Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2018) 42: 402-411
© TÜBİTAK
doi:10.3906/tar-1709-57

A comparative study of fuzzy linear regression and multiple linear regression in agricultural studies: a case study of lentil yield management

Karim SORKHEH[1,*], **Ahmad KAZEMIFARD**[2], **Shakiba RAJABPOOR**[3]

[1]Department of Agronomy and Plant Breeding, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[2]Department of Mathematics, Faculty of Mathematics and Computer Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[3]Department of Biology, Payme-Noor University, Tehran, Iran

Received: 15.09.2017 • **Accepted/Published Online:** 07.05.2018 • **Final Version:** 11.12.2018

Abstract: This study investigates the advantages of two fuzzy linear regression (FLR) models, namely the Tanaka and the Savic and Pedrycz models, over multiple linear regression (MLR) for lentil yield management. We used a fuzzy approach to model the yield of lentil genotypes in which the input is crisp and the output fuzzy. Moreover, after finding FLR equations, we estimated the output corresponding to the collection of fuzzy inputs by using fuzzy algebraic operations and an appropriate defuzzification method known as the center of area method. Results showed the superiority of the Tanaka model over MLR because of reducing the included variables, especially variables available after harvest. The study also emphasizes the advantage of the Savic and Pedrycz model in comparison to the other two models with a smaller error rate.

Key words: Fuzzy linear regression, multiple linear regression, *Lens culinaris*

1. Introduction

For millennia, human beings have been aware of the importance of agricultural products in their life and have benefited from plants and their different parts like fruits, leaves, and seeds, for food, clothing, medicine, and animal feed (Ercisli, 2009; Erturk et al., 2010; Canan et al., 2016; Hricova et al., 2016; Yazici and Sahin, 2016).

From among crop plants, lentil (*Lens culinaris*) from the legume family is an important source of protein, fiber, energy, and minerals for both humans and animals (Torres et al., 2016). Based on FAO statistics (www.fao.org), Iran ranks as the 11th largest producer of lentil. Total production of lentil in Iran was 334,000 t (Ministry of Jihad-e-Agriculture of Iran, 2014).

One of the important areas of research in agriculture is yield management. In this area, several investigations have been carried out, such as those of Kravchenko and Bullock (2000), Kitchen et al. (2003), Jiang and Thelen (2004), Park et al. (2005), and Singh et al. (2013). In these studies, different approaches like crop models, statistical tools, and algorithms have been used to evaluate yield. Multiple linear regression (MLR) has been widely used to predict yield and determine factors influencing yield (Kravchenko and Bullock, 2000; Park et al., 2005; Huang et al., 2010). According to Kitchen et al. (2003), MLR fails to properly describe the relationship between the relevant parameters and variables if they are not linear and hence the results may not be trusted. Jiang and Thelen (2004), Huang et al. (2010), and Fortin et al.(2010) combined multivariate techniques, like principal component analysis and factor analysis, with multiple regressions to reduce the problems and to facilitate selecting a set of variables from a large data set. Some studies also applied artificial intelligence in yield management (Huang et al., 2010).

Regression analysis is one of the standard tools in analyzing data. The obtained mathematical equation can explain the relationship between the dependent and independent variables. Its explanatory power lies in its multivariate nature. It is available in computer packages and is widely used in different fields (Agresti, 1996). As is known, there are two types of regression, linear and nonlinear. Rousseeuw et al. (2004) listed some difficulties arising from nonlinear approaches that may lead to inconsistent and biased estimation.

Although nowadays, some researchers are interested in nonlinear modeling statistics rather than linear modeling, it is highly desirable to have a linear relationship between a dependent variable (such as yield) and independent variables affecting it, because linear relations are of a simple nature and mathematically are easier to work

* Correspondence: karimsorkheh@gmail.com

with. However, in practice and in many cases, such a relationship is nonlinear. If we look at the relationship between a dependent variable and independent variables from a fuzzy point of view, it is possible to consider some nonlinear relations as “fuzzy linear relations” in classical mathematics.

Keeping the above point in mind, the aim of the present study was to compare the fuzzy linear regression (FLR) model and MLR model in evaluating the relationship of lentil yield and independent variables. We will also show how the fuzzy approach can be employed in working with imprecise data. The FLR method was used for yield estimation and its inference capabilities were compared with the MLR tool.

2. Materials and methods

2.1. Data collection

Data were collected from kernel yield and its components in rain-fed lentil from a field experiment conducted in the 2010 and 2014 growing seasons at different sites of an experimental research field in Ahvaz, Iran. Five rows per meter width were planted and 8 seeds per row meter were placed (40 seeds/m^2), and fertilizers and herbicides were applied manually in planting and stem elongation.

Monthly minimum and maximum air temperatures (°C) and rainfall (mm) at the experimental site are illustrated in Table 1. Yield and yield components were recorded at maturity.

We studied and analyzed 25 predictor variables to detect the yield of lentil genotypes. After data analysis, we found that only 12 of those 25 variables were of more significance ($P < 0.05$) for lentil yield estimation. Statistical descriptions of studied variables are shown in Table 2. They are as follows: hundred-kernel weight (HKW), pod number (PN), kernel number per pod (KNPP), branch number (BN), leaf area index (LAI), length of the internodes (LI), plant height (H), harvest index (HI), biological yield (BY), days to flowering (DF), total dry matter (TDM), and kernel yield (KY), which were measured at the dry seed stage. Kernel yield was measured using samples of 5 m^2 randomly cut from each plot. Days to flowering were calculated based on days after sowing.

Following Yuan et al. (2017), leaf area was measured using a leaf area meter (Delta-T Device, UK) on four plants when one open flower at any node emerged. When one mature pod emerged at every node, shoot length, node number, and internode lengths were recorded. By using node number divided by shoot length, internode length was calculated. It should be mentioned that sowing time to harvest corresponded to the regular growing period for field-grown lentil in the experimental site. Then harvest index was calculated.

2.2. Methodology

To analyze the data, the statistical software SPSS 20, Excel 2010, and Lingo ver. 5 were used. Lingo was used for model optimization in order to obtain fuzzy linear regression equations. For every variable, whenever necessary, close data were organized as fuzzy numbers. To find MLR equations, SPSS 20 was used.

2.2.1. Multiple linear regression (MLR)

Regression analysis was first introduced by Galton in the 19th century. He developed a mathematical description in which regression describes statistical relations between variables (Kutner, 2004):

$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j \quad X_{ij} + \varepsilon_i(\beta) =$
$\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_k X_{ik} + \varepsilon_i(\beta), i = 1, \dots, n$.
Or $Y = X\beta + \varepsilon$.

The function for the least squares method is:

$$S(\beta_0, \beta_1, \beta_2, \dots, \beta_k) = S(\beta) = \sum_{j=1}^{d} \varepsilon_j^2 \; or \; \varepsilon^T \varepsilon.$$

From Eq. (1), $\varepsilon(\beta) = Y - X\beta$.
Then, $S(\beta) = (Y - X\beta)^T (Y - X\beta)$
$= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$.

To minimize $S(\beta)$, we have to differentiate $S(\beta)$ with respect to β where $\frac{\delta S}{\delta \beta} | \beta$ is equal to 0:

$$\frac{\delta S}{\delta \beta} | \beta = -2X^T Y + 2X^T X\beta = 0.$$

Hence, the least square estimator is:

$$\beta^{\wedge} = (X^T X)^{-1} X^T Y.$$

The value fit by the equation $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots \beta_k X_{ik}$ is denoted i, and the residuals ε_i are equal to $y_i - y^{\wedge}$, the difference between the observed and fitted values.

2.2.2. Fuzzy theory

The concept of fuzziness was first introduced by Zadeh (1965, 1975a, 1975b, 1975c). According to him, fuzzy theory emerged after it was found that traditional techniques of systems analysis are incapable of dealing with problems in which the relationships between variables are too vague or complex. Such problems are common in many fields such as biology, economics, social sciences, linguistics, and of course agriculture studies. According to Zadeh, “a common thread that runs through problems of this type is the un-sharpness of class boundaries and the concomitant imprecision, uncertainty, and partiality of truth” (Bojadziev, 2007).

Fuzzy theory is a framework within which the imprecision, vagueness, and uncertainty of the real world are modeled (Zadeh, 1975a). In classical set theory, membership in a set is binary. An element either belongs or does not belong to the set. However, membership in fuzzy sets is gradual, not binary. Elements in fuzzy sets have degrees of membership. The cornerstone of fuzzy theory is the concept of fuzzy sets.

Fuzzy sets are a generalization of classical sets, in which the membership of every element can be considered a number in interval [0,1] instead of 0 or 1. The membership function of a fuzzy set $\tilde{A}$, $\underset{\tilde{A}}{\mu}(x)$, specifies the grade or degree to which any element x belongs to the fuzzy set $\tilde{A}$. We will identify any fuzzy set with its membership function and use these two concepts interchangeably.

A fuzzy subset $\tilde{A}$ of the real numbers $\mathbb{R}$ is said to be convex if for all $x, y \in \mathbb{R}$ and for every real number λ satisfying $0 \leq \lambda \leq 1$, we have

$$\mu_{\tilde{A}}(\lambda x + (1-\lambda)y) \geq min\left\{\underset{\tilde{A}}{\mu}(x), \underset{\tilde{A}}{\mu}(y)\right\}.$$

Now a fuzzy number $\tilde{A}$ is defined on the universe $\mathbb{R}$ a convex and normalized fuzzy set with semicontinuous membership function $\underset{\tilde{A}}{\mu}(x)$ in which $\tilde{A}$ is called normalized when at least one $x \in \mathbb{R}$ attains the maximum membership grade 1. We denote the set of all fuzzy numbers by $\tilde{\mathbb{R}}$.

Since in the real world, data sometimes cannot be precisely recorded or collected, fuzzy theory is naturally an appropriate tool in modeling when fuzzy data have been observed. In this regard, the concept of fuzzy numbers is vital. In particular, let r be a real number. Then the trapezoidal fuzzy set $\tilde{r}$ defined by:

$$\mu_{\tilde{r}}(x) = \begin{cases} \frac{x-a_1}{r-\varepsilon_1-a_1} & a_1 \leq x \leq r-\varepsilon_1 \\ 1 & r-\varepsilon_1 \leq x \leq r+\varepsilon_2 \\ \frac{a_2-x}{a_2-r-\varepsilon_2} & p+\varepsilon_2 \leq x \leq a_2 \\ 0 & o.w \end{cases}.$$

This is a fuzzy number, which can be expressed as being about r or approximately equal to r, and it can be denoted by $\tilde{r} = (a_1, r-\varepsilon_1, r+\varepsilon_2, a_2)$. If $\varepsilon_1 = \varepsilon_2 = 0$ then the trapezoidal fuzzy number is called a triangular fuzzy number and it can be written in the form of (a_1, r, a_2) instead of (a_1, r, r, a_2).

In addition, a triangular fuzzy number (a_1, a_2, a_3) is symmetrical if

$$a_2 = \frac{a_1+a_3}{2}.$$

In such a case it is customary that the symmetrical fuzzy number $(a-\varepsilon\ , a, a+\varepsilon\ )$ be written in the form $(a, \varepsilon\ )$. Triangular fuzzy numbers are very often used in the applications. Note that every real number a can be expressed as the triangular fuzzy number $a = (a, a, a)$. In this way, every real number can be regarded as a fuzzy number, i.e. $\mathbb{R} \subset \tilde{\mathbb{R}}$.

Let:
$L(x)$ $(and\ R(x))$: $[0,\infty) \rightarrow [0,1]$ be decreasing, shape function with :

- $L(0) = 1;\ L(x) < 1$ for all $x > 0$,
- $L(x) > 0$ for all $x \in [0,1)$,
- $L(1) = 0\ or\ \ L(x) > 0$ for all x and $\lim_{x\to\infty} L(x) = 0$.

Then a fuzzy number $\tilde{A}$ is called L-R type if for $m, \alpha \geq 0, \beta \geq 0$ in $\mathbb{R}$ we have the following:

$$\underset{\tilde{A}}{\mu}(x) = \begin{cases} L(\frac{m-x}{\alpha}) & x \leq m \\ R(\frac{x-m}{\beta}) & x \geq m \\ 0 & o.w \end{cases}$$

Here, m is called the mode of $\tilde{A}$, and α, β are called the left and right spreads, respectively, and we denote the set of all L-R fuzzy numbers by $\tilde{\mathbb{R}}_{L-R}$. In this notation, $\tilde{A}$ can be shown as $(m, \alpha, \beta)_{L-R}$.

A particular case is when $\tilde{A} = (a, b, c) = (b, b-a, \mathrm{c}-\mathrm{b})_{L-R}$, a triangular fuzzy number. Here
$L(x) = R(x) = \max\{0,$

$$1-|x|\ \}= \begin{cases} 1-|x| & -1 \leq x \leq 1 \\ 0 & o.w \end{cases}$$

Therefore,

$$\underset{\tilde{A}}{\mu}(x) = \begin{cases} L(^{b-x}/_{b-a}) & a \leq x \leq b \\ R(^{x-b}/_{c-b}) & b \leq x \leq c \\ 0 & o.w \end{cases}$$

Let $\odot$ be any binary operation between fuzzy numbers $\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_n$. Then the membership function of $\tilde{A}_1 \odot \tilde{A}_2 \odot \ldots \odot \tilde{A}_n$ is defined by:

$$\mu_{\tilde{A}_1 \odot \tilde{A}_2 \odot \ldots \odot \tilde{A}_n}(z) = \sup \min\left\{\mu_{\tilde{A}_1}(a_1), \mu_{\tilde{A}_2}(a_2), \ldots, \mu_{\tilde{A}_n}(a_n) \,|\, a_1 \odot a_2 \odot \ldots \odot a_n = z\right\}.$$

Using the extension principle, in particular if

$$\tilde{A} = (m, \alpha, \beta)_{L-R}, \tilde{A}' = (m', \alpha', \beta')_{L-R}\quad then$$

$$\tilde{A} \oplus \tilde{A}' = (m+m', \alpha+\alpha', \beta+\beta')_{L-R}$$

and

$$\tilde{A} \otimes \tilde{A}' = (mm', m\alpha' + m'\alpha, m\beta' + m'\beta)_{L-R}, \tilde{A}, \tilde{A}' \geq 0$$

$\tilde{A} \otimes \tilde{A}' = (mm', m'\alpha - m\beta', m'\beta - m\alpha')_{L-R}$, $\tilde{A} \leq 0, \tilde{A}' \geq 0$,

$\tilde{A} \otimes \tilde{A}' = (mm', -m'\beta - m\beta', -m'\alpha - m\alpha')_{L-R}$, $\tilde{A}, \tilde{A}' \leq 0$ (Wu, 2003).

2.2.3. Defuzzification

Defuzzification is the process of mapping a fuzzy set onto a crisp value. A number of defuzzification methods have been developed and the most widely used one is the center of area method (CAM) (Ni, 2005), as follows:

Consider $\underset{\tilde{A}}{\mu}(x)$, the membership function of a fuzzy set $\tilde{A}$. The defuzzified value CAM, $\underset{CAM}{d}$, is defined as the first coordinate of the gravity center of the area under the curve of $\underset{\tilde{A}}{\mu}(x)$ (Ni, 2005). Indeed:

$$\underset{CAM}{d} = \frac{\int_{-\infty}^{+\infty} x \underset{\tilde{A}}{\mu}(x)dx}{\int_{-\infty}^{+\infty} \underset{\tilde{A}}{\mu}(x)dx}.$$

2.2.4. Fuzzy linear regression (FLR)

Statistical regression has many applications, but it causes serious problems if the data are too small, the relationship between variables is unclear or the verification of the normal distribution of error is difficult, there is ambiguity in the event or the linearity is an inappropriate assumption, or some of the values of independent or dependent variables are fuzzy. These are the situations addressed by FLR (Shapiro, 2005).

The FLR model was first introduced by Tanaka et al. (1982) by using linear programming to determine the regression coefficients as fuzzy numbers. FLR provides tools to study the relationship between variables when some of the assumptions of MLR fail.

After Tanaka et al., FLR has been studied extensively by many authors (e.g., Bardossy, 1990; Savic and Pedrycz, 1991; Ishibuchi, 1992; Chang et al., 1994; Peters, 1994; Redden and Woodal, 1994; Ayyub et al., 1997; Diamond et al., 1997; Mann et al., 2011; Taheri and Kelkinnama, 2012; Torres et al., 2016). There are two models to develop:

1) Models in which the relationship of the variables is fuzzy;

2) Models in which the variables are fuzzy (Shapiro, 2005).

In this article, to model the yield of lentil genotypes, we used a MLR model and two FLR models, one introduced by Tanaka et al. (1982) and the other by Savic and Pedrycz (1991). It should be mentioned that these two FLR models are both developed based on a minimum fuzziness method.

According to both, we set

$$\tilde{y} = \tilde{A}_0 \oplus \sum_{j=1}^{k} \tilde{A}_j \otimes X_j,$$

in which $\tilde{A}_0, \tilde{A}_1, \dots, \tilde{A}_k$ are symmetrically triangular fuzzy numbers.

Our aim is to determine these fuzzy numbers. For this, we must solve the following linear programming problem:

Min $\beta = mc_0 + \sum_{j=1}^{k} \sum_{j=1}^{k} c_j\, x_{ij}$

s.t.:

$p_0 + \sum_{i=1}^{k} p_i\, x_{ij} - (1-h)\left(c_0 + \sum_{i=1}^{k} c_i\, x_{ij}\right) \leq y_i - (1-h)e_i$

$p_0 + \sum_{i=1}^{k} p_i\, x_{ij} + (1-h)\left(c_0 + \sum_{i=1}^{k} c_i\, x_{ij}\right) \geq y_i + (1-h)e_i$

$$c_0, c_1, \dots, c_k \geq 0$$

where m is the size of the data set, k is the number of independent variables, $\{x_{i1}, x_{i2}, \dots, x_{ik}\}$ denotes the *i*th observation, $\{c_0, c_1, \dots, c_k\}$ denotes the half spreads of fuzzy regression coefficients, $\{p_0, p_1, \dots, p_k\}$ denotes the centers of fuzzy regression coefficients, and h is a possibility level predetermined by a decision-maker (Figure 1) (Tanaka et al., 1982). However, it is necessary to mention that there exists a subtle difference between the Savic and Pedrycz model and the Tanaka model.

In the Savic and Pedrycz FLR model (1991), which was developed by integrating minimum fuzziness into MLR, the centers of fuzzy regression coefficients are exactly those coefficients that appear in the MLR model, whereas in the Tanaka model, $\{p_0, p_1, \dots, p_k\}$ and $\{c_0, c_1, \dots, c_k\}$ are calculated independently of MLR.

3. Results and discussion

According to our data, the estimated MLR model, after defuzzification of the dependent variable (using CAM) for lentil yield, is as follows:

Y = −1063.421 + 2.2X_1(=HKW) + 10.105X_2(=PN) + 1.024X_3(=KNPP) −0.125X_4(=BN) + 4.175X_5(=LAI) + 5.746X_6(=LI) − 0.717X_7(=H) + 24.680X_8(=HI) + 0.689X_9(=BY) − 0.130X_{10}(=DF) − 2.367X_{11}(=KY) + 3.350X_{12}(=TDM).

Now we will apply 2 models of FLR to our data, the Tanaka and the Savic and Pedrycz, respectively.

Applying the Tanaka model with $h = 0.5$ leads to the following results:

$$\tilde{y} = \tilde{A}_0 \oplus \sum_{i=1}^{12} \tilde{A}_i \otimes X_i,$$

where:

$\tilde{A}_i = (p_i, c_i) = (0,0)$, i = 0, 1, 4, 7, 11, 12,

and

$\tilde{A}_1 = (p_1, c_1) = (7.895, 0)$,

$\tilde{A}_3 = (p_3, c_3) = (4.33, 0)$,

$\tilde{A}_5 = (p_5, c_5) = (1.61, 0)$,

$\tilde{A}_6 = (p_6, c_6) = (3.04, 11.47)$,
$\tilde{A}_8 = (p_8, c_8) = (0.68, 0)$,
$\tilde{A}_9 = (p_9, c_9) = (1.9, 3.1)$,
$\tilde{A}_{10} = (p_{10}, c_{10}) = (11.56, 0)$.
In fact the only two fuzzy coefficients (Figure 2) are

$$\widetilde{A_6} = (3.04,11.47),\ \underset{\widetilde{A_6}}{\mu}(x) = \begin{cases} \dfrac{x+8.43}{11.47} & -8.43 \le x \le 3.04 \\ \dfrac{14.51-x}{11.47} & 3.04 \le x \le 14.51 \\ 0 & o.w \end{cases}$$

and

$$\widetilde{A_9} = (1.9,3.1),\ \underset{\widetilde{A_9}}{\mu}(x) = \begin{cases} \dfrac{x+1.2}{3.1} & -2.1 \le x \le 1.9 \\ \dfrac{5-x}{3.1} & 1.9 \le x \le 5 \\ 0 & o.w \end{cases}$$

We have:
$\tilde{Y} = 7.895 \otimes PN \oplus 4.33 \otimes KNPP \oplus 1.61 \otimes LAI \oplus (3.04, 11.47) \otimes LI \oplus 0.68 \otimes HI \oplus (1.9,3.1) \otimes BY \oplus 11.56 \otimes DF$.

On the other hand, applying the Savic and Pedrycz model to our data collection, we conclude that:

$$\tilde{y} = \tilde{A}_0 \oplus \textstyle\sum_{i=1}^{12} \tilde{A}_i \otimes X_i,$$

where

$$\widetilde{A_0} = (-1063.421,20.34),$$

$$\underset{\widetilde{A_0}}{\mu}(x) = \begin{cases} \dfrac{x+1063.421}{20.34} & -1063.421 \le x \le -1043.081 \\ \dfrac{-1022.741-x}{20.34} & -1043.081 \le x \le -1022.741 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_1} = (2.2,0.45),\ \underset{\widetilde{A_1}}{\mu}(x) = \begin{cases} \dfrac{x-1.75}{0.45} & 1.75 \le x \le 2.2 \\ \dfrac{2.65-x}{0.45} & 2.2 \le x \le 2.65 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_2} = (10.105,11.382),\ \underset{\widetilde{A_2}}{\mu}(x) = \begin{cases} \dfrac{x+1.275}{11.382} & -1.275 \le x \le 10.105 \\ \dfrac{21.487-x}{11.382} & 10.105 \le x \le 21.487 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_3} = (1.024,0.231),\ \underset{\widetilde{A_3}}{\mu}(x) = \begin{cases} \dfrac{x+1.024}{0.231} & -1.024 \le x \le -0.793 \\ \dfrac{-0.562-x}{0.231} & -0.793 \le x \le -0.562 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_4} = (-0.125,0.05),\ \underset{\widetilde{A_4}}{\mu}(x) = \begin{cases} \dfrac{x+0.130}{0.05} & -0.130 \le x \le -0.125 \\ \dfrac{-0.120-x}{0.05} & -0.125 \le x \le -0.120 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_5} = (4.175,0.21),\ \underset{\widetilde{A_5}}{\mu}(x) = \begin{cases} \dfrac{x-3.965}{0.21} & 3.965 \le x \le 4.175 \\ \dfrac{4.385-x}{0.21} & 4.175 \le x \le 4.385 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_6} = (5.746,6.201),\ \underset{\widetilde{A_6}}{\mu}(x) = \begin{cases} \dfrac{x+0.455}{6.201} & -0.455 \le x \le 5.746 \\ \dfrac{11.947-x}{6.201} & 5.746 \le x \le 11.947 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_7} = (-0.717,0.132),\ \mu_{\widetilde{A_7}}(x) = \begin{cases} \dfrac{x+0.849}{0.132} & -0.849 \le x \le -0.717 \\ \dfrac{-0.585-x}{0.132} & -0.717 \le x \le -0.585 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_8} = (24.680,5.169),\ \mu_{\widetilde{A_8}}(x) = \begin{cases} \dfrac{x-19.511}{5.169} & 19.511 \le x \le 24.680 \\ \dfrac{29.849-x}{5.169} & 24.680 \le x \le 29.849 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_9} = (0.689,0.123),\ \mu_{\widetilde{A_9}}(x) = \begin{cases} \dfrac{x-0.566}{0.123} & 0.566 \le x \le 0.689 \\ \dfrac{0.812-x}{0.123} & 0.689 \le x \le 0.812 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_{10}} = (-0.130,0.023),\ \mu_{\widetilde{A_{10}}}(x) = \begin{cases} \dfrac{x+0.153}{0.023} & -0.153 \le x \le -0.130 \\ \dfrac{-0.107-x}{0.023} & -0.130 \le x \le -0.107 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_{11}} = (-2.367,0.521),\ \mu_{\widetilde{A_{11}}}(x) = \begin{cases} \dfrac{x+2.897}{0.521} & -2.897 \le x \le -2.376 \\ \dfrac{-1.855-x}{0.521} & -2.376 \le x \le -1.855 \\ 0 & o.w \end{cases}$$

$$\widetilde{A_{12}} = (3.350,0.428),\ \mu_{\widetilde{A_{12}}}(x) = \begin{cases} \dfrac{x-2.922}{0.428} & 2.922 \le x \le 3.350 \\ \dfrac{3.778-x}{0.428} & 3.350 \le x \le 3.778 \\ 0 & o.w \end{cases}$$

Hence, we have:

$$\tilde{Y} = (-1063.421,20.34) \oplus (2.2,0.45) \otimes \mathrm{HKW} \oplus (10.105,11.382) \otimes PN \oplus (1.024,0.231) \otimes KNPP \oplus (-0.125,0.05) \otimes BN \oplus (4.175,0.21) \otimes LAI \oplus (5.746,6.201) \otimes LI \oplus (-0.717,0.132) \otimes H \oplus (24.680,5.169) \otimes HI \oplus (0.689,0.123) \otimes BY \oplus (-0.130,0.023) \otimes DF \oplus (-2.367,0.521) \otimes KY \oplus (3.350,0.428) \otimes TDM.$$

In Table 3, the root mean square errors (RMSEs) of these 3 models and the number of variables involved in each model are displayed. Moreover, the number of fuzzy coefficients that appear in every equation are given (Table 3).

According to Table 3, every FLR model has an advantage in comparison to MLR. Tanaka proved to be the best model based on the number of involved variables, which leads to the facility and easiness of calculations. Savic and Pedrycz was the best model regarding RMSE, which is the smallest value of error.

Now, in the following example, we see how we can estimate lentil yield corresponding to a collection of data using FLR and algebraic operations on arbitrary fuzzy numbers:

Example: Let

$\widetilde{HKW}=(22.8,2.3,2.2.1)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{PN}=(3.6,1.4,1.8)_{L-R}$, $L(x) = R(x) = \frac{1}{1+|x|}$,
$\widetilde{KNPP}=(1.8,.6,.6)_{L-R}$, $L(x) = R(x) = e^{-x^2}$,
$\widetilde{BN}=(151,8,10)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{LAI}=(124,5,6)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{LI}=(5,2,3)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{H}=(22,4,6)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{HI}=(36.74,5.6,4.8)_{L-R}$, $L(x)=\frac{1}{1+x^2}$, $R(x) = \frac{1}{1+2|x|}$,
$\widetilde{BY}=(73.45,5.8,4.2)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$,
$\widetilde{DF}=(22.37,3.4,4.1)_{L-R}$, $L(x) = R(x) = \frac{1}{1+|x|}$,
$\widetilde{KY}=(21.8,3.5,3.1)_{L-R}$, $L(x) = R(x) = e^{-x^2}$,
$\widetilde{TDM}=(45,6,8)_{L-R}$, $L(x) = R(x) = \max\{0, 1-|x|\}$.

Then, using the CAM defuzzification method, we conclude that:

$HKW \approx 21.73$, $PN \approx 4.93$, $KNPP \approx 1.8$, $BN \approx 151.66$, $LAI \approx 124.33$, $LI \approx 5.33$, $H \approx 22.33$, $HI \approx 36.15$, $BY \approx 72.92$, $DF \approx 20.36$, $KY \approx 21.63$, $TDM \approx 45.66$.

Now, by using the fuzzy regression equation, we have:

$$\tilde{Y} = (-1063.421,20.34) \oplus (2.2,0.45) \otimes 21.73 \oplus (10.105,11.382) \otimes 4.93 \oplus (1.024,0.231) \otimes 1.8 \oplus (-0.125,0.05) \otimes 151.66 \oplus (4.175,0.21) \otimes 124.33 \oplus (5.746,6.201) \otimes 5.33 \oplus (-0.717,0.132) \otimes 22.33 \oplus (24.68,5.169) \otimes 36.15 \oplus (0.689,0.123) \otimes 72.92 \oplus$$

Table 1. Some weather parameters during the study at the experimental site.

Month	Rainfall (mm)	Tmax (°C)	Tmin (°C)	Mean temperature (°C)	Number of days under 0 °C	Relative humidity (%)
September	0.00	33.40	1.80	18.90	0.00	26.00
October	69.60	27.20	1.20	13.00	0.00	60.00
November	71.60	20.80	–4.60	8.30	6.00	65.00
December	70.80	16.60	–7.00	4.10	22.00	63.00
January	40.80	18.80	–7.60	5.90	15.00	61.00
February	68.40	22.40	–2.00	10.60	3.00	61.00
March	86.90	29.80	–1.40	13.00	1.00	57.00
April	22.80	33.60	5.80	19.50	0.00	48.00
May	2.00	36.60	9.60	23.40	0.00	35.00

Table 2. Evaluation of quantitative traits assessed in study

Traits[a]	Min	Max	Mean	Coefficient variation
HKW (g)	25.09	43.06	33.26	8.28
PN	36.16	64.60	48.54	3.65
KNPP	20.00	43.40	31.83	5.56
BN	1.00	4.50	2.70	8.70
LAI (mm^2)	22.20	31.50	26.95	8.65
LI (mm)	10.25	35.65	21.70	16.25
H (cm)	35.56	57.60	48.09	6.18
HI (%)	34.44	58.40	46.82	11.55
BY (kg ha^{-1})	1441.8	2650.44	2060.25	8.61
DF (days)	25.00	43.40	31.82	5.65
TDM (g)	118.6	300.87	173.82	13.10
KY (kg ha^{-1})	527.64	1391.16	969.26	16.08

[a] Hundred-kernel weight (HKW), pod number (PN), kernel number per pod (KNPP), branch number (BN), leaf area index (LAI), length of the internodes (LI), plant height (H), harvest index (HI), biological yield (BY), days to flowering (DF), total dry matter (TDM), and kernel yield (KY).

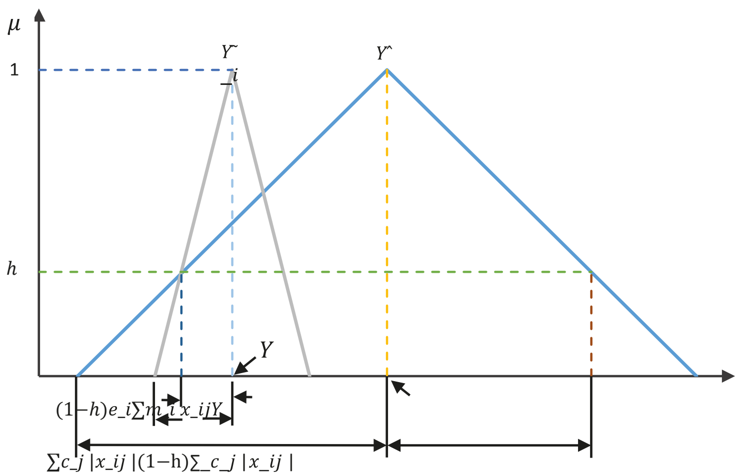

Figure 1. Membership functions for $\tilde{y}$ and $\hat{\tilde{y}}$.

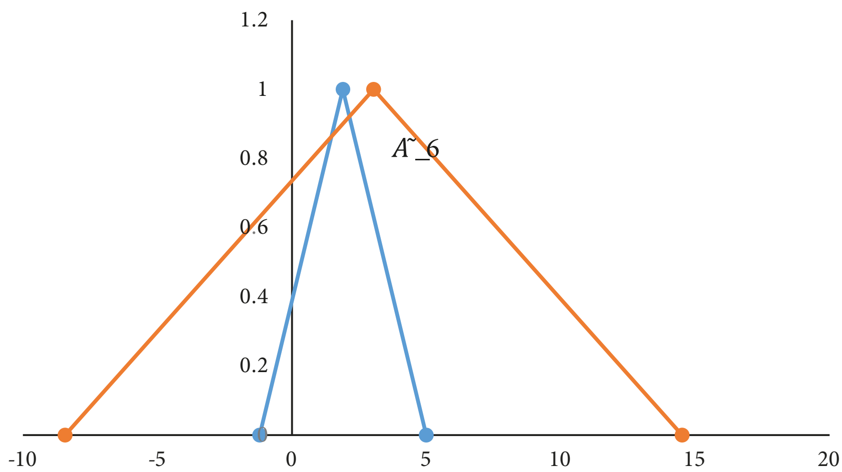

Figure 2. Fuzzy coefficients for Tanaka model.

Table 3. A summary of models.

Models of linear regression	RMSE	Number of involved variables
MLR	43.703	12
FLR (Tanaka)	68.216	7
FLR (Savic and Pedrycz)	40.928	12

$(-0.130,0.023)\otimes 20.36 \oplus$
$(-2.367,0.521)\otimes 21.63 \oplus$
$(3.350,0.428)\otimes 45.66 = (592.441,191.663)_{L-R}$, in which $L(x) = R(x) = \max\{0, 1 - |x|\}$.
Finally, by using the CAM method, we conclude that $\tilde{Y} \approx 592.441$ (kg ha^{-1}).

In this paper, our purpose was to emphasize the advantages of FLR over MLR in lentil yield management. Findings showed that FLR has more advantages than MLR. By using the Tanaka model instead of MLR, and with the fuzzification of only 2 coefficients, 5 variables from among 12 variables and the constant coefficient were put aside from the regression equation. The Tanaka model facilitates calculations and creates a more simple relationship between independent and dependent variables. In addition, 4 out of 5 variables that are not necessary to be included in calculations are those that are available only after harvest, namely HKW, BN, TDM, and KY, and interestingly the two other variables are small. This is evidence that the Tanaka FLR model outperforms MLR in our case study with simplicity, easiness, and reduction of included variables. In addition to enjoying the general advantages of other FLR models, the Savic and Pedrycz model proved to have the smallest value of RMSE. In the given example, we showed how to

Finally, even if we prefer MLR over FLR, if even one variable, be it input or output, is not crisp, the fuzzy approach should be undertaken for defuzzification since MLR is only proper for crisp data. Likewise, in our case study, we had to defuzzify yield in order to find the MLR equation. This confirms the importance and capability of fuzzy approaches in such studies.

In conclusion, this paper aimed at a comparison of FLR and MLR in lentil yield management. Based on our results, first the present study showed that a fuzzy approach including FLR is a suitable framework in lentil yield management. MLR is inefficient if some of its preconditions are not satisfied (see Section 2.2.4), as in the case of our data, in which some of the variables were fuzzy numbers. Even if MLR is applicable, the fuzzy approach has some advantages. For instance, when the input is huge, it is possible to categorize the close data in some limited fuzzy numbers. Second, as we have seen in our example, using FLR, the set of outputs corresponding to a collection of data can be estimated as a fuzzy number. We also found that although it indicates a bigger value of error, the Tanaka model outperforms MLR in lentil yield estimation approach fuzzy data. Here the important point is that if we have a collection of close data related to a variable, instead of involving every individual member of this set in the calculation, we can calculate the yield once by using a fuzzy approach. For this, first we describe such a collection in the form of a fuzzy number. Then we use CAM to obtain a crisp value from this fuzzy number, and then we continue the calculation with this number, which represents the collection. For instance, in our example, the values of BN were a collection of numbers close to 151 describable as a triangular fuzzy number $\widetilde{\mathrm{BN}}= (151,8,10)_{\mathrm{L-R}}$ in which $\mathrm{L(x)} = \mathrm{R(x)} = \max\{0, 1, |x|\}$, which here means a fuzzy number "around 151" from which the number 151.66 is obtained after defuzzification. For other variables of the example, a similar process is done.

based on the number of variables appearing in the equation, since it has reduced the involved variables and has facilitated the calculations. Finally, the Savic and Pedrycz model is better than MLR with regard to RMSE.

Acknowledgment

The authors wish to acknowledge Shahid Chamran University of Ahvaz, Iran, for providing financial support for research and development in the Faculty of Agriculture (Grant No. SHU/FAG/1396) and Faculty of Mathematics and Computer Science (Grant No. SHU/FMCS/1396).

References

Agresti A (1996). An Introduction to Categorical Data Analysis. New York, NY, USA: John Wiley & Sons, Inc.

Ayyub BM, Gupta MM (1997). Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach. Dordrecht, the Netherlands: Kluwer Academic.

Bardossy A (1990). Note on fuzzy regression. Fuzzy Set Syst 5: 65-75.

Bojadziev G (2007). Fuzzy Logic for Business, Finance, and Management (Vol. 23). Singapore: World Scientific.

Canan İ, Gündoğdu M, Seday U, Oluk CA, Karaşahin Z, Eroğlu EÇ, Yazıcı E, Ünlü M (2016). Determination of antioxidant, total phenolic, total carotenoid, lycopene, ascorbic acid, and sugar contents of *Citrus* species and mandarin hybrids. Turk J Agric For 40: 894-899.

Chang PT, Lee ES (1994). Fuzzy linear regression with spreads unrestricted in sign. Comp Math Appl 28: 61-70.

Diamond P, Korner R (1997). Extended fuzzy linear models and least squares estimates. Comput Math Appl 33: 15-32.

Ercisli S (2009). Apricot culture in Turkey. Sci Res Essays 4: 715-719.

Erturk Y, Ercisli S, Haznedar A, Cakmakci R (2010). Effects of plant growth promoting rhizobacteria (PGPR) on rooting and root growth of kiwifruit (*Actinidia deliciosa*) stem cuttings. Biol Res 43: 91-98.

Fortin JG, Anctil F, Parent LÉ, Bolinder MA (2010). A neural network experiment on the site-specific simulation of potato tuber growth in Eastern Canada. Comput Electron Agr 73: 126-132.

Hricova A, Fejer J, Libiakova G, Szabova M, Gazo J, Gajdosova A (2016). Characterization of phenotypic and nutritional properties of valuable *Amaranthus cruentus* L. mutants. Turk J Agric For 40: 761-771.

Huang Y, Lan Y, Thomson SJ, Fang A, Hoffmann WC, Lacey RE (2010). Development of soft computing and applications in agricultural and biological engineering. Comput Electron Agr 71: 107-127.

Ishibuchi H (1992). Fuzzy regression analysis. Fuzzy Theo Syst 4: 137-148.

Jiang P, Thelen KD (2004). Effect of soil and topographic properties on crop yield in a north-central corn-soybean cropping system. Agro J 96: 252-258.

Kitchen NR, Drummond ST, Lund ED, Sudduth KA, Buchleiter GW (2003). Soil electrical conductivity and topography related to yield for three contrasting soil-crop systems. Agro J 95: 483-495.

Kravchenko AN, Bullock DG (2000). Correlation of corn and soybean grain yield with topography and soil properties. Agro J 92: 75-83.

Kutner M, Nachtsheim C, Neter J, Li W (2004). Applied Linear Statistical Models. 5th ed. New York, NY, USA: McGraw-Hill/Irwin.

Mann KK, Schumann AW, Obreza TA (2011). Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data. Precis Agric 12: 457-472.

Ministry of Jihad-e-Agriculture of Iran (2014). Annual Agricultural Statistics. Tehran, Iran: Ministry of Jihad-e-Agriculture of Iran.

Ni Y (2005). Fuzzy Correlation and Regression Analysis. Norman, OK, USA: University of Oklahoma Graduate College.

Park SJ, Hwang CS, Vlek PLG (2005). Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agri Syst 85: 59-81.

Peters G (1994). Fuzzy linear regression with fuzzy intervals. Fuzzy Sets Syst 63: 35-43.

Redden DT, Woodal WH (1994). Properties of certain fuzzy linear regression methods. Fuzzy sets Syst 64: 361-375.

Rousseeuw PJ, Van Alest S, Van Driessen K, Agulló J (2004). Robust multivariate regression. Technometrics 46: 293-305.

Savic D, Pedrycz W (1991). Evaluation of fuzzy regression models. Fuzzy Sets Syst 39: 51-63.

Shapiro AF (2005). Fuzzy Regression Models. State College, PA, USA: Pennsylvania State University.

Singh AK, Meena MK, Bharati RC, Gade RM (2013). Effect of sulphur and zinc management on yield, nutrient uptake, changes in soil fertility and economics in rice (*Oryza sativa*)–lentil (*Lens culinaris*) cropping system. Indian J Agri Sci 83: 344-348.

Taheri SM, Kelkinnama M (2012). Fuzzy linear regression based on least absolutes deviations. Iranian J Fuzzy Syst 9: 121-140.

Tanaka H, Uejima S, Asai K (1982). Linear regression analysis with fuzzy model. IEEE T Syst Man Cyb 12: 903-907.

Torres J, Rutherfurd SM, Muñoz LS, Peters M, Montoya CA (2016). The impact of heating and soaking on the in vitro enzymatic hydrolysis of protein varies in different species of tropical legumes. Food Chem 194: 377-382.

Wu HC (2003). Linear regression analysis for fuzzy input and output data using the extension principle. Comput Math Appl 45: 1849-1859.

Yazıcı K, Şahin A (2016). Characterization of pomegranate (*Punica granatum* L.) hybrids and their potential use in further breeding. Turk J Agric For 40: 813-824.

Yuan HY, Saha S, Vandenberg A, Bett KB (2017). Flowering and growth responses of cultivated lentil and wild lens germplasm toward the differences in red to far-red ratio and photosynthetically active radiation. Front Plant Sci 8: 386.

Zadeh LA (1965). Fuzzy sets. Information Control 8: 338-353.

Zadeh LA (1975a). The concept of linguistic variables and its application to approximate reasoning I. Inform Sciences 8: 199-249.

Zadeh LA (1975b). The concept of linguistic variables and its application to approximate reasoning II. Inform Sciences 8: 301-357.

Zadeh LA (1975c). The concept of linguistic variables and its application to approximate reasoning III. Inform Sciences 9: 43-80.