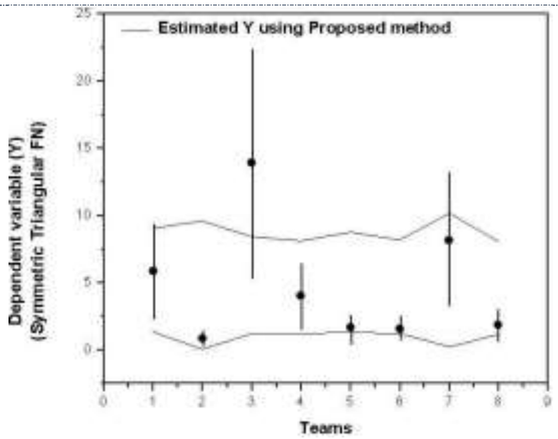

Fig.5 The proposed fuzzy regression equation for the data set in Table 4

Table 2. Kim and Bishu's data set [14] using the proposed approach

Obs.	Dependent variable (response time)	Independent variable (Inside control room experience)	Independent variable(outside control room experience)	Independent variable (Education)	Estimated $(\tilde{y}_i, \gamma_i)$	Similarity measure	Goodness of fit
Team 1	(5.83,3.56)	2	0	15.25	(5.206,3.868)	0.615612	0.99296
Team 2	(0.85,0.52)	0	5	14.13	(4.823.4.752)	0.201074	0.566704
Team 3	(13.93,8.5)*	1.13	1.5	14.13	(4.823,3.584)	0.098944	0.566709
Team 4	(4, 2.44)	2	1.25	13.63	(4.653,3.458)	0.605107	0.98783
Team 5	(1.65,1.01)	2.19	3.75	14.75	(5.035,3.752)	0.228055	0.602016
Team 6	(1.58,0.96)	0.25	3.5	13.75	(4.694,3.488)	0.243096	0.612613
Team 7	(8.18,4.99)	0.75	5.25	15.25	(5.206,4.99)	0.25161	0.915005
Team 8	(1.85,1.13)	4.25	2	13.5	(4.608,3.425)	0.266085	0.692983
				Average		**0.313698**	**0.742102**

* indicates the outlier data

Table. 3 Comparison of different models available in the literature for the data set in Table 2

Method	Estimated fuzzy regression function	Similarity measure using [24]
This work	$\tilde{Y} = (0,1.2346) \oplus_w (0,0.7213) \otimes_w X_1 \oplus_w (0,0.9505) \otimes_w X_2 \oplus_w (0.3414,0.2537) \otimes_w X_3$	**0.314**
Choi and Buckley's [30]	$\tilde{Y} = -2.8273 \oplus 0.3878 \otimes X_1 \oplus 1.0125 \otimes X_2 \oplus 0.6185 \otimes X_3 \oplus (0,1.0696,2.0042)$	**0.2155**
Chen Hseuh [16]	$\tilde{Y} = -16.7956 \oplus 1.0989 \otimes X_1 \oplus 1.1798 \otimes X_2 \oplus 1.8559 \otimes X_3 \oplus (0,2.8888)$	**0.1222**
Hassanpour et al. [31]	$\tilde{Y} = (-2.8273,0.0000) \oplus (0.3877,0.0000) \otimes X_1 \oplus (1.0125,0.000) \otimes X_2 \oplus (0.6185,0.1790) \otimes X_3$	**0.1630**
Taheri and Kelkinnama [18]	$\tilde{Y} = -15.5578 \oplus (0.2444,0) \otimes X_1 \oplus (0.9976,0) \otimes X_2 \oplus (1.5142,0) \otimes X_3 + (0,1.13)$	**0.2019**

CASE STUDY

To verify the performance of the proposed fuzzy linear regression method, we apply the method on the following practical problem of investigating the relationship between Diesel Fuel properties and emissions from engines. The air pollution has become more serious due to automotive emissions in the last few years. The manufacturers must introduce new techniques to control the emissions. Investigation of the effect of the fuel properties on the emissions from engines is important for controlling the emission from the engines and thus decreasing the air pollution. Selecting the fuel properties as the independent variables and the emissions from engines as the dependent variables, linear regression equations can be constructed with the statistical method. The relationship between the fuel properties and emissions can be analyzed with the equation. Since the fuel properties correlative with each other, when studying the effect of one of the properties must be separated from the other properties. The relationship between the fuels properties and emissions can be analyzed with the equation. Since the fuel properties correlative with each other, when studying the effect of one of the properties must be separated from the other properties. The relationship between the fuel property and the emission is fuzzy.

This study constructs the fuzzy relationship between fuel properties and the emission using the proposed method. The results obtained with the equations are analyzed and discussed.

The sample data used in this paper come from the paper [32]. Eleven fuels with different properties were given in Table 4. Table 5 gives the emissions from engines HC, CO NO_X, PM.

Table 4 The properties of the fuels

No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Q/kg/m³	829.2	828.8	857.0	855.1	828.8	855.5	826.9	855.1	855.4	826.6	827
P/%	1.0	7.7	1.1	7.4	7.1	7.6	1	7.3	8	1.1	0.9
CN/ τ	51.0	50.2	50.0	50.3	50.6	50.2	49.5	54.8	59.1	58	57.1
T_{95}/⁰C	344	349	348	344	346	371	326	345	344	347	329

Table 5The emissions form Engines /g/km

No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
HC	0.085	0.085	0.111	0.103	0.085	0.099	0.089	0.083	0.073	0.063	0.066
CO	0.432	0.431	0.551	0.517	0.446	0.513	0.454	0.434	0.378	0.331	0.327
NO_x	0.556	0.564	0.532	0.546	0.556	0.539	0.554	0.552	0.559	0.534	0.551
PM	0.048	0.052	0.061	0.063	0.051	0.066	0.045	0.065	0.066	0.051	0.049

By analyzing the properties of the fuels, fuel 1, 3, 4, 5, 7, 8, 9, 11 along with the corresponding results of the emissions are selected as the simulation samples. They are used to construct the fuzzy linear regression equations. Fuels 2, 6, 10 along with the corresponding results of the emissions are selected as the prediction sample. They are used to analyze the prediction ability of the Fuzzy linear regression equation.

Selecting density, polyaromatics, Cetane number and back end distillation as the independent variables and one of the emissions HC, CO NO_X or PM as the dependent variable. The fuzzy linear regression equation obtained in the form $\tilde{Y} = A_0 \oplus_w A_1 \otimes_w X_1 \oplus_w A_2 \otimes_w X_2 \oplus_w A_3 \otimes_w X_3 \oplus_w A_4 \otimes_w X_4$ where $A_i = (a_i, \alpha_i)$

The solution of the above equations is transformed into the solution of the following linear programming problem with the simulation samples:

$$\min \sum_{i=1}^{m} \left(\left| y_i - a^T x_i \right| + 0.5 \left| e_i - \max_{1 \le j \le n} \left(\left| a_j \right| \gamma_{ij}, \left| x_{ij} \right| \alpha_j \right) \right| \right) + \sum_{i=1}^{m} 1 - \left(\frac{1}{1 + abs(a^T x_i - y_i)} \right)$$

Subject to

$$\sum_{j=0}^{n} a_j x_{ij} + L^{-1}(h) \max_{1 \le j \le p} \left(\left| a_j \right| \gamma_{ij}, \left| x_{ij} \right| \alpha_j \right) \ge y_i - L^{-1}(h) e_i$$

$$\sum_{j=0}^{n} a_j x_{ij} - L^{-1}(h) \max_{1 \le j \le p} \left(\left| a_j \right| \gamma_{ij}, \left| x_{ij} \right| \alpha_j \right) \le y_i + L^{-1}(h) e_i$$

$$\max_{1 \le j \le p} \left(\left| a_j \right| \gamma_{ij}, \left| x_{ij} \right| \alpha_j \right) \ge 0, \quad \forall i = 1, 2, \ldots m$$

Because the sizes of different independent variables vary widely, in order to remove the effect of the sizes of them on the regression coefficients, the data is transformed into deviations from the average.

Table 6 Fuzzy regression coefficients

No.		**X0**	**X1**	**X2**	**X3**	**X4**	**Average Similarity measure**	**MAPE**
HC	a_i	0.8583×10^{-1}	0.2188×10^{-3}	0	0	0.3835×10^{-3}	98.96%	0.126

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

(h=0.28)	α_i	0.1698×10^{-1}	0.1763×10^{-2}	0	0	0		
CO (h=0.6)	a_i	0.4365	0.6752×10^{-3}	0	0	0.3485×10^{-2}	95.42%	0.115
	α_i	0.8403×10^{-1}	0	0	0	0.5336×10^{-2}		
NO_x (h=0.996)	a_i	0.5512	0	0	0.124×10^{-2}	0	99.49%	0.0095
	α_i	1.6841	0	0	0.2005	0.5287		
PM (h=0.95)	a_i	0.5601×10^{-1}	0.4639×10^{-3}	0.4991×10^{-3}	0.4056×10^{-3}	0.7769×10^{-4}	99.96%	0.0066
	α_i	0.1054×10^{-2}	0	0	0.2282	0		

Table.7 The simulation results calculated with the Fuzzy linear regression equations

Simulation samples		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Similarity measure**
HC	**LL**	0.0674	0.0651	0.0665	0.0681	0.06	0.0674	0.0662	0.0612	
	Center	0.0843	0.0919	0.0899	0.085	0.0769	0.0904	0.0901	0.07809	98.96%
	UL	0.1012	0.1187	0.1133	0.1019	0.0938	0.1138	0.114	0.0950	
CO	**LL**	0.3553	0.388	0.3728	0.362	0.291	0.3763	0.3729	0.3015	
	Center	0.4393	0.472	0.4568	0.446	0.375	0.4603	0.4569	0.3855	95.42%
	UL	0.5233	0.566	0.5408	0.53	0.459	0.5443	0.5409	0.4695	
NOx	**LL**	-1.1694	-3.2853	-1.1701	-2.2271	-1.137	-1.6932	-1.1592	-1.1276	
	Center	0.5489	0.5477	0.5481	0.5485	0.5471	0.5537	0.559	0.5565	99.49%
	UL	2.2672	4.3807	2.2663	3.3241	2.2312	2.8006	2.2772	2.2406	
PM	**LL**	0.04799	0.0608	0.0629	0.0508	0.0449	0.0603	0.0656	0.0473	
	Center	0.0481	0.0609	0.063	0.0509	0.045	0.0649	0.067	0.0483	99.96%
	UL	0.0482	0.061	0.0631	0.051	0.0451	0.0695	0.0684	0.0493	

from the average, the average of the dependent variable can be explained by the size of the fuzzy center of the coefficients Comparing with Table 7, because the data of each independent variable has been transformed into deviations for the constant x_0. The simulation accuracy of the equations is decided by the fuzzy centers and fuzzy widths of the coefficients. The fuzzy centers decide the biases of the center values of the simulation results to the actual values. The fuzzy width decides the sizes of the intervals of the simulation results.

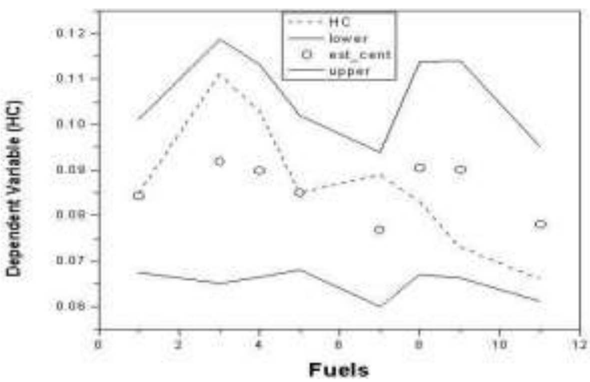

Fig. 6 Estimated emission of HC

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

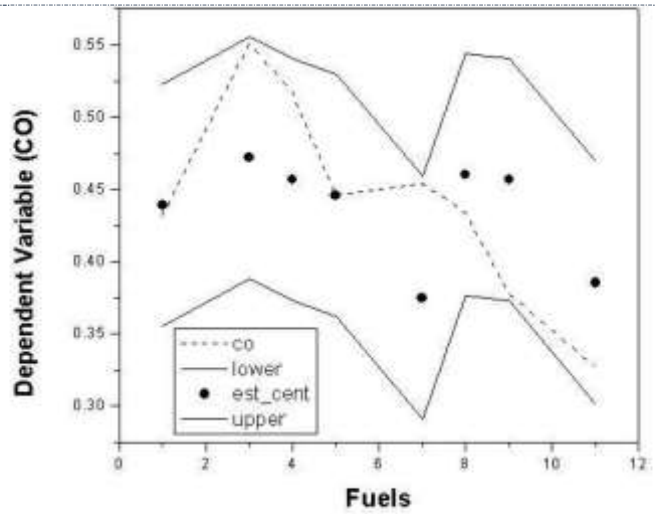

Fig. 7 Estimated emission of CO

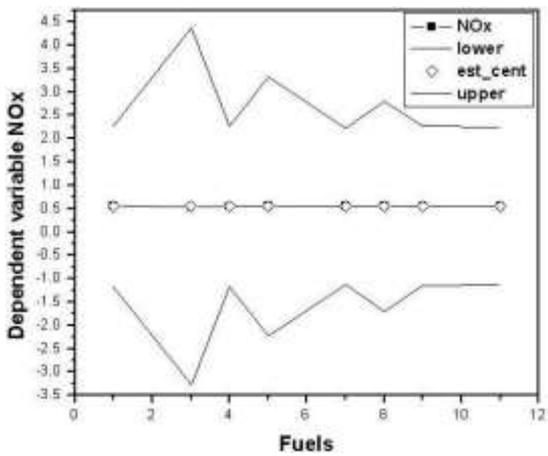

Fig. 8 Estimated emission of NOx

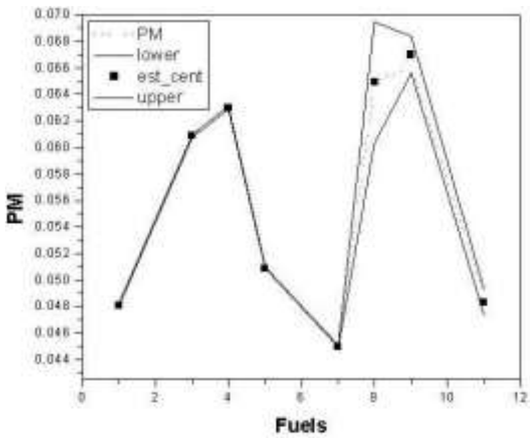

Fig. 9. Estimated emission of PM

Figures 6, 7, 8 and 9 shows that estimated emission HC, CO, NOx and PM values which are in fuzzy have more than 95% of similarity measure with the observed ones and the forecasted emissions have MAPE value almost nearest to zero. The estimated center coincides with the observed crisp values in the Fig. 6, 7, 8, 9. The trend of estimated center values represents the general trend of the overall system which implies that the system has the ability to estimate the trend of the system fuzziness, but also the trend of the system center. The

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

estimated fuzzy function using the proposed method gives the conventional regression equation which results accurate estimated values of emissions.

VI CONCLUSION

The above used methods are based on LAD using conjunction problem, symmetric difference and Hausdorff distance with weakest t-norm based operations and the estimated response is compared with the methods that exists in the literature. It is proved that these methods with Tw- norm based operations are efficient and more accurate when they are used with different types of input data and even with outlier dataset. The fuzzy regression model with LAD with Hausdorff distance is illustrated with crisp input and fuzzy output, data, fuzzy input and fuzzy output data and also with crisp multiple inputs. Selecting the fuel properties as the independent variables, and the emissions from engines as the dependent variables, the fuzzy linear regression equations are constructed; the simulation results calculated with the equations are analyzed.

VII. REFERENCES

[1] J. Andres, Claim reserving with fuzzy regression and Taylor’s geometric separation method, Insurance: Mathematics and Economics 40 (2007) 145–163.
[2] J. Andres, Calculating insurance claim reserves with fuzzy regression, Fuzzy Sets and Systems 157 (2006) 3091–3108.
[3] K. Chena, M.J. Rysb, E.S. Lee, Modeling of thermal comfort in air conditioned rooms by fuzzy regression analysis, Mathematical and Computer Modelling 43 (2006) 809–819.
[4] Y.Q. He, L.K. Chan, M.L. Wu, Balancing productivity and consumer satisfaction for profitability: statistical and fuzzy regression analysis, European Journal of Operational Research, 176 (2007), pp. 252–263.
[5] C.Y. Huang, G.H. Tzeng, Multiple generation product life cycle predictions using a novel two-stage fuzzy piecewise regression analysis method,Technological Forecasting and Social Change 75 (2008) 12–31.
[6] S.J. Mousavia, K. Ponnambalamb, F. Karrayb, Inferring operating rules for reservoir operations using fuzzy regression and ANFIS, Fuzzy Sets and Systems 158 (2007) 1064–1082.
[7] J. D. Sanchez, A.T. Gomez, Applications of fuzzy regression in actuarial analysis, Journal of Risk Insurance 70 (2003) 665–699.
[8] B. Wu, N.F. Tseng, A new approach to fuzzy regression models with application to business cycle analysis, Fuzzy Sets and Systems 130 (2002) 33–42.
[9] H. Tanaka and J. Watada, “Possibilistic linear systems and their applications to the linear regression model”, Fuzzy Sets and Systems, Vol. 27, Pp. 275-289, 1988.
[10] H. Tanaka, I. Hayashi and J. Watada, “Possibilistic linear regression analysis for fuzzy data”, Eur. J.Opera. Res. Vol. 40, Pp. 389-396, 1980.
[11] H. Tanaka, “Fuzzy data analysis by possibilistic linear models”, Fuzzy Sets and Systems, Vol. 24, Pp. 363-375, 1987.
[12] D. T. Redden and W. H. Woodall, “Properties of certain fuzzy linear regression models”, Fuzzy Sets and Systems, Vol. 44, Pp. 361-375, 1994.
[13] E. S. Lee and P.-T. Chang, “Fuzzy linear regression analysis with spread un- constrained in sign,” Computers and Mathematics with Applications, vol. 28, no. 4, pp. 61–70, 1994
[14] B. Kim and R. R. Bishu, “Evaluation of fuzzy linear regression models by comparing membership functions,” Fuzzy Sets and Systems, vol. 100, no. 1-3, pp. 343–352, 1998
[15] H. Torabi and J. Behboodian, (2007) “Fuzzy least absolutes estimates in linear regression models”, Communi. Stat. – Theory methods, Vol. 36, pp. 1935-1944L.-H.
[16] L. H. Chen and C.-C. Hsueh, “Fuzzy regression models using the least squares method based on the concept of distance,” IEEE Transactions on Fuzzy Systems, vol. 17, no. 6, pp. 1259–1272, 2009.
[17] Choi S. H. & Buckley J. J, (2008) “Fuzzy regression using least absolute deviation estimators”, Soft Comput. Vol. 12, pp. 257-263.
[18] S.M. Taheri and M. Kelkinnama, (2012) “Fuzzy linear regression based on least absolute deviations”, Iranian J. fuzzy system Vol. 9, pp. 121-140.
[19] H. Tanaka, S. Uejima, and A. Asai, “Linear regression analysis with fuzzy model,” IEEE Transactions on Systems, Man and Cybernetics, vol. 12, pp. 903–907, 1982.

[20] D. H. Hong, S. Lee and D.Y. Do, (2001) "Fuzzy linear regression analysis for fuzzy input-output data using shape preserving operations", Fuzzy sets Syst. Vol.122, pp. 513-526.
[21] L. A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353
[22] B. Pushpa, R. Vasuki, "Estimation of confidence interval level 'h' in FLR analysis using shape preserving operations, International journal of Computer applications, Vol. 68, No. 17 , 2013.
[23] B. Pushpa, S. Murugananandam, " Development of Tw norm based novel fuzzy regression model, Asian Journal of Fuzzy and applied Mathematics, Vol. 01, Issue 04 (2013).
[24] C.H. Hsieh and S.H. Chen, (1999) "Similarity of generalized fuzzy numbers with graded mean integration representation", in Proceedings of Eighth international fuzzy system association world congress, vol2, Taipei, Taiwan, Republic of China, pp. 551-555
[25] Sakawa, M. and Yano, Y., "Multiobjective fuzzy linear regression analysis for fuzzy input and fuzzy output data", Fuzzy Sets and Systems, 47, 173-181, 1992.
[26] Nasrabedi, M. M. and Nasrabedi, E., "Mathematical programming approaches to fuzzy linear regression analysis", Applied Mathematics and Computation, 155, 873-881, 2004.
[27] Diamond, P., "Fuzzy least squares", Information Sciences, 46, 141-157, 1988.
[28] Kao, C. and Chyu, C. L., "Least-squares estimates in fuzzy regression analysis", European Journal of Operational Research 148, 426–435, 2003.
[29] M. Hojati, C. R. Bector, K. Smimou, A Simple method for computation of fuzzy linear regression, European Journal Of Operational Research (2004).
[30] S. H. Choi and J.J. Buckley, (2008) "Fuzzy regression using least absolute deviation estimators", Soft Comput. Vol. 12, pp. 257-263.
[31] H. Hassanpour, A Goal Programming Approach to Fuzzy Linear Regression. Ph. D thesis, Shahid Bahonar University of Kerman, Kerman, Iran, 2008.
[32] M. Hublin, P. G. Gadd, D. E. Hall, etc. European Programmes on Emissions, Fuels and Engine Technologies (EPEFE) —Light Duty Diesel Study, SAE961073. Journal of Fuels &Lubricants 1996; 4: 527-56.