Fuzzy Regression Model With Interval-Valued Fuzzy Input-Output Data

Mohammad Reza Rabiei*†, Naser Reza Arghami*, S. Mahmoud Taheri‡ and Bahram Sadeghpour*
*Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, 91775 Iran
†Corresponding author, Email: mo_ra519@stu-mail.um.ac.ir, TelFax: +985118828605
‡Department of Basic Engineering Science, College of Engineering, University of Tehran, Tehran, Iran,
and Department of Mathematical Sciences, Isfahan University of Technology, Isfahan 84156-83111, Iran

***Abstract*—A novel approach is introduced to construct a fuzzy regression model when both input data and output data are interval-valued fuzzy numbers. Using a distance on the space of interval-valued fuzzy numbers, a least-squares method is developed. Also, a nonlinear programming model is proposed to estimate the crisp parameters for the interval-valued fuzzy regression model. A real example demonstrates the feasibility and efficiency of the proposed method. Moreover, two goodness of fit indices are introduced and employed for more evaluation of such fuzzy interval-valued regression models.**

Keywords—Interval-valued fuzzy number, fuzzy regression, least-squares method, goodness of fit.

I. Introduction

Since fuzzy set theory was introduced by Zadeh [1], many new approaches and theories treating imprecision and uncertainty have been proposed. Specifically, the intuitionistic fuzzy set theory pioneered by Atanassov [2] and the interval-valued fuzzy set theory suggested by Grozafczany [3] and Turksen [4] are two well-known generalizations of the fuzzy set theory. In fact, it has been pointed out that there is a strong connection between Atanassov's intuitionistic fuzzy sets and the interval-valued fuzzy sets [5], [6], [7]. Over the last decades, the theory of interval-valued fuzzy sets has been developed in different directions. In this introduction we shall briefly review some works on this topic.

Gorzalczany [8] investigated approximate reasoning based on interval-valued fuzzy sets. Wang and Li [9], [10] presented the applications of interval-valued fuzzy numbers and interval-distribution numbers in pseudo-probability metric spaces. They investigated the concept of interval-valued fuzzy number and studied some of its properties and presented a method for calculating correlation between and information energy of interval-valued fuzzy numbers. Hong and Lee [11] presented some algebraic properties and a distance measure for interval-valued fuzzy numbers. Grzegorzewski [12] studied some distances between interval-valued fuzzy sets based on Hausdorff metric. Wang et al. [13] investigated the combination and normalization of the interval-valued belief structures. Deschrijver [14] investigated some arithmetic operators in interval-valued fuzzy set theory. Chen and Chen [15] presented a method for handling information filtering problems based on interval-valued fuzzy numbers and presented a similarity measure between interval-valued fuzzy numbers. In Chen [16], Chen presented a method for handling the similarity measure problems of interval-valued fuzzy numbers. Chachi and Taheri[17] investigated two general classes of similarity measures between intuitionistic fuzzy sets. Chen and Ouyang [18] investigated an inventory model by fuzzifying the carrying cost rate and interest earned rate, simultaneously, based on the interval-valued fuzzy numbers.

Although, there has been a lot of research on fuzzy regression analysis (see, e.g. [19], [20], [21], [22]), so far as the authors know, there is no work on regression analysis for interval-valued fuzzy data. In this paper we introduce an approach to the problem of regression modeling when the available data of the response variable (output) and independent variables (inputs) are interval-valued fuzzy numbers. To do this, we consider a regression model, in which the coefficients are crisp. We then use a distance on the space of interval-valued fuzzy numbers and a least-squares method to obtain coefficients of the proposed model.

The rest of this paper is organized as follows. In Section II, we review some preliminaries of interval-valued fuzzy sets theory. In Section III, a distance between interval-valued fuzzy numbers is introduced. In Section IV, we state the proposed model and explain how the coefficients are obtained. In Section V, the proposed model is illustrated via a real world data set in the filed of soil science. The obtained models are evaluated by using some indices in Section VI. A brief onclusion is given in the last section.

II. Preliminaries

In this section, we review some elementary definitions and a well-known result of the interval-valued fuzzy sets and interval-valued fuzzy numbers, due to Wang and Li [9], Hong and Lee [11] and Zhixin and Hongmei [23]. Let $I=[0,1]$ and $[I]=\{[a,b] | a\leq b, a,b\in I\}$. For any $a\in I$, define $\bar{a}=[a,a]$.

Definition 1. *We define* $\vee_{t\in T}a_t=\sup\{a_t : t\in T\}$ *and* $\wedge_{t\in T}a_t=\inf\{a_t : t\in T\}$, *where* $a_t\in I, t\in T$. *We also define for* $[a_t,b_t]\in[I], t\in T$,

1) $\bigvee_{t\in T}[a_t,b_t]=[\bigvee_{t\in T}a_t,\bigvee_{t\in T}b_t]$,
$\bigwedge_{t\in T}[a_t,b_t]=[\bigwedge_{t\in T}a_t,\bigwedge_{t\in T}b_t]$,
2) $[a_1,b_1]=[a_2,b_2]$ *iff* $a_1=a_2, b_1=b_2$,
$[a_1,b_1]\leq[a_2,b_2]$ *iff* $a_1\leq a_2, b_1\leq b_2$,
$[a_1,b_1]<[a_2,b_2]$ *iff* $[a_1,b_1]\leq[a_2,b_2]$, *but* $[a_1,b_1]\neq[a_2,b_2]$.

Definition 2. *Let* X *be an ordinary nonempty set. Then*

- *The mapping* $A: X\to[I]$ *is called an interval-valued* *fuzzy set (IVFS) on X. The set of all IVFS on X is denoted by IF(X).*

- *For $A \in IF(X)$, let $A(x) = [A^-(x), A^+(x)]$, for all $x \in X$. Then two fuzzy sets $A^- : X \to I$ and $A^+ : X \to I$ are called lower fuzzy set and upper fuzzy set of A, respectively.*

- *The value of $\Pi_A(x) = A^+(x) - A^-(x)$ is called the degree of non-determinancy of the element $x \in X$ in the IVFS A.*

Definition 3. *Let $A \in IF(X)$ and $[\lambda_1, \lambda_2] \in [I]$. We call $A_{[\lambda_1,\lambda_2]} = \{x \in X : A^-(x) \geq \lambda_1, A^+(x) \geq \lambda_2\}$ and $A_{(\lambda_1,\lambda_2)} = \{x \in X : A^-(x) > \lambda_1, A^+(x) > \lambda_2\}$ the $[\lambda_1, \lambda_2]$-level set of A and the (λ_1, λ_2)-level set of A, respectively.*

Definition 4. *Let $A \in IF(R)$, where R is the real line. Assume the following conditions are satisfied*

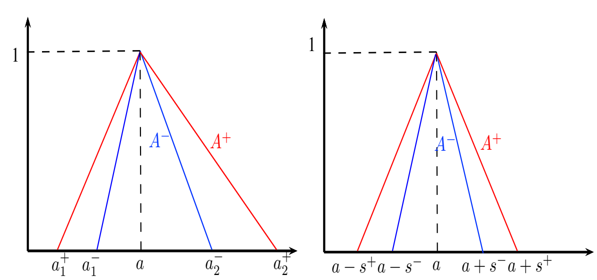
- *A is normal, i.e., there exists $x_0 \in R$, s.t. $A(x_0) = \bar{1}$,*

- *For arbitrary $[\lambda_1, \lambda_2] \in [I]^+ = [I] - \{\bar{0}\}$, $A_{[\lambda_1,\lambda_2]}$ is a closed bounded interval.*

Then we call A an interval-valued fuzzy number (IVFN). We denote the set of all IVFNs by $IF^*(R)$.

Definition 5. *[9] Let $A, B \in IF(R)$ and $\circledast \in \{+, -, \cdot, \div\}$. We define the extended operations by $(A \circledast B)(z) = \bigvee_{z=x\circledast y}(A(x) \wedge B(y))$. For each $[\lambda_1, \lambda_2] \in [I]^+$, we write $A_{[\lambda_1,\lambda_2]} \circledast B_{[\lambda_1,\lambda_2]} = \{x \circledast y : x \in A_{[\lambda_1,\lambda_2]}, y \in B_{[\lambda_1,\lambda_2]}\}$.*

Definition 6. *A triangular IVFN is presented as $A = [A^-, A^+] = [(a_1^-, a, a_2^-),\ (a_1^+, a, a_2^+)]$, where A^- and A^+ denote the lower and upper triangular fuzzy numbers of A, $A^- \subseteq A^+$. Also, A is denoted by $A = [A^-, A^+] = [(a_1^+, a_1^-), a, (a_2^-, a_2^+)]$ where $a_1^+ \leq a_1^- \leq a \leq a_2^- \leq a_2^+$ (see Figure 1 (a)). Specifically, A is called symmetric if $A = [A^-, A^+] = [(a - s^-, a, a + s^-), (a - s^+, a, a + s^+)]$. In such a case A is shown by $A = [A^-, A^+] = (a, s^-, s^+)$ where $0 \leq s^- \leq s^+$ (see Figure 1 (b)).*



(a) (b)

Fig. 1. Two typical triangular interval-valued fuzzy numbers

Proposition 1. *[24] Let $A = [(a_1^+, a_1^-), a, (a_2^-, a_2^+)]$ and $B = [(b_1^+, b_1^-), b, (b_2^-, b_2^+)]$ are two triangular IVFNs. Then*
- *Extended addition is obtained as*

$$A \oplus B = [(a_1^+ + b_1^+, a_1^- + b_1^-), a + b, (a_2^- + b_2^-, a_2^+ + b_2^+)]$$

- *Extended scalar multiplication is obtained as*

$$\lambda A = \begin{cases} \left[\left(\lambda a_1^+, \lambda a_1^-\right), \lambda a, \left(\lambda a_2^-, \lambda a_2^+\right)\right], & \lambda \in [0, \infty); \\ \left[\left(\lambda a_2^+, \lambda a_2^-\right), \lambda a, \left(\lambda a_1^-, \lambda a_1^+\right)\right], & \lambda \in (-\infty, 0). \end{cases}$$

Corollary 2. *Let $A = (a, s_a^-, s_a^+)$ and $B = (b, s_b^-, s_b^+)$ be two symmetric triangular IVFNs. Then*

$$\begin{aligned} A \oplus B &= (a + b, s_a^- + s_b^-, s_a^+ + s_b^+), \quad (1)\\ \lambda A &= (\lambda a, |\lambda| s_a^-, |\lambda| s_a^+), \quad \lambda \in R. \quad (2) \end{aligned}$$

III. A NEW DISTANCE BETWEEN INTERVAL-VALUED FUZZY NUMBERS

Based on Definition 3.2 in [16] and the presented distances between two fuzzy numbers in [25], we propose the following definition of distance between IVFNs.

Definition 7. *Let $A, B \in IF^*(R)$. The $D^*_{p,f}$ distance between A and B is defined as*

$$D^*_{p,f}(A, B) = \max\{D_{p,f}(A^-, B^-), D_{p,f}(A^+, B^+)\} \quad (3)$$

in which for two fuzzy sets A° and B° ($\circ \in \{-, +\}$)

$$D_{p,f}(A^\circ, B^\circ) = (\int_0^1 f(\lambda) d^p(A^\circ_\lambda, B^\circ_\lambda) d\lambda)^{1/p}, \quad (4)$$

and

$$\begin{aligned} d^p(A^\circ_\lambda, B^\circ_\lambda) &= |a_1(\lambda) - b_1(\lambda)|^p + |a_2(\lambda) - b_2(\lambda)|^p, \\ A^\circ_\lambda &= [a_1(\lambda), a_2(\lambda)], \quad B^\circ_\lambda = [b_1(\lambda), b_2(\lambda)] \quad (5) \end{aligned}$$

whrer $a_1(\lambda), a_2(\lambda)$ are the lower and upper bounds of the λ-cut A° and $b_1(\lambda), b_2(\lambda)$ are the lower and upper bounds of the λ-cut B°. Also, $f(\lambda)$ is an increasing function on $[0, 1]$ with $f(0) = 0$ and $\int_0^1 f(\lambda) d\lambda = \frac{1}{2}$ (see [25]).

Specifically, for $p = 2$, we have

$$d^2(A^\circ_\lambda, B^\circ_\lambda) = (a_1(\lambda) - b_1(\lambda))^2 + (a_2(\lambda) - b_2(\lambda))^2. \quad (6)$$

In the following, we put $f(\lambda) = \lambda$ and we denote $D_{p,f}$ and $D^*_{p,f}$ by D_p and D^*_p, respectively.

In the following theorem, we prove that $D^*_{p,f}$ is a metric on the space of IVFNs. At first, we need to express the following lemma.

Lemma 3. *If a, b, c and d are real numbers, then*

$$\max\{a + b, c + d\} \leq \max\{a, c\} + \max\{b, d\} \quad (7)$$

Proof: See Appendix A. ∎

Theorem 4. *$D^*_{p,f}$ is a metric on $IF^*(R)$.*

Proof: See Appendix B. ∎

Proposition 5. *Let $A = (a_1, a, a_2)$ and $B = (b_1, b, b_2)$ be two triangular fuzzy numbers. Then*

$$\begin{aligned} D_2^2(A, B) &= \frac{(a - b)^2}{2} + \frac{1}{12}[(a_2 - b_2)^2 + (a_1 - b_1)^2] + \\ &\quad \frac{1}{6}(a - b)[(a_2 - b_2) + (a_1 - b_1)]. \quad (8) \end{aligned}$$

Proof: See Appendix C. ∎

Corollary 6. *Let $A = (a, s_a)$ and $B = (b, s_b)$ be two symmetric triangular fuzzy numbers. Then*

$$D_2^2(A, B) = (a - b)^2 + \frac{1}{6}(s_a - s_b)^2. \quad (9)$$

Proof: In Proposition 5, it is enough to note that $A = (a - s_a, a, a + s_a)$ and $B = (b - s_b, b, b + s_b)$. ∎

Theorem 7. *Let* $A = ((a_1^+, a_1^-), a, (a_2^-, a_2^+))$ *and* $B = ((b_1^+, b_1^-), b, (b_2^-, b_2^+))$ *be two triangular IVFNs. Then*

$$\begin{aligned} D_2^{*^2}(A, B) \;\; = \;\; & \frac{(a-b)^2}{2} + \\ & \max\{\frac{1}{12}[(a_2^- - b_2^-)^2 + (a_1^- - b_1^-)^2] + \\ & \frac{1}{6}(a-b)[(a_2^- - b_2^-) + (a_1^- - b_1^-)], \\ & \frac{1}{12}[(a_2^+ - b_2^+)^2 + (a_1^+ - b_1^+)^2] + \\ & \frac{1}{6}(a-b)[(a_2^+ - b_2^+) + (a_1^+ - b_1^+)]\} \end{aligned} \quad (10)$$

Proof: In view of Eq. (3) and Proposition 5 the proof is straightforward. ∎

Corollary 8. *Let* $A = (a, s_a^-, s_a^+)$ *and* $B = (b, s_b^-, s_b^+)$ *be two symmetric triangular IVFNs. Then*

$$D_2^{*^2}(A, B) = (a-b)^2 + \frac{1}{6}\max\{(s_a^- - s_b^-)^2, (s_a^+ - s_b^+)^2\} \quad (11)$$

Proof: In Theorem 7, it is enough to note that $A = ((a - s_a^+, a - s_a^-), a, (a + s_a^-, a + s_a^+))$ and $B = ((b - s_b^+, b - s_b^-), b, (b + s_b^-, b + s_b^+))$. ∎

IV. The Proposed Regression Model

Suppose that we have a data set denoted by $(\mathbf{y}_i, \mathbf{x}_{i1}, ..., \mathbf{x}_{in})$ $(i = 1, ..., m; m > n)$, where $\mathbf{y}_i, \mathbf{x}_{ij} \in IF(R)$ $(i = 1, ..., m,\ j = 1, \cdots, n)$. We wish to find, in an optimal way, the coefficients of the regression model

$$\mathbf{Y} = \beta_0 \oplus \beta_1 \mathbf{x}_1 \oplus ... \oplus \beta_n \mathbf{x}_n, \quad (12)$$

where $\mathbf{Y}, \mathbf{x}_i, i = 1, ..., m$ are IVFNs and $\beta_0, \beta_1 \cdots, \beta_n$ are crisp numbers.

To achieve this, we have to minimize the sum of squared distances between the estimated and observed IVF response variable , i.e.

$$Q(\beta_0, \beta_1, ..., \beta_n) = \sum_{i=1}^{m} D_2^{*^2}(\beta_0 \oplus \beta_1 \mathbf{x}_{i1} \oplus ... \oplus \beta_n \mathbf{x}_{in}, \mathbf{y}_i). \quad (13)$$

Writing $\mathbf{y}_i = (y_i, s_{y_i}^-, s_{y_i}^+)$ $(i = 1, ..., m)$ and $\mathbf{x}_{ij} = (x_{ij}, s_{x_{ij}}^-, s_{x_{ij}}^+)$ $(i = 1, \cdots, m, j = 1, ..., n)$, we have

$$\begin{aligned} & \beta_0 \oplus \beta_1 \mathbf{x}_{i1} \oplus ... \oplus \beta_n \mathbf{x}_{in} = \\ & (\beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}, \sum_{j=1}^{n} |\beta_j| s_{x_{ij}}^-, \sum_{j=1}^{n} |\beta_j| s_{x_{ij}}^+). \end{aligned}$$

Thus by Theorem 7, the sum of squared distances (13) can be rewritten as

$$Q(\beta_0, \beta_1, ..., \beta_n) = \sum_{i=1}^{m} (\beta_0 + \sum_{j=1}^{n} \beta_j x_{ij} - y_i)^2$$

$$+ \frac{1}{6}\sum_{i=1}^{m} \max\{(\sum_{j=1}^{n} |\beta_j| s_{x_{ij}}^- - s_{y_i}^-)^2, (\sum_{j=1}^{n} |\beta_j| s_{x_{ij}}^+ - s_{y_i}^+)^2\}. \quad (14)$$

By minimizing the sum of squared distances, one can estimate $\beta_0, \beta_1, \cdots, \beta_n$. To solve the above optimalization problem, we used Mathematica 6.0 [26].

Proposition 9. *For the IVF regression model (12), Let* $\mathbf{Y}_i = (Y_i, s_{Y_i}^-, s_{Y_i}^+)$ *and* $\mathbf{y}_i = (y_i, s_{y_i}^-, s_{y_i}^+), i = 1, ..., m$ *be the estimated and observed symmetric triangular IVF response for the* i*th observation, respectively. Then, for* $p = 2, f(\lambda) = \lambda$ *and* $i = 1, \cdots, m$*, we have*

$$D_{p,f}^*(\mathbf{Y}_i, \mathbf{y}_i) = (Y_i - y_i)^2 + \frac{1}{6}\max\{(s_{Y_i}^- - s_{y_i}^-)^2, (s_{Y_i}^+ - s_{y_i}^+)^2\}. \quad (15)$$

Proof: By Eq. (30) and Eq. (10), the proof is straightforward. ∎

Definition 8. *For the IVF regression model (12), the mean of distances between estimated and observed values is defined by*

$$MD_{f,p}^* = \frac{1}{m}\sum_{i=1}^{m} D_{f,p}^*(\mathbf{Y}_i, \mathbf{y}_i). \quad (16)$$

Note that, the above index, in some sense, is similar to the mean of squared errors in the statistical regression. So, one can use such an index to compare the fit of different fuzzy regression models which are obtained based on different data sets.

In below section, we provide an applied example to explain how the proposed method is applicable to deriving the regression model for interval-valued fuzzy observations.

V. Application to Soil Science

In soil science studies, sometimes, problems arise in measurement of physical, chemical and/or biological soil properties. The problem results from the difficulty, time and cost of direct measurements. Pedomodels (derived from Greek root of pedo as soil) have become a popular topic in soil science and environmental research. They are predictive functions of certain soil properties based on other easily or cheaply measured properties [27]. In this article, two pedomodels including one and two independent variables are studied to develop the relationships between different chemical and physical soil properties by means of interval-valued fuzzy least squares regression technique. Based on a study in a part of Silakhor plain (situated in a province west of Iran), a total of 24 core samples were obtained from 0.0 to 25-cm depth [28]. The data set is given in Table I and Table II.

1) Pedomodel of ESP-SAR: We first wish to provide a relationship between exchangeable sodium percentage (ESP), as the dependent variable, and sodium absorption ratio (SAR), as an independent variable. The exchange sodium percentage, ESP, governs the source/sink phenomenon for ionic constituents, i.e., sodium, as a contaminant in sodic soils, is calculated from the ratio of exchangeable sodium, Na_x, to cation exchangeable capacity, CEC. In soil science, cation-exchange capacity (CEC) is the maximum quantity of total cations, of any class, that a soil is capable of holding, at a given pH value, available for exchange with the soil solution. CEC is used as a measure of fertility, nutrient retention capacity, and the capacity to protect groundwater from cation contamination. It is expressed as milliequivalent of hydrogen per 100 g of dry

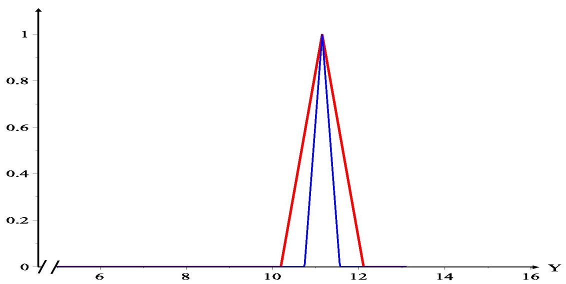


Fig. 2. Prediction of the EPS by IVF regression model for $SAR = (1.50, 0.06, 0.14)$

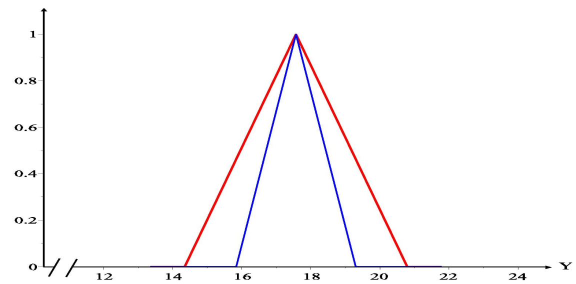


Fig. 3. Prediction of the CEC using IVF regression model (Eq. 18) for $SAND = (35, 1.48, 3.65)$ and $OM = (1.38, 0.54, 0.93)$

soil(meq+/100g), or the SI unit centi-mol per kg (cmol+/kg). The numeric expression is coincident in both units. All these soil parameters, measured on soil colloidal surface, are time consuming and costly. Due to close relationship between the distribution of cations in the exchange and solution phases, it is preferred to estimate ESP from sodium adsorption ratio, SAR, i.e., $Na/(Ca + Mg/2)^{0.5}$, in soil solution [29], [27].

In this case, ESP is considered as cost and time variable, therefore the need for less expensive indirect measurement is emphasized. Measurements of SAR have been related to ESP due to low cost, simplicity, and the possibility of relating measurements to the quantity and quality parameters. But, due to some impreciseness in related experimental environment, the observations of response variable (ESP) are given in fuzzy form. Thus, we may use a interval-valued fuzzy method for modeling such a data set [28] (see Table I).

According to the proposed method, the estimated coefficients are obtained as $\beta_0 = 0.835$ and $\beta_1 = 6.879$, and the IVF regression model is, therefore

$$Y = 0.835 \oplus 6.879\mathbf{x}. \tag{17}$$

The above IVF regression model can be applied to predict the ESP for a new case. For example, if for a new case, $SAR = (1.50, 0.06, 0.14)$ then, by Eq. (17), we predict the amount of ESP as $Y = (11.15, 0.41, 0.96)$. The membership functions of $\mathbf{Y}$ are shown in Fig. 2.

2) Pedomodel of CEC-OM-SAND: The second model provides a relationship between cation exchange capacity (CEC), as a function of two soil variables namely percentage of sand content (SAND) and organic matter content (OM) (Table II). In the soil, organic matter can enhance the CEC, while the sand content has negative effect on the cation exchange capacity [28].

According to the proposed method, the estimated coefficients are obtained as $\beta_0 = 21.97$, $\beta_1 = 2.57$ and $\beta_2 = -0.23$, and the IVF regression model is, therefore

$$\mathbf{Y} = 21.97 \oplus 2.57\mathbf{x}_1 \oplus (-0.23)\mathbf{x}_2. \tag{18}$$

The above IVF regression model can be used to predict the CEC of a new case. For example, if for a new case, $SAND = (35, 1.48, 3.65), OM = (1.38, 0.54, 0.93)$, then by Eq. (18), we predict the CEC as $Y = (17.57, 1.73, 3.22)$. The membership functions of Y are shown in Fig. 3.

VI. EVALUATION BY OTHER DISTANCES

In the following, we introduce two distances between interval-valued fuzzy numbers based on Hausdorff metric, for evaluating the goodness of fit of the IVF regression model. Let $u = [u_1, u_2]$ and $v = [v_1, v_2]$ be two closed intervals. The Hausdorff metric between u and v is defined by [30]

$$d_H(u, v) = \max\{|u_1 - v_1|, |u_2 - v_2|\}. \tag{19}$$

Definition 9. *[16] Let $A, B \in IF^*(R)$. The D_p^* distance between A and B is defined as*

$$D_p^*(A, B) = \max\{D_p(A^-, B^-), D_p(A^+, B^+)\} \tag{20}$$

where for fuzzy sets A° and B°

$$D_p(A^\circ, B^\circ) = (\int_0^1 d_H^p(A_\lambda^\circ, B_\lambda^\circ)d\lambda)^{1/p}. \tag{21}$$

Since A° and B° are fuzzy numbers, so for each $\lambda \in (0, 1]$, A_λ° and B_λ° are bounded closed intervals, i.e. $A_\lambda^\circ = [a_1(\lambda), a_2(\lambda)], B_\lambda^\circ = [b_1(\lambda), b_2(\lambda)]$. Therefore, from Eq. (19), we have

$$d_H(A_\lambda^\circ, B_\lambda^\circ) = \max\{|a_1(\lambda) - b_1(\lambda)|, |a_2(\lambda) - b_2(\lambda)|\}, \tag{22}$$

where $\circ \in \{-, +\}$.

Theorem 10. *[16] D_p^* is a metric on $IF^*(R)$.*

Proposition 11. *For the IVF regression model (12), Let $\mathbf{Y}_i = (Y_i, s_{Y_i}^-, s_{Y_i}^+)$ and $\mathbf{y}_i = (y_i, s_{y_i}^-, s_{y_i}^+)$, $i = 1, ..., m$ be the estimated and observed triangular IVF response for the ith observation, respectively. Then, for $i = 1, \cdots, m$, $D_p^*(\mathbf{Y}_i, \mathbf{y}_i)$ is obtained as*

$$D_p^*(\mathbf{Y}_i, \mathbf{y}_i) = \max\{D_p(\mathbf{Y}_i^-, \mathbf{y}_i^-), D_p(\mathbf{Y}_i^+, \mathbf{y}_i^+)\}, \tag{23}$$

where

$$\begin{aligned} D_p(\mathbf{Y}_i^-, \mathbf{y}_i^-) = (\int_0^1 & \max\{|(Y_i - y_i) - (1-\lambda)(s_{Y_i}^- - s_{y_i}^-)|^p, \\ & |(Y_i - y_i) + (1-\lambda)(s_{Y_i}^- - s_{y_i}^-)|^p\}d\lambda)^{1/p}, \\ D_p(\mathbf{Y}_i^+, \mathbf{y}_i^+) = (\int_0^1 & \max\{|(Y_i - y_i) - (1-\lambda)(s_{Y_i}^+ - s_{y_i}^+)|^p, \\ & |(Y_i - y_i) + (1-\lambda)(s_{Y_i}^+ - s_{y_i}^+)|^p\}d\lambda)^{1/p}. \end{aligned}$$

Proof: Proof. By Eq. (30) and Eq. (22), the result is obviously held. ■

Definition 10. *For the IVF regression model (12), the mean distance between the estimated and the observed values is defined by*

$$MD_p^* = \frac{1}{m}\sum_{i=1}^{m} D_p^*(\mathbf{Y}_i, \mathbf{y}_i). \tag{24}$$

Definition 11. *[16] Let $A, B \in IF^*(R)$. The D_∞^* distance between A and B is defined as*

$$D_\infty^*(A,B) = \max\{D_\infty(A^-,B^-), D_\infty(A^+,B^+)\} \tag{25}$$

where for fuzzy sets A° and B°

$$D_\infty(A^\circ,B^\circ) = \sup_{\lambda\in[0,1]} d_H(A_\lambda^\circ, B_\lambda^\circ), \tag{26}$$

for $\circ \in \{-,+\}$ and $d_H(A_\lambda^\circ, B_\lambda^\circ)$ can be obtained by Eq. (22).

Theorem 12. *[16] D_∞^* is a metric on $IF^*(R)$.*

Proposition 13. *For the IVF regression model (12), Let $\mathbf{Y}_i = (Y_i, s_{Y_i}^-, s_{Y_i}^+)$ and $\mathbf{y}_i = (y_i, s_{y_i}^-, s_{y_i}^+)$, $i = 1, ..., m$ be the estimated and observed triangular IVF response for the ith observation, respectively. Then, for $i = 1, \cdots, m$, $D_\infty^*(\mathbf{Y}_i, \mathbf{y}_i)$ is obtained as*

$$D_\infty^*(\mathbf{Y}_i, \mathbf{y}_i) = \max\{D_\infty(\mathbf{Y}_i^-, \mathbf{y}_i^-), D_\infty(\mathbf{Y}_i^+, \mathbf{y}_i^+)\}, \tag{27}$$

where

$$\begin{aligned} D_\infty(\mathbf{Y}_i^-, \mathbf{y}_i^-) = \sup_{\lambda\in[0,1]} \max\{&|(Y_i - y_i) - (1-\lambda)(s_{Y_i}^- - s_{y_i}^-)|,\\ &|(Y_i - y_i) + (1-\lambda)(s_{Y_i}^- - s_{y_i}^-)|\},\\ D_\infty(\mathbf{Y}_i^+, \mathbf{y}_i^+) = \sup_{\lambda\in[0,1]} \max\{&|(Y_i - y_i) - (1-\lambda)(s_{Y_i}^+ - s_{y_i}^+)|,\\ &|(Y_i - y_i) + (1-\lambda)(s_{Y_i}^+ - s_{y_i}^+)|\}. \end{aligned}$$

Proof: Proof. By Eq. (30) and Eq. (22), the result is obviously held. ■

Definition 12. *For the IVF regression model (12), the mean distance between the estimated and the observed values is defined by*

$$MD_\infty^* = \frac{1}{m}\sum_{i=1}^{m} D_\infty^*(\mathbf{Y}_i, \mathbf{y}_i). \tag{28}$$

By the two indices in (24) and (28), the goodness of fit of the obtained models were examined.

A. Evaluation of the pedomodels by the $MD_{f,p}^*$ and MD_∞^*

For $p = 2$, the indices MD_p^* and MD_∞^* between the observed values and the estimated values for two soil models are shown in Table I and Table II. As we see, the MD_p^* and MD_∞^* for the proposed model of ESP-SAR are 1.55 and 1.57, respectively, which are very close to 1.54, i.e. the $MD_{f,p}^*$. Also, the MD_p^* and MD_∞^* for the proposed model of CEC-OM-SAND are 1.92 and 2.53, respectively, which are very close to 1.41, i.e. the $MD_{f,p}^*$.

VII. Conclusion

In this work, we proposed a new approach to IVF regression analysis, based on least-squares method, for IVF input-IVF output data. The applicability of the proposed approach was investigated by using a real data set in soil science. By two indices, based on some distances between IVF numbers, the goodness of fit of the obtained models were examined. The extension of the proposed model to IVF input-IVF output data when they are nonsymmetric, is a potential topic for future work.

Appendix A
Proof of Lemma 3

We have 24 possible permutations of a, b, c and d. We prove inequality (7) for two cases.

- Let $a \le b \le c \le d$. Then $a + b \le c + d$, and therefore $\max\{a+b, c+d\} = c+d$, $\max\{a, c\} = c$ and $\max\{b, d\} = d$.
- Let $b \le c \le d \le a$. Then $\max\{a, c\} = a$ and $\max\{b, d\} = d$. If $\max\{a + b, c + d\} = c + d$, then

$$\begin{aligned} &c \le a \Rightarrow c + d \le a + d\\ &\quad i.e.\ \max\{a + b, c + d\} \le \max\{a, c\} + \max\{b, d\}. \end{aligned}$$

If $\max\{a + b, c + d\} = a + b$, then

$$\begin{aligned} &b \le d \Rightarrow a + b \le a + d\\ &\quad i.e.\ \max\{a + b, c + d\} \le \max\{a, c\} + \max\{b, d\}. \end{aligned}$$

The proof for the remaining 22 permutations is similar.

Appendix B
Proof of Theorem 4

Suppose $A, B, C \in IF^*(R)$.

- It is obvious that $D_{p,f}^*(A, B) \ge 0$.
- If $A = B$, then $D_{p,f}^*(A, B) = 0$. Conversely, if $D_{p,f}^*(A, B) = 0$, then $D_{p,f}(A^-, B^-) = D_{p,f}(A^+, B^+) = 0$. Therefore, $\forall x \in R, A^-(x) = B^-(x)$ and $A^+(x) = B^+(x)$, and so $A = B$.
- The symmetry property i.e. $D_{p,f}^*(A, B) = D_{p,f}^*(B, A)$ is clearly held.
- Triangular inequality Since $D_{p,f}(A^-, B^-)$ and $D_{p,f}(A^+, B^+)$ are metrics on the space of $F(R)$ ([31], [25]), if A^-, B^-, C^-, A^+, B^+, C^+ are fuzzy numbers, then

$$\begin{aligned} D_{p,f}(A^-, B^-) &\le D_{p,f}(A^-, C^-) + D_{p,f}(C^-, B^-),\\ D_{p,f}(A^+, B^+) &\le D_{p,f}(A^+, C^+) + D_{p,f}(C^+, B^+). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \max\{D_{p,f}(A^-, B^-), D_{p,f}(A^+, B^+)\} \le\\ \max\{D_{p,f}(A^-, C^-) + D_{p,f}(C^-, B^-),\\ D_{p,f}(A^+, C^+) + D_{p,f}(C^+, B^+)\}. \end{aligned} \tag{29}$$

By using relation (29) and Lemma 3, we have

$$\begin{aligned} D_{p,f}^*(A, B) &= \max\{D_{p,f}(A^-, B^-), D_{p,f}(A^+, B^+)\}\\ &\le \max\{D_{p,f}(A^-, C^-) + D_{p,f}(C^-, B^-),\\ &\qquad D_{p,f}(A^+, C^+) + D_{p,f}(C^+, B^+)\}\\ &\le \max\{D_{p,f}(A^-, C^-), D_{p,f}(A^+, C^+)\} +\\ &\qquad \max\{D_{p,f}(C^-, B^-), D_{p,f}(C^+, B^+)\}\\ &= D_{p,f}^*(A, C) + D_{p,f}^*(C, B). \end{aligned}$$

Appendix C
Proof of Proposition 5

The $\lambda-$level sets of triangular fuzzy numbers A and B can be expressed as

$$\begin{aligned} A_\lambda &= [a_1 + \lambda(a - a_1), a_2 - \lambda(a_2 - a)], \\ B_\lambda &= [b_1 + \lambda(b - b_1), b_2 - \lambda(b_2 - b)]. \end{aligned} \tag{30}$$

According to Eq. (5), we have

$$\begin{aligned} D_2^2(A, B) &= \int_0^1 \lambda[(a_1 - b_1) + \lambda((a - a_1) - (b - b_1))]^2 d\lambda \\ &\quad + \int_0^1 \lambda[(a_2 - b_2) - \lambda((a_2 - a) - (b_2 - b))]^2 d\lambda \\ &= \frac{(a-b)^2}{2} + \frac{1}{12}[(a_2 - b_2)^2 + (a_1 - b_1)^2] + \\ &\quad \frac{1}{6}(a - b)[(a_2 - b_2) + (a_1 - b_1)], \end{aligned}$$

and the proof is complete.

References

[1] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, pp. 338–353, 1965.

[2] K. Atanassov, *Intuitionistic Fuzzy Sets, Theory and Applications*. New York: Physica-Verlag, 1999.

[3] B. Gorzafczany, “Approximate inference with interval-valued fuzzy sets-an outline,” in *Proc. of the Polish Symp. on Interval and Fuzzy Math*, Poznan, Poland, 1983, pp. 89–95.

[4] I. B. Turksen, “Interval valued fuzzy sets based on normal forms,” *Fuzzy Sets and Systems*, vol. 20, no. 2, pp. 191–210, 1986.

[5] G. J. Wang and Y. Y. He, “Intuitionistic fuzzy sets and l-fuzzy sets,” *Fuzzy Sets and Systems*, vol. 110, no. 2, pp. 271–274, 2000.

[6] G. Deschrijver and E. E. Kerre, “On the relationship between some extensions of fuzzy set theory,” *Fuzzy Sets and Systems*, vol. 133, no. 2, pp. 227–235, 2003.

[7] ——, “On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision,” *Information Sciences*, vol. 177, no. 8, pp. 1860–1866, 2007.

[8] M. B. Gorzalczany, “A method of inference in approximate reasoning based on interval-valued fuzzy sets,” *Fuzzy Sets and Seystems*, vol. 21, pp. 1–17, 1987.

[9] G. Wang and X. Li, “The applications of interval-valued fuzzy numbers and interval-distribution numbers,” *Fuzzy Sets and Systems*, vol. 98, no. 3, pp. 331–335, 1998.

[10] ——, “Correlation and information energy of interval-valued fuzzy numbers,” *Fuzzy Sets and Systems*, vol. 103, no. 1, pp. 169–175, 1999.

[11] D. H. Hong and S. Lee, “Some algebraic properties and a distance measure for interval-valued fuzzy numbers,” *Information Sciences*, vol. 148, no. 1–4, pp. 1–10, 2002.

[12] P. Grzegorzewski, “Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on the hausdorff metric,” *Fuzzy Sets and Systems*, vol. 148, no. 2, pp. 319–328, 2004.

[13] Y. M. Wang, J. B. Yang, D. L. Xu, and K. S. Chin, “On the combination and normalization of interval-valued belief structures,” *Information Sciences*, vol. 177, pp. 189–200, 2007.

[14] G. Deschrijver, “Arithmetic operators in interval-valued fuzzy set theory,” *Information Sciences*, vol. 177, no. 14, pp. 2906–2924, 2007.

[15] J. H. Chen and S. M. Chen, “A new method to measure the similarity between interval-valued fuzzy numbers,” in *Proc. of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong*, vol. 3, 2007, pp. 1403–1408.

[16] L. Chen, “Distances between interval-valued fuzzy sets,” in *Proc. of the 28th IEEE North American Fuzzy Information Processing Society Annual Conference (NAFIPS2009), Cincinnati, Ohio, USA*, 2009, pp. 1–3.

[17] J. Chachi and S. M. Taheri, “A unified approach to similarity measures between intuitionistic fuzzy sets,” *International Journal of Intelligent Systems*, 2013.

[18] L. H. Chen and L. Y. Ouyang, “Fuzzy inventory model for deteriorating items with permissible delay in payment,” *Applied Mathematics and Computation*, vol. 182, no. 1, pp. 711 – 726, 2006.

[19] C. C. Yao and P. T. Yu, “Fuzzy regression based on asymmetric support vector machines,” *Applied Mathematics and Computation*, vol. 182, no. 1, pp. 175 – 193, 2006.

[20] A. A. Ramli, J. Watada, and W. Pedrycz, “Real-time fuzzy regression analysis: A convex hull approach,” *European Journal of Operational Research*, vol. 210, no. 3, pp. 606 – 617, 2011.

[21] M. Kelkinnama and S. M. Taheri, “Fuzzy least-absolutes regression using shape preserving operations,” *Information Sciences*, vol. 214, pp. 105 – 120, 2012.

[22] O. Kocadagli, “A novel nonlinear programming approach for estimating capm beta of an asset using fuzzy regression,” *Expert Systems with Applications*, vol. 40, no. 3, pp. 858–865, 2013.

[23] L. Zhixin and J. Hongmei, “Effectiveness and relevancy measures under modal cardinality for interval-valued fuzzy sets,” in *Proc. of the 3rd IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE), Chengdu, China*, vol. 1, 2010, pp. 400–402.

[24] S. J. Chen and S. M. Chen, “Fuzzy risk analysis based on measures of similarity between interval-valued fuzzy numbers,” *Computers and Mathematics with Applications*, vol. 55, no. 8, pp. 1670–1685, 2008.

[25] R. Xu and C. Li, “Multidimensional least-squares fitting with a fuzzy model,” *Fuzzy Sets and Systems*, vol. 119, no. 2, pp. 215–223, 2001.

[26] R. Grzymkowski, A. Kapusta, T. Kuboszek, and D. Slota, *Mathematica 6*. Jacka Skalmierskiego, 2008.

[27] A. L. Page *et al.*, *Methods of soil analysis. Part 2. Chemical and microbiological properties.* American Society of Agronomy, Soil Science Society of America, 1982.

[28] J. Mohammadi and S. M. Taheri, “Pedomodels fitting with fuzzy least squares regression,” *Iranian Journal of Fuzzy Systems*, vol. 1, no. 2, pp. 45–61, 2004.

[29] R. W. Miller, R. L. Donahue *et al.*, *Soils: an introduction to soils and plant growth.* Prentice-Hall International Inc., 1990, no. Ed. 6.

[30] R. Moore, *Methods and Applications of Interval Analysis*. Philadelphia: SIAM, 1979.

[31] R. Xu, “A linear regression model in fuzzy environment,” *Advance Modelling Simulation*, vol. 27, pp. 31–40, 1991.

TABLE I. OBSERVED AND PREDICTED INTERVAL-VALUED FUZZY VALUES OF SAR AND ESP AND THEIR DISTANCES

No.	SAR (x, s_x^-, s_x^+)	ESP (y, s_y^-, s_y^+)	Predicted ESP (Y, s_Y^-, s_Y^+)	$D_{f,p}^*$	D_p^*	D_∞^*
1	(0.78,0.05,0.08)	(3.08,0.43,0.57)	(6.20,0.35,0.58)	3.12	3.16	3.20
2	(0.64,0.14,0.15)	(2.86,0.16,0.34)	(5.24,0.98,1.04)	2.40	2.80	3.20
3	(0.62,0.06,0.14)	(6.25,0.18,0.27)	(5.10,0.38,0.96)	1.18	1.51	1.85
4	(0.49,0.04,0.06)	(4.11,0.16,0.26)	(4.21,0.29,0.40)	0.11	0.17	0.24
5	(1.10,0.07,0.08)	(1.04,0.32,0.41)	(8.40,0.50,0.54)	7.36	7.45	7.54
6	(0.61,0.08,0.08)	(2.71,0.37,0.57)	(5.03,0.55,0.57)	2.32	2.41	2.51
7	(0.74,0.07,0.09)	(4.45,0.53,0.60)	(5.93,0.51,0.61)	1.48	1.48	1.49
8	(1.15,0.07,0.15)	(6.92,0.18,0.59)	(8.75,0.47,1.07)	1.84	2.07	2.30
9	(1.08,0.12,0.13)	(7.41,0.37,0.60)	(8.26,0.84,0.93)	0.88	1.10	1.32
10	(0.38,0.07,0.13)	(9.08,0.32,0.51)	(3.45,0.51,0.87)	5.63	5.81	5.99
11	(0.61,0.05,0.06)	(6.56,0.18,0.32)	(5.03,0.33,0.43)	1.53	1.60	1.67
12	(0.98,0.10,0.10)	(5.05,0.33,0.61)	(7.58,0.66,0.68)	2.53	2.70	2.86
13	(0.71,0.04,0.07)	(5.23,0.16,0.58)	(5.72,0.30,0.45)	0.49	0.56	0.63
14	(0.50,0.05,0.07)	(5.16,0.47,0.51)	(4.27,0.35,0.47)	0.89	0.95	1.00
15	(0.77,0.12,0.13)	(11.10,0.19,0.22)	(6.13,0.85,0.92)	4.98	5.32	5.67
16	(0.99,0.11,0.13)	(4.47,0.23,0.34)	(7.65,0.77,0.88)	3.18	3.45	3.71
17	(3.56,0.10,0.12)	(28.84,0.24,0.41)	(25.33,0.71,0.84)	3.52	3.76	3.99
18	(0.86,0.12,0.15)	(9.43,0.40,0.52)	(6.75,0.82,1.05)	2.69	2.95	3.21
19	(0.61,0.07,0.13)	(4.50,0.24,0.55)	(5.03,0.48,0.92)	0.55	0.72	0.89
20	(0.64,0.05,0.05)	(9.30,0.50,0.51)	(5.24,0.32,0.36)	4.06	4.15	4.24
21	(0.71,0.15,0.15)	(9.48,0.41,0.57)	(5.72,1.01,1.06)	3.77	4.07	4.37
22	(0.61,0.10,0.12)	(3.65,0.22,0.38)	(5.03,0.67,0.81)	1.39	1.61	1.83
23	(0.63,0.04,0.13)	(10.14,0.46,0.49)	(5.17,0.30,0.91)	4.97	5.18	5.39
24	(1.13,0.06,0.11)	(3.00,0.33,0.57)	(8.61,0.39,0.74)	5.61	5.70	5.78
		Mean of distances		1.54	1.55	1.57

TABLE II. OBSERVED AND PREDICTED INTERVAL-VALUED FUZZY VALUES OF SAND, OM AND CEC AND THEIR DISTANCES

No.	OM $(x_1, s_{x_1}^-, s_{x_1}^+)$	SAND $(x_2, s_{x_2}^-, s_{x_2}^+)$	CEC (y, s_y^-, s_y^+)	Predicted CEC (Y, s_Y^-, s_Y^+)	$D_{f,p}^*$	D_p^*	D_∞^*
1	(0.88,0.03,0.11)	(35,1.72,3.55)	(16.5,0.89,2.19)	(16.28,0.46,1.10)	0.50	0.83	1.31
2	(1.13,0.09,0.15)	(37,0.51,5.25)	(18.6,1.80,2.17)	(16.47,0.35,1.56)	2.21	2.88	3.57
3	(1.31,0.11,0.16)	(27,0.84,3.39)	(19.3,0.76,2.38)	(19.21,0.48,1.18)	0.50	0.77	1.29
4	(1.98,0.16,0.25)	(29,2.32,4.23)	(20.3,1.03,2.79)	(20.47,0.95,1.61)	0.51	0.83	1.34
5	(1.02,0.07,0.14)	(38,2.46,3.92)	(17.3,0.25,2.56)	(15.96,0.73,1.24)	1.44	2.03	2.66
6	(1.29,0.04,0.18)	(32,0.27,3.72)	(20.4,1.56,2.96)	(18.02,0.16,1.31)	2.47	3.24	4.03
7	(1.52,0.13,0.17)	(29,1.09,3.47)	(19.3,1.40,2.59)	(19.29,0.58,1.23)	0.56	0.80	1.37
8	(1.33,0.06,0.16)	(18,0.29,2.08)	(21.9,1.62,2.82)	(21.30,0.23,0.89)	0.99	1.66	2.53
9	(1.71,0.05,0.24)	(40,3.44,5.36)	(15.9,1.53,1.64)	(17.28,0.90,1.83)	1.40	1.70	2.01
10	(2.00,0.07,0.24)	(28,0.29,2.84)	(18.3,1.55,1.88)	(20.75,0.24,1.25)	2.51	3.13	3.76
11	(1.68,0.15,0.17)	(13,0.48,1.92)	(22.6,1.62,2.98)	(23.34,0.49,0.88)	1.13	1.89	2.84
12	(2.15,0.18,0.30)	(19,0.27,1.90)	(23.7,2.28,2.88)	(23.18,0.54,1.21)	0.88	1.48	2.26
13	(3.52,0.21,0.40)	(31,1.64,4.13)	(24.4,0.34,2.96)	(23.98,0.92,1.96)	0.58	0.96	1.41
14	(2.33,0.20,0.33)	(31,1.88,4.08)	(21.8,1.49,3.05)	(20.92,0.93,1.77)	1.02	1.56	2.16
15	(1.71,0.16,0.19)	(17,1.20,2.24)	(23.8,1.45,2.61)	(22.51,0.68,0.99)	1.45	2.15	2.91
16	(1.14,0.03,0.11)	(14,0.04,1.94)	(20.8,1.92,2.31)	(21.72,0.10,0.74)	1.18	1.91	2.74
17	(0.99,0.09,0.10)	(19,1.08,1.96)	(17.5,0.02,2.58)	(20.20,0.48,0.71)	2.81	3.67	4.57
18	(1.14,0.02,0.16)	(28,0.33,3.02)	(17.8,1.12,2.50)	(18.54,0.13,1.11)	0.93	1.49	2.13
19	(1.46,0.09,0.20)	(26,2.21,2.66)	(20.2,0.73,2.13)	(19.82,0.74,1.12)	0.56	0.93	1.39
20	(1.81,0.06,0.23)	(32,1.47,3.76)	(20.0,1.13,2.63)	(19.36,0.49,1.44)	0.80	1.28	1.83
21	(1.38,0.07,0.14)	(10,0.50,1.39)	(22.8,1.39,2.28)	(23.25,0.30,0.68)	0.79	1.33	2.05
22	(0.84,0.07,0.11)	(38,2.91,4.18)	(19.1,1.60,2.12)	(15.50,0.84,1.24)	3.62	4.05	4.48
23	(1.48,0.07,0.16)	(49,0.96,6.47)	(12.1,1.09,1.73)	(14.65,0.40,1.89)	2.57	2.90	3.24
24	(1.08,0.04,0.16)	(42,1.14,5.52)	(12.8,0.88,1.90)	(15.21,0.37,1.65)	2.42	2.67	2.93
			Mean of distances		1.41	1.92	2.53