A forward selection based fuzzy regression for new product development that correlates engineering characteristics with consumer preferences

Kit Yan Chan[a,*] and Sai Ho Ling [b]
[a]*Department of Electrical and Computer Engineering, Curtin University, Australia*
[b]*Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia*

Abstract. Fuzzy regression models have commonly been used to correlate engineering characteristics with consumer preferences regarding a new product. Based on the models, product developers can determine optimal engineering characteristics of the new product in order to satisfy consumer preferences. However, they have a common limitation in that they cannot guarantee to include significant regressors with significant engineering characteristics or significant nonlinear terms. The generalization capability of the model can be reduced, when too few significant regressors are included and too many insignificant regressors are included. In this paper, a forward selection based fuzzy regression (FS-FR) is proposed based on the statistical forward selection to determine significant regressors. After the significant regressors are determined, the fuzzy regression is used to generate the fuzzy coefficients which address the uncertainties due to fuzziness and randomness caused by consumer preference evaluations. The developed model includes only significant regressors which attempt to improve the generalization capability. A case study of a tea maker design demonstrated that the FS-FR was able to generate consumer preference models with better generalization capabilities than the other tested fuzzy regressions. Also simpler consumer preference models can be provided for the new product development.

Keywords: fuzzy regression, statistical forward selection, new product development, consumer preferences, engineering characteristics, tea maker design, overfitting, generalization capability

1. Introduction

For the past two decades, the role of product developers has been extended to include both sounding out consumers and determining the optimal engineering characteristics of a new product that will satisfy consumer preferences [1]. When the correlation between consumer preferences and engineering characteristics is developed, a successful new product is more likely to be produced as consumer requirements can be satisfied based on the truly specified engineering characteristics [2]. Based on quality function deployment (QFD) [1,3], these correlations can be represented by a matrix, namely houses of quality (HOQ), which relates consumer preferences to engineering characteristics. However, development of the HOQ associated with engineering characteristics is a complex decision-making process as nonlinearities exist between engineering characteristics and consumer preferences; moreover, an evaluation of the degrees of consumer preferences is fuzzy as it is normally accomplished in a subjective or heuristic manner. Therefore, the inclusion of correlated engineering characteristics cannot be guaranteed when the consumer preferences cannot be accurately determined. Alternatively, the empirical consumer preference models have commonly been developed in order to represent the relationship between consumer preferences and engineering characteristics [4]. These consumer preference models are developed using

*Corresponding author. Kit Yan Chan. E-mail: kit.chan@curtin.edu.au ; Telephone number is 61 8 92662945.

experimental or consumer survey data which illustrates correlations between engineering characteristics to consumer preferences.

Artificial neural networks can be used to develop models which correlate engineering characteristics to consumer preferences, when certain amount of consumer preference samples is given. However, artificial neural networks generally lack transparency and they are represented as a black-box model. Explicit information for consumer preferences cannot be indicated in the artificial neural networks. To generate explicit models, multivariate regression approaches are commonly used [5,6]. In general, statistical methods are more preferred as more explicit information can be found than using the artificial neural networks. Compared with artificial neural networks, more explicit information can be found in statistical regression models which are in a polynomial form. Hence, variable significances and variable interactions can be determined in the polynomial of the regression models. However, to develop these regression models, it is necessary to assume that deviations between collected samples and model estimates are randomly distributed. As deviations between collected samples and model estimates can be caused by indefinite knowledge or imprecise evaluations of consumer preferences, errors in the models can be fuzzy since the models are unable to capture the fuzziness of consumer preferences. To address the fuzziness of customers' perceptual evaluations on a product, the approaches of fuzzy modelling [7,8,9] have been commonly developed. However, these approaches can generate only implicit consumer preference models of which explicit information cannot be indicated. These methods are not widely utilized by product developers, as they reveal no explicit reasons for new product development. Analytical information such as significances, sensitivities, and interactions for engineering characteristics cannot be illustrated on the models.

Fuzzy regressions are commonly used for developing consumer preference models, as explicit and analytical information can be indicated within the model [10]; and also fuzziness caused by subjective evaluations of consumer preferences can be addressed by the model [11,12]. Different versions of fuzzy regression have been developed for new product development. Kim et al. [11] and Sekkeli et al. [12] have applied fuzzy linear regression to generate consumer preference models, whereby fuzzy coefficients of the models consist of symmetric triangular memberships. To increase the flexibility for addressing fuzziness, advanced versions of fuzzy linear regressions were developed based on asymmetric triangular fuzzy memberships [13-16]. To address uncertainties caused by fuzziness and randomness in consumer preference observations, fuzzy linear least-squares regression has been developed [17]. More recently, the chaos optimization method [18] has been integrated with fuzzy regression [19], whereby the chaos optimization method is used to generate the nonlinear polynomial structure of the model. Based on the polynomial structure, fuzzy coefficients are generated in order to address nonlinearities for consumer preferences.

Although all these fuzzy regression methods are able to address partially the characteristics and fuzziness of consumer preferences, these approaches have a common limitation: they cannot guarantee the model consists only of significant engineering characteristics and significant nonlinear terms; and they also cannot guarantee insignificant engineering characteristics can be excluded in the model. Using all regressors to develop a model is not effective for the development of the models [20]. Including too many insignificant regressors could reduce the model generalization capabilities and may cause an overfitted model [21]. Excluding significant one could lead to the model that cannot learn significant patterns for consumer preferences. Therefore, it is necessary to determine significant regressors which consist only of significant engineering characteristics and significant nonlinear terms. Based on the model consists of significant regressors, strong correlation between engineering characteristics and consumer preferences can be developed. Accurate prediction of consumer preferences is more likely to be determined by the model.

Although the heuristic method namely genetic programming has been used to develop the models involving with significant regressors, the genetic programming requires a lot more computational model evaluations than the statistical regression methods do [22]. Therefore, in this paper, a novel fuzzy modelling method, namely forward selection based fuzzy regression (FS-FR) which incorporates the approaches of fuzzy least square regression [23] and forward selection regression [24], is proposed. The FS-FR attempts to develop the consumer preference model which consists only of significant regressors. In the FS-FR, the mechanism of forward selection regression [24] is used to identify the significant regressors, where the forward selection regression has commonly been used on system modelling such as manufacturing process systems [25], product development [26] and biomedical development [7] etc.

Also we have compared the genetic programming [22,27] and the proposed FS-FR based on a case study of new product development. The case study shows that better generalization capability can be obtained by the proposed FS-FR.

In the FS-FR, a model with no regressor is first initialized. The FS-FR then adds the significant regressors one by one to the model, until the model cannot be improved by adding one more regressor. Finally, a model which includes only the significant regressors is developed. Based on the selected regressors, the fuzzy coefficients are determined using the fuzzy least square regression [23] in order to address the uncertainties caused by the quantitative evaluations of consumer preferences [27]. Hence, the resulting model consists of significant regressors only, and is able to address nonlinearities and sample uncertainties caused by consumer preference evaluations. The effectiveness of the FS-FR is evaluated based on a case study of a tea maker design, as the consumer preferences for tea maker design are nonlinear. Also the collected samples of the consumer preferences are contaminated with uncertainties which are caused by fuzziness and randomness of the quantitative evaluations. The results obtained by the FS-FR are compared with those obtained by the commonly-used fuzzy regression methods and the recently developed fuzzy regression method for new product development [4]. Better generalization capabilities can be obtained by the models developed by the FS-FR compared with the tested methods. Also less numbers of engineering characteristics are involved on developed the models. Hence, more effective usage of engineering characteristics can be obtained by FS-FR.

2. Fuzzy regression for consumer preference models

In the new product development, the consumer preference model [4] described in (1) is essential to estimate the consumer preference, y:

$$y = f_{CPM}\left(x_1, x_2, ..., x_m\right), \tag{1}$$

where x_j is the *j*-th engineering characteristic with $j = 1,2,...,m$, and x_j is correlated to y; m is the number of engineering characteristics involved on the new product development; and f_{CPM} represents the functional relationship between x_j with $j = 1,2,...,m$ to y. f_{CPM} can be formulated by the fuzzy polynomial form in (2), in order to address nonlinearities and fuzziness on y.

$$\tilde{y} = \tilde{A}_0 + \tilde{A}_1 z_1 + \tilde{A}_2 z_2 + ... + \tilde{A}_k z_k + ... + \tilde{A}_{N_R} z_{N_R}, \tag{2}$$

where $\tilde{y}$ is the fuzzy estimate of the consumer preference and $\tilde{y}$ is denoted by the fuzzy number, $\tilde{y} = \left(\tilde{y}^C, \tilde{y}^R, \tilde{y}^L\right)$ with $\tilde{y}^C$ be the center, $\tilde{y}^R$ be the right spread and $\tilde{y}^L$ be the left spread. N_R is the number of regressors z_k with $k = 1,2,..., N_R$ in (2); z_k are aligned with the linear, interaction and the high-order terms of the engineering characteristic. z_k are represented as:

$z_1 = x_1$, $z_2 = x_2$,…, $z_m = x_m$, $z_{m+1} = x_1 \cdot x_1$,

$z_{m+2} = x_1 \cdot x_2$, $z_{m+3} = x_1 \cdot x_3$,…,

$z_{m^2} = x_m \cdot x_m$,…, $z_k = x_{I(1)} \cdot x_{I(2)} \cdot ... \cdot x_{I(d)}$ with

$I(1), I(2), ..., I(d) \in \{1,2,...,m\}$, $m^2 + 1 \le k < N_{NR}$

and $3 \le d < m$,…, and $z_{N_{NR}} = x_1 \cdot x_2 \cdot ... x_m$.

(2) can be rewritten as (3) by substituting with the centers, left spreads and right spreads for the fuzzy coefficients.

$$\tilde{y} = \left(a_0^C, a_0^L, a_0^R\right) + \left(a_1^C, a_1^L, a_1^R\right) z_1 + \left(a_2^C, a_2^L, a_2^R\right) z_2, \\ + ... + \left(a_k^C, a_k^L, a_k^R\right) z_k + ... + \left(a_{N_R}^C, a_{N_R}^L, a_{N_R}^R\right) z_{N_R}, \tag{3}$$

Including all z_k may not be effective to estimate $\tilde{y}$, as some z_k is not significant in contributing the estimate of $\tilde{y}$. The estimate may not be better than only including the significant z_k on the model, and also the model may be overfitted as some insignificant z_k are included. Therefore, (3) is reformulated by (4) which only includes the significant regressors.

$$\tilde{y} = \left(a_{I(1)}^C, a_{I(1)}^L, a_{I(1)}^R\right) z_{I(1)} + \left(a_{I(2)}^C, a_{I(2)}^L, a_{I(2)}^R\right) z_{I(2)} + \\ \left(a_{I(3)}^C, a_{I(3)}^L, a_{I(3)}^R\right) z_{I(3)} + ... + \\ \left(a_{I(N_{Term})}^C, a_{I(N_{Term})}^L, a_{I(N_{Term})}^R\right) z_{I(N_{Term})} \tag{4}$$

where N_{Term} is the number of significant regressors; and $I(j)$ is the index of significant regressor which is significantly correlated to $\tilde{y}$, and the indexes are given by:

$$I(j) \in \{0,1,2,..., N_R\}, \text{ but } I(j) \ne I(k) \text{ with } j \ne k \\ \text{and } j,k = 1,2,..., N_{Term}. \tag{5}$$

Section 3 proposes an approach, namely forward selection based fuzzy regression (FS-FR), which incorporates the mechanisms of fuzzy regression [23] and forward selection [24], in order to develop the model formulated in (5) of which the significant regressors are only included. Also, the consumer preference model attempts to address uncertainties caused by the quantitative evaluations and measures of consumer preferences. As (2) is involved with linear terms, linear correlation between consumer preference and engineering characteristics can be addressed by the linear terms, z_1 , z_2 ,…, and z_m . Also (2) is involved with high order and interaction terms, z_{m+1} , z_{m+2} ,…,and $z_{N_{NR}}$. Based on [21], those high order and interaction terms can be used to address system nonlinearity. Hence, the nonlinear correlation between consumer preference and engineering characteristics can be generated by the FS-FR.

3. Forward selection based fuzzy regression

3.1. Mechanism of regressor selection

The FS-FR is proposed to generate the fuzzy regression models which only consist of significant regressors. The FS-FR determines N_{Term} and $I(j) \in \{0,1,2,...,N_R\}$ with j =1, 2, …, N_{Term}, when the original data, namely D_o^i, is given; D_o^i consists of the samples which correlate the i-th consumer preference, $\tilde{y}_i$, and the m engineering characteristics. Here we consider the number of consumer preferences can be larger than or equal to one. Hence, $i \geq 1$. The FS-FR first uses D_o^i to generate the full data set for each regressor, namely D_F^i , which represents the relationship between the regressors, z_j with $j = 1,2,...,N_R$, and $\tilde{y}_i$. Based on D_F^i , the FS-FR uses the mechanism of forward selection [24] to identify the significant regressors. The FS-FR then develops an 'empty' consumer preference model with no regressor. It adds the significant regressors one-by-one to the consumer preference model, where the significances of the regressors are determined based on the hypothesis test with respect to $\tilde{y}_i$. The FS-FR stops adding regressors until the consumer preference model cannot significantly be improved by adding another regressor. When N_{Term}, and $I(j) \in \{0,1,2,...,N_R\}$ with j =1, 2, …, N_{Term} are determined, the fuzzy coefficients for the significant regressors, $z_{I(j)}$ with j =1, 2, …, N_{Term}, are generated based on the fuzzy least square regression [4,23] in order to address interaction and high-order terms for engineering characteristics, and uncertainties caused by the quantitative evaluations and measures of consumer preferences. The mechanism of the FS-FR is detailed as follows:

Step 0: Generate the full data set, D_F^i , based on the original data set,

$$D_o^i = \{\tilde{y}_i(k), x_1(k), x_2(k), ..., x_m(k) \mid k = 1,2,...,N_D\} \quad (6)$$

where N_D is the number of collected samples; $\tilde{y}_i(k) = \left(y_i^C(k), y_i^R(k), y_i^L(k)\right)$ is the k-th fuzzy sample with respect to the i-th consumer preference; and $x_j(k)$ is the k-th data with respect to the j-th engineering characteristic. D_F^i , is given as:

$$D_F^i = \left\{\left(\tilde{y}_i^C(k), \tilde{y}_i^R(k), \tilde{y}_i^L(k)\right), z_0(k), z_1(k), ..., z_{N_R}(k) \right| k = 1,2,...,N_D\} \quad (7)$$

where N_R is given by (4):

$z_1(k) = x_1(k)$, $z_2(k) = x_2(k)$,…, $z_m(k) = x_m(k)$;

$z_{m+1}(k) = x_1(k) \cdot x_1(k)$, $z_{m+2}(k) = x_1(k) \cdot x_2(k)$,…,

$z_{m+m^2}(k) = x_m(k) \cdot x_m(k)$; …,

$z_j(k) = x_{p(1)}(k) \cdot x_{p(2)}(k) \cdot ... \cdot x_{p(d)}(k)$ with $p(1), p(2), ..., p(d) \in \{1,2,...,m\}$, $3 \leq d < m$ and $\left(m^2 + m + 1\right) \leq j < N_{NR}$,…; and

$z_{N_{NR}}(k) = x_1(k) \cdot x_2(k) \cdot ... \cdot x_m(k)$.

Step 1: Perform the hypothesis test for H_0^1 and H_A^1 :

$$H_0^1: z_{I(k)} \text{ is insignificant to } \tilde{y}_i, \ \forall\, I(k) \in I_k^R \text{ with } k = 1 \quad (8)$$

$$H_A^1: z_{I(k)} \text{ is significant to } \tilde{y}_i, \ \exists\, I(k) \in I_k^R \text{ with } k = 1 \quad (9)$$

where I_k^R is a set of possible regressor indexes not being picked out before *Step k*; here $k = 1$ and hence I_1^R contains all regressors being considered and is given by $I_1^R = [0,1,2,...,N_R]$. The hypothesis test for H_0^k and H_A^k is discussed in Section 3.1.

If H_0^1 **is accepted**, the procedure of F-SR terminates and no regressor is correlated to $\tilde{y}_i$.

If H_A^1 **is accepted**, at least one regressor is correlated to $\tilde{y}_i$. Then, **Step 1.1** and **Step 1.2** are conducted:

Step 1.1: The index with the most significant regressor, $I(1)$, is picked out from I_k^R with k=1, where $I(1)$ is identified from the hypothesis test for H_0^k and H_A^k with k=1.

Step 1.2: Set k=2 and **Step k** is performed.

Step k: (with $k \geq 2$**):** Perform the hypothesis test for H_0^k and H_A^k:

$$H_0^k: z_{I(k)} \text{ is insignificant to } \left(\tilde{y}_i - \sum_{j=1}^{k-1} \tilde{B}_{I(j)}^i \cdot z_{I(j)}\right), \quad \forall\, I(k) \in I_k^R \tag{10}$$

$$H_A^k: z_{I(k)} \text{ is significant to } \left(\tilde{y}_i - \sum_{j=1}^{k-1} \tilde{B}_{I(j)}^i \cdot z_{I(j)}\right), \quad \exists\, I(k) \in I_k^R \tag{11}$$

where I_k^R is identical to I_{k-1}^R except that $I(k-1)$ is excluded from I_{k-1}^R, and I_k^R is given as:

$$I_k^R = \{I(k), I(k+1), ..., I(N_R)\}. \tag{12}$$

If H_0^k **is accepted**, no significant regressor can be found in I_k^R, $K = k-1$ is set, and the **Final Step** is performed.

If H_A^k **is accepted**, at least one regressor is significant to $\tilde{y}_i$. Then, **Step k.1**, and **Step k.2** are conducted:

Step k.1: The index with the most significant regressor, $I(k)$, is picked out from I_k^R where $I(k)$ is given by the hypothesis test for H_0^k and H_A^k.

Step k.2: Set k =k+1 and **Step k** is performed again.

Final Step: The F-SR returns the number of significant regressors as N_{Term}=k. Also, it returns the indexes of the significant regressors, $\{I(1), I(2), ..., I(N_{term})\}$, which can be used to develop the consumer preference model given in (14) based on the fuzzy least square regression in (Chan et al. 2014):

$$\tilde{y}_i = \left(b_{I(1)}^{i,C}, b_{I(1)}^{i,L}, b_{I(1)}^{i,R}\right) \cdot z_{I(1)} + \left(b_{I(2)}^{i,C}, b_{I(2)}^{i,L}, b_{I(2)}^{i,R}\right) \cdot z_{I(2)} + ... + \left(b_{I(N_{term})}^{i,C}, b_{I(N_{term})}^{i,L}, b_{I(N_{term})}^{i,R}\right) \cdot z_{I(N_{term})} \tag{13}$$

3.2. Hypothesis test for H_0^k and H_0^k

The hypothesis test for H_0^k and H_0^k can be performed, when D_F^i is given and the samples of each of the three numbers, $y_i^C(k)$, $y_i^R(k)$, and $y_i^L(k)$, with $k = 1, 2, ..., N_D$ are assumed to be distributed normally. Based on the regressor indexes, $I_k^R = \{I(1), I(2), ..., I(N_{term})\}$ given in (13), the consumer preference model represented by (14) can be developed by the fuzzy least square regression [2,23].

$$\tilde{y}_i = \left(b_{I(k)}^{i,C}, b_{I(k)}^{i,L}, b_{I(k)}^{i,R}\right) \cdot z_{I_k^R(k)} + \left(b_{I(k+1)}^{i,C}, b_{I(k+1)}^{i,L}, b_{I(k+1)}^{i,R}\right) \cdot z_{I(k+1)} + ... + \left(b_{I(N_R)}^{i,C}, b_{I(N_R)}^{i,L}, b_{I(N_R)}^{i,R}\right) \cdot z_{I(N_R)} \tag{14}$$

The significance of the $I(j)$-th regressor, $z_{I(j)}$ with $j = k, (k+1), ..., N_R$, in (14) is determined based on the t-values of the regressor with respect to the center, left spread and right spread respectively as given by (15a), (15b) and (15c) respectively, where t_j^C, t_j^L and t_j^R are used to address the significances of the center, left spread and right spread for the j-th regressor respectively. More discussions about the t-values can be found in [20]. The larger the three t-values, the more significant the regressor is.

$$t_j^C = \frac{b_{I(j)}^{i,C}}{C_j^C}, \text{ with } j = k, (k+1), ..., N_R \text{ and}$$

$$\mathrm{C_j^C} = \sqrt{\lambda_j \times \mathrm{s^C}}, \text{ if } \mathrm{s^C} \neq 0, \tag{15a}$$

$$t_j^L = \frac{b_{I(j)}^{i,L}}{C_j^L}, \text{ with } j = k, (k+1), ..., N_R \text{ and}$$

$$C_j^L = \sqrt{\lambda_j \cdot s^L}, \text{ if } s^L \neq 0, \tag{15b}$$

$$t_j^R = \frac{b_{I(j)}^{i,R}}{C_j^R}, \text{ with } j = k, (k+1), ..., N_R \text{ and}$$

$$C_j^R = \sqrt{\lambda_j \cdot s^R}, \text{ if } s^R \neq 0, \tag{15c}$$

where λ_j is the j-th diagonal element of $\left(\left(\mathbf{z}^D\right)^T \cdot \mathbf{z}^D\right)^{-1}$ and $\left(\mathbf{z}^D\right)^T$ is the transpose of $\mathbf{z}^D$; $\mathbf{z}^D$ is developed based on D_F^i of which only the data with the indexes, $I(k)$, $I(k+1)$, …, $I(N_R)$ are included and $\mathbf{z}^D$ is given as:

$$\mathbf{z}^D = \begin{pmatrix} z_{I(k)}(1) & z_{I(k+1)}(1) & \dots & z_{I(N_R)}(1) \\ z_{I(k)}(2) & z_{I(k+1)}(2) & \dots & z_{I(N_R)}(2) \\ \vdots & \vdots & \vdots & \vdots \\ z_{I(k)}(N_D) & z_{I(k+1)}(N_D) & \dots & z_{I(N_R)}(N_D) \end{pmatrix} \tag{16}$$

s^C, s^L and s^R in (15a), (15b) and (15c) are the hybrid errors of estimates with respect to center, left spread and right spread, and they are given in (17a), (17b) and (17c) respectively:

$$s^C = \frac{\left|\left(\left(\mathbf{y}_i^{D,C}\right)^T \cdot \mathbf{y}_i^{D,C}\right) - \left(\mathbf{b}^C \cdot \left(\mathbf{z}^D\right)^T \cdot \mathbf{y}_i^{D,C}\right)\right|}{d^f}, \tag{17a}$$

$$s^L = \frac{\left|\left(\left(\mathbf{y}_i^{D,L}\right)^T \cdot \mathbf{y}_i^{D,L}\right) - \left(\mathbf{b}^C \cdot \left(\mathbf{z}^D\right)^T \cdot \mathbf{y}_i^{D,L}\right)\right|}{d^f}, \tag{17b}$$

$$s^R = \frac{\left|\left(\left(\mathbf{y}_i^{D,R}\right)^T \cdot \mathbf{y}_i^{D,R}\right) - \left(\mathbf{b}^C \cdot \left(\mathbf{z}^D\right)^T \cdot \mathbf{y}_i^{D,R}\right)\right|}{d^f}. \tag{17c}$$

where d^f is the degree of freedom [20]. It is illustrated in the t-values formulated in (15a), (15b) and (15c) that t-values, t_j^C, t_j^L and t_j^R, are large when the corresponding j-th diagonal element, λ_j is small, and the hybrid errors of estimates, s^C, s^L and s^R, are small. When $s^R = s^L = s^C = 0$, all regressors with the index in I_k^R are significant to $\tilde{y}_i$;

$\mathbf{y}_i^{D,C}$, $\mathbf{y}_i^{D,L}$ and $\mathbf{y}_i^{D,R}$ in (18a), (18b) and (18c) are given by the centers, left spreads and right spreads of the consumer preferences describing in D_F^i. They are given respectively as:

$$\mathbf{y}_i^{D,C} = \left(y_i^C(1) \quad y_i^C(2) \quad \dots \quad y_i^C(N_D)\right)^T \tag{18a}$$

$$\mathbf{y}_i^{D,L} = \left(y_i^C(1) + y_i^L(1) \quad y_i^C(2) + y_i^L(2) \quad \dots y_i^C(N_D) + y_i^L(N_D)\right)^T \tag{18b}$$

$$\mathbf{y}_i^{D,R} = \left(y_i^C(1) + y_i^R(1) \quad y_i^C(2) + y_i^R(2) \quad \dots y_i^C(N_D) + y_i^R(N_D)\right)^T \tag{18c}$$

$\mathbf{b}^C$, $\mathbf{b}^L$ and $\mathbf{b}^R$ in (18a), (18b) and (18c) are given by the centers, left spreads and right spreads of the fuzzy coefficients. They are given respectively as:

$$\mathbf{b}^C = \left(b_{I(k)}^{i,C} \quad b_{I(k+1)}^{i,C} \quad \dots \quad b_{I(N_R)}^{i,C}\right)^T \tag{19a}$$

$$\mathbf{b}^L = \left(b_{I(k)}^{i,C} + b_{I(k)}^{i,L} \quad b_{I(k+1)}^{i,C} + b_{I(k+1)}^{i,L} \quad b_{I(N_R)}^{i,C} + b_{I(N_R)}^{i,L}\right)^T \tag{19b}$$

$$\mathbf{b}^R = \left(b_{I(k)}^{i,C} + b_{I(k)}^{i,R} \quad b_{I(k+1)}^{i,C} + b_{I(k+1)}^{i,R} \quad b_{I(N_R)}^{i,C} + b_{I(N_R)}^{i,R}\right)^T \tag{19c}$$

For all $m \in \{k, (k+1), ..., N_R\}$, if the t-values for the center, left spread and right spread, t_m^C, t_m^L and t_m^R, are larger than a threshold value, $T_{threshold}$, and t_m^C, t_m^L and t_m^R are the largest among all the others, i.e.:

$$t_m^C \ge t_j^C,\ t_m^L \ge t_j^L,\ t_m^R \ge t_j^R, \text{ and } t_m^C, t_m^L, t_m^R > T_{threshold},$$
$$\forall j \in \{k, (k+1), ..., N_R\} \text{ but } m \ne j, \tag{20}$$

then the hypothesis, H_A^k, is accepted. Hence, the hypothesis concerning the three fuzzy components including center, left spread and right spread at a time of which a confidence region is satisfactory. Then, the index of $I(m)$ is picked out from I_k^R and the index, $I(m)$, is returned as the solution of the hypothesis test, which is the most significant regressor among all the tested regressors. Otherwise, if all t_j^C, t_j^L and t_j^R are smaller than $T_{threshold}$, then the hypothesis, H_0^k, is accepted. Hence, no significant regressor can be found.

4. Evaluation of Forward selection based fuzzy regression

The effectiveness of the FS-FR is evaluated using the case study of a tea maker design, as the consumer preferences regarding the tea makers are nonlinear and also the collected samples of the consumer preferences are uncertainties caused by the fuzziness and randomness of the quantitative evaluations and measures. The effectiveness of the FS-FR is compared with that achieved by the commonly-used fuzzy regression methods for developing consumer preference models.

4.1. A Case study of tea maker design

The tea maker design is involved with, three consumer preferences, namely catechin content, y_1, concentration, y_2, and tea temperature, y_3. y_1 represents the amount of antioxidant which was found in great abundance in the tea leaves. Its health benefits have been under close examination. y_2 represents the three tea ratings in terms of aroma, texture and overall taste. y_3 represents tea temperature after the tea brewing process. The following five steps are involved for brewing a tea. Also, the steps are involved with the five engineering characteristics namely the reheating temperature, x_1, the number of dips in the first brewing, x_2, the dipping time, x_3, the number of dips in the second brewing, x_4, and the dipping time in the second brewing, x_5.

- **Step 1) Heating the fresh water:** Fresh water in the tea container is heat to 98 degrees Celsius.
- **Step 2) Loading the tea leaves and reheating the water:** 7 grams of tea leaves are loaded into the tea infuser and are placed into the tea maker container. The water is reheated to the reheating temperature x_1.
- **Step 3) First brewing:** The tea infuser is dropped into the water for x_2 times. In each drop, the tea infuser is dipped for 10 seconds and is elapsed for 10 seconds before the next drop.
- **Step 4) Tea dipping:** The tea infuser is immersed in the water for x_3 times.
- **Step 5) Second brewing:** The second brewing cycle is similar to the first brewing cycle. The tea infuser is dipped into the water for x_4 times. The dipping time is x_5.

Before developing the three consumer preference models for y_1, y_2, and y_3, experiments were conducted by setting different values for the five engineering characteristics, x_1, x_2, x_3, x_4, and x_5. In the experiments, the ranges of the engineering characteristics are given and are quantized into four levels as illustrated in Table 1. Table 2 show the 16 experimental configurations of the orthogonal array, $L_{16}(4^5)$, which studies the effects of x_1, x_2, x_3, x_4, and x_5. In order to study the fuzziness of evaluating the three consumer preferences, the experiments configured with $L_{16}(4^5)$ were repeated twice; the fuzzy observations for catechin content, tea concentration, and tea temperature are shown in Table 3. The details of the experiments can be referred to [2].

4.2. Experimental results and comparisons

The consumer preference models for y_1, y_2, and y_3 can be evaluated by investigating the mean absolute errors, which indicate the difference between the fuzzy observations and the fuzzy estimates of the models. The mean absolute errors are defined by e^i_{MAE}, with i=1, 2 and 3, where e^1_{MAE}, e^2_{MAE}, and e^3_{MAE} represent the errors for y_1, y_2, and y_3 respectively:

$$e^i_{MAE} = \frac{\sum_{k=1}^{N_D} \left| \tilde{y}_i^c\left(I(k)\right) - \tilde{f}_i*\left(x_1\left(I(k)\right),...,x_5\left(I(k)\right)\right) \right|}{N_D \cdot \tilde{y}_i*\left(I(k)\right)} \tag{21}$$

N_D is the number of observations to be investigated; $x_1\left(I(k)\right)$, $x_2\left(I(k)\right)$, $x_3\left(I(k)\right)$, $x_4\left(I(k)\right)$ and $x_5\left(I(k)\right)$ are parameter values for the $I(k)$-th experimental configuration of $L_{16}\left(4^5\right)$, where $I(k) \in \left[1,2,...,N_D\right]$, but $I(k) \neq I(j)$ with $k \neq j$ and $j,k = 1,2,...,N_D$. Here $N_D = 16$, as 16 experiments have been conducted; $\tilde{f}_i*\left(x_1\left(I(k)\right),...,x_5\left(I(k)\right)\right)$ and $\tilde{y}_i*\left(I(k)\right)$ are the crisp values for the $I(k)$-th fuzzy estimate and the fuzzy observation respectively. The crisp value, $\tilde{z}*$, is defuzzified from a fuzzy number, $\tilde{z} = \left(z^C, z^L, z^R\right)$, based on the weighted fuzzy arithmetic [28] which is particularly developed for defuzzification of triangular fuzzy numbers generated by fuzzy regressions. The weighted fizzy arithmetic integrates the fuzzy interval with respect to the membership level of the fuzzy number, and divides the integral of the membership function of the fuzzy number. It performs defuzzification of $\tilde{z}$ as:

$$\tilde{z}* = \frac{\int_0^1 z_h^L \cdot h \, dh + \int_0^1 z_h^R \cdot h \, dh}{\int_0^1 h \, dh} \tag{22}$$

where $\left[z_h^L, z_h^R\right] = \left[z^C - (1-h)z^L, z^C - (1-h)z^R\right]$ is the fuzzy interval with respect to $\tilde{z}$ and h is the membership level. By simplifying (22), defuzzification of $\tilde{z}*$ can be given as:

$$\tilde{z}* = z^C + \frac{1}{6}\cdot\left(z^R - z^L\right) \tag{23}$$

Using the 16 fuzzy observations shown in Tables 3, the consumer preference models for y_1, y_2, and y_3 are developed based on the proposed FS-FR and are given by (24), (25) and (26) respectively. The model for y_1 in fuzzy numbers is developed as:

$$\tilde{y}_1 = (1.568,\ 2.465, 2.472) + (0.17, 0.497, 0.496) \cdot x_1 \tag{24}$$

where the training error involved with the 16 fuzzy observations with respect to e^1_{MAE} is 4.034%. The model for y_2 in fuzzy numbers is developed as:

$$\tilde{y}_2 = (1.634, 1.567,\ 1.57) + (0.300, 0.139,\ 0.14) \cdot x_1 + (0.0872,\ 0.625,\ 0.626) \cdot x_1 \cdot x_2 \tag{25}$$

where training error is 2.382%. The model for y_3 in fuzzy numbers is developed as:

$$\tilde{y}_3 = (86.345, 35.45, 35.37) + (3.379, 9.788, 9.805) \cdot x_1 \tag{26}$$

where the training error is 2.382%. To evaluate the generalization capability of the models developed by the five methods, four commonly-used approaches for generating consumer preference models were employed to compare the results obtained by the proposed FS-FR:

a) **Tanaka and Watada's fuzzy regression** (TS-FR) [29] can generate models based on a small amount of samples. It has been used to generate the consumer preference models [14,15].

b) **Peters' fuzzy regression** (P-FR) [30] is a new version of T-FR, where the estimated interval on the generated model is bounded by all samples and the generated model is effective on detecting presence of outliers. P-FR has been used to develop consumer preference models for mobile phone design [22].

c) **Hybrid fuzzy least square regression** (H-FLSR) [2] can be used to address the uncertainties caused by fuzzy and random natures of the samples. H-FLSR was used to develop consumer preference models for packing machines [17].

d) **Genetic programming-based fuzzy regression** (GP- FR) [22] can be used to generate fuzzy nonlinear regression models. In the GP-FR, the polynomial structures are generated by genetic programming, where polynomial structures are involved with the linear, high order and interactions terms for the dependent variables. The polynomial structures attempts to represent the nonlinear correlation between dependent and independent variables. The fuzzy coefficients of the polynomial structures are determined by TS-FR [29]. The GP-FR was implemented based on the routines of the GP Matlab package [31] which is available for the public. As GP-FR is a stochastic method, 31 runs were performed on GP-FR and the median results were used as the comparison. Madar et al. [31] showed that the GP parameters and mechanisms given in Table 4 are able to find good solutions for various problems. Hence, these parameters and mechanisms are used in the GP-FR. The detailed operations of the GP-FR can be referred to [22].

Table 4
GP parameters setting of the GP-FR

Population size	100
Generation number	500
Crossover operation	One-point crossover with two parents
Mutation operation	Point mutation
Crossover rate	0.5
Mutation rate	0.5

Cross-validations based on the 16 experimental samples given in Tables 5, 6 and 7 are conducted in order to evaluate the generalization capability of the proposed FS-FR, and the other four algorithms (i.e. TS-FR, P-FR, H-FLSR and GP-FR). The cross-validation uses two samples from the whole sampling set as the validation samples, and the remaining samples as the training samples for the algorithms. The pair sets for the two samples are given in the first column of the tables. For example, the first validation used the 1st and 2nd samples as the validation samples, and it used the rest of the samples, 3rd to 16th samples, as the training samples. The second validation used the 2nd and 3rd samples as the validation samples, and it used the 1st, 4th to 16th samples as the training samples. The cross validations were repeated until all pair sets were used as the validation samples. For modelling the Catechin content, the cross-validation results are shown in Table 5 which summarizes the generalization errors of the five methods, and the mean generalization error of each method. Also, the ranks of each method are shown in the table. It indicates that the generalization errors obtained by the proposed FS-FR are generally smaller than those of the other four tested methods, TS-FR, Peters-FR, H-FLSR and GP-FR. Also, the rank of the proposed FS-FR is the first. Similar cross-validation results are shown in Table 6 and Table 7 regarding the tea concentration and the tea temperature respectively. The ranks of the proposed FS-FR for both consumer preferences are both the first. The generalization errors obtained by the proposed FS-FR are generally smaller than the other four tested methods. Therefore, better generalization capability can be achieved by the proposed FS-FR in modelling the consumer preferences. Better generalization capability can be explained by the mechanisms of the proposed FS-FR of which only significant engineering characteristics are included in the models. The other four tested methods may include insignificant engineering characteristics and may exclude the significant ones from the models.

Table 5
Generalization errors (%) for Catechin content obtained by the tested methods

Cross validation number	FS-FR (%)	TS-FR (%)	Peters-FR (%)	H-FLSR (%)	GP-FR (%)
1,2	6.796	13.605	16.094	3.898	6.170
2,3	6.011	5.256	13.808	2.788	5.676
3,4	12.024	8.141	14.233	3.203	7.959
4,5	5.308	8.888	5.197	6.772	5.167
5,6	3.258	9.418	7.107	6.078	8.744
6,7	5.852	3.892	3.577	1.608	7.842
7,8	5.501	4.622	8.426	6.485	6.455
8,9	2.170	8.559	3.130	9.392	4.102
9,10	4.897	8.869	2.426	3.347	5.494
10,11	3.403	0.8617	3.728	2.569	4.994
11,12	1.271	3.342	7.448	9.814	2.838
12,13	1.214	7.370	9.603	9.153	5.347
13,14	9.666	5.362	27.027	3.008	7.941
14,15	7.407	9.738	10.624	7.240	9.394
15,16	6.795	5.737	26.927	10.785	7.668
16,1	6.476	7.385	7.998	6.534	7.481
Mean errors	**5.503**	**6.940**	**10.46**	**5.792**	**6.454**
Rank	**1**	**4**	**5**	**2**	**3**

Figure 1 shows the relative improvements when each of the four tested methods, TS-FR, Peters-FR, H-FLSR and GP-FR, is compared with the proposed FS-FR, where the relative improvement (Rel. imp.) is the difference between the mean error obtained by the proposed FS-FR (Err. FS-FR) and the one obtained by the other tested method (Err. Other algor.). The relative improvement is given by the following formulation:

$$\text{Rel. imp.} = \frac{100\% \times \left((\text{Err. FS-FR}) - (\text{Err. other algor.}) \right)}{(\text{Err. other algor.})}$$

For the three consumer preferences, the proposed FS-FR obtained the improvements with more than 15% relatively to Peters-FR, and GP-FR. Also, the proposed FS-FR obtained the improvement with more than 5% relatively to H-FLSR and TS-FR. These improvements further indicate the better generalization capabilities of the models developed by the proposed FS-FR.

Table 6
Generalization errors (%) for tea concentration obtained by the tested methods

Cross validation number	FS-FR (%)	TS-FR (%)	Peters-FR (%)	H-FLSR (%)	GP-FR (%)
1,2	6.235	4.758	6.146	9.070	6.384
2,3	11.531	3.084	13.514	7.629	12.009
3,4	13.008	20.144	14.246	5.075	13.309
4,5	2.091	10.525	8.486	12.15	3.836
5,6	5.158	2.311	8.145	5.749	5.281
6,7	3.792	4.169	3.278	3.792	2.781
7,8	2.285	1.183	7.236	6.632	4.531
8,9	4.784	4.035	10.221	6.238	5.153
9,10	2.416	16.841	29.130	2.416	3.525
10,11	1.295	1.089	1.692	1.295	1.283
11,12	1.694	5.369	15.710	6.494	4.714
12,13	3.615	5.829	7.817	3.615	8.734
13,14	5.468	3.887	16.721	5.468	8.449
14,15	4.963	9.924	10.478	3.501	6.182
15,16	5.468	5.093	5.115	5.468	8.135
16,1	6.014	9.780	9.485	7.902	9.272
Mean errors	**4.988**	**6.751**	**10.464**	**5.781**	**6.473**
Rank	**1**	**4**	**5**	**2**	**3**

Table 7
Generalization errors for tea temperature obtained by the tested methods

Cross validation number	FS-FR (%)	TS-FR (%)	Peters-FR (%)	H-FLSR (%)	GP-FR (%)
1,2	2.937	3.714	3.077	3.165	5.873
2,3	1.442	0.7550	1.822	2.551	5.187
3,4	0.734	4.364	6.174	1.065	5.931
4,5	1.451	1.697	3.988	1.606	3.976
5,6	1.896	2.361	3.088	1.743	2.331
6,7	1.410	2.841	1.816	1.410	2.623
7,8	2.679	4.158	2.513	5.241	3.527
8,9	3.705	3.823	2.514	4.473	5.459
9,10	6.126	3.007	3.344	1.761	7.358
10,11	5.128	3.459	3.150	1.928	6.334
11,12	2.090	0.602	4.831	3.374	3.083
12,13	2.592	3.667	3.389	6.548	4.390
13,14	1.811	3.102	3.204	6.576	3.239
14,15	1.968	1.977	4.862	1.915	3.486
15,16	2.937	2.679	5.049	1.170	4.700
16,1	3.418	3.034	4.926	3.043	5.137
Mean errors	**2.645**	**2.827**	**3.609**	**2.973**	**4.5396**
Rank	**1**	**4**	**5**	**2**	**3**

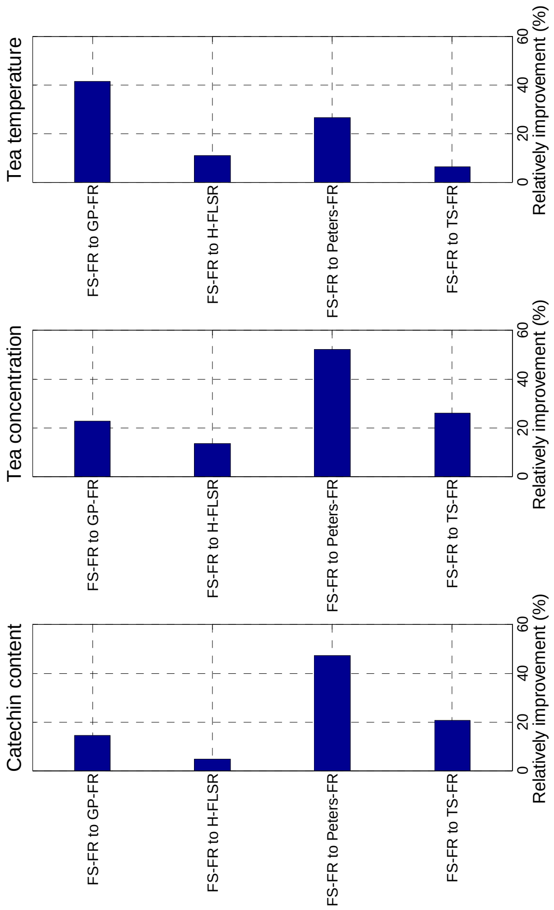

Fig. 1. Relative improvements between the proposed FS-FR to TS-FR, Peters-FR, H-FLSR and GP-FR

Also, the average numbers of engineering characteristics involved in the models developed by the five methods are shown in Figure 2 for the three consumer preferences. Figure 2 illustrates that the models developed by TS-FR, Peters-FR, and H-FLSR all have five engineering characteristics, as the methods require all the five engineering characteristics for developing the models. For the GP-FR, the numbers of engineering characteristics are smaller than those used by TS-FR, Peters-FR and H-FLSR, where on average 3.267, 4.533 and 4.333 engineering characteristics are used by the GP-FR when modelling catechin content, tea concentration, and tea temperature respectively. The number of engineering characteristics used by the proposed FS-FR is the smallest, where on average only 2.267 engineering characteristics are used for catechin content; only 2.000 engineering characteristics are used for tea concentration; and only 1.267 engineering characteristics are used for tea temperature. Therefore, the simplest configuration can be produced by the proposed FS-FR as fewer engineering characteristics are involved in the models.

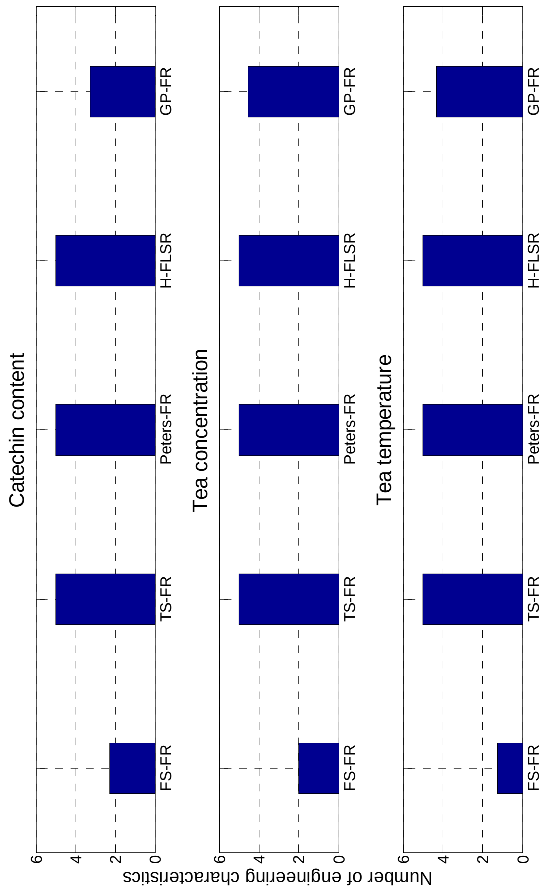

Fig. 2. Numbers of engineering characteristics involved in the models developed by the proposed FS-FR, TS-FR, Peters-FR, H-FLSR and GP-FR

5. Conclusion

In this paper, a FS-FR is proposed by incorporating the approaches of fuzzy least square regression and statistical forward selection. The proposed FS-SR attempts to develop a consumer preference model which includes only significant regressors of engineering characteristics. It attempts to develop consumer preference models with better generalization capabilities than those developed by the other fuzzy regression methods. It overcame the limitations of the commonly-used fuzzy regression which cannot guarantee that the models only consist of significant regressors and do not consist of insignificant regressors. The proposed FS-FR uses the statistical forward selection to determine the polynomial model which only includes significant regressors. Then it uses the fuzzy least square regression to determine the fuzzy coefficients in order to address the uncertainties due to fuzziness and randomness caused by consumer preference evaluations. A case study of a tea maker design demonstrated that the proposed FS-FR was able to generate consumer preference models with smaller generalization errors than the other tested fuzzy regression methods. Also, the proposed FS-FR was able to generate consumer preference models with fewer engineering characteristics than those generated by the other tested methods. Hence, the proposed FS-FR was able to develop simpler and more effective consumer preference models.

References

[1] J.R. Hauser and D. Clausing, The house of quality, *Harvard Business Review* **66** (1988) 63–73.

[2] K.Y. Chan, C. K. Kwong, and M. C. Law, A fuzzy ordinary regression method for modeling customer preference in tea maker design, *Neurocomputing* **142**(22) (2014) 147-154.

[3] A. Griffin, and J.R. Hauser, The voice of the customer, *Marketing Sciences* **12**(1) (1993) 1-27.

[4] K.Y. Chan, C.K. Kwong, and T.S. Dillon, Computational Intelligence Techniques for New Product Design. Springer: Studies in Computational Intelligence, 2012.

[5] E. C. Grigoroudis, V.A. Moustakis, Y. Politis, and L. Tsironis, The assessment of user-perceived web quality: Application of a satisfaction benchmarking approach, *European Journal of Operational Research* **187**(3) (2008) 1346-1357.

[6] H. You, T. Ryu, K. Oh, M.H. Yun and K.J. Kim, Development of customer satisfaction models for automotive interior materials, *International Journal of Industrial Ergonomics* **36**(4) (2006) 323-330.

[7] Z. Liu, Y. Shen and J. Ott, Multilocus association mapping using generalized ridge logistic regression, *BMC Bioinformatics* **12** 384 (2011).

[8] X. Liu, X. Zeng, Y. Xu, and L. Koehl, A fuzzy model for customer satisfaction index in e-commerce, *Mathematics and Computers in Simulation* **77**(5-6) (2008) 512-521.

[9] J. Park, and S.H. Han, A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design, *International Journal of Industrial Ergonomics* **34**(1) (2004) 31-47.

[10] T.E. Karakasidis, D.N. Georgiou and J.J. Nieto, Fuzzy regression analysis: An application on tensile strength of materials and hardness scales, *Journal of Intelligent & Fuzzy Systems* **23**(5) (2012) 177-186.

[11] K.J. Kim, H. Moskowitz, A. Dhingra and G. Evans, Fuzzy multi-criteria models for quality function deployment, *European Journal of Operational Research* **121**(3) (2000) 504-518.

[12] G. Sekkeli, G. Koksal, I. Batmaz and O.T. Bayrak, Classification models based on Tanaka's fuzzy linear regression approach: the case of customer satisfaction modeling, *Journal of Intelligent & Fuzzy Systems* **21**(5) (2010) 341-351.

[13] L.H. Chen, C.C. Hsueh, and C.J. Chang. A two-stage approach for formulating fuzzy regression models, Knowledge-Based Systems **52** (2013) 302-310.

[14] Y. J. Chen, Tang, R.Y.K. Fung, and Z. Ren, Apply fuzzy regression to model functional relationship in product planning, *Proceed. the IEEE Conf. Decis. Cont.* (2003) 4747-4752.

[15] Y. J. Chen, J. Tang, R.Y.K. Fung, and Z. Ren, Fuzzy regression-based mathematical programming model for quality function deployment, *International Journal of Production Research* **42**(5) (2004) 1009-1027.

[16] R.Y.K. Fung, Y. Chen, J. Tang, and Y. Tu, Estimating functional relationships for product planning under uncertainties, *Fuzzy Sets and Systems* **157**(1) (2006) 98-120.

[17] C.K. Kwong, Y. Chen, K.Y. Chan, and X. Luo, A generalised fuzzy least-squares regression approach to modelling relationships in QFD, *Journal of Engineering Design* **21**(5) (2010) 601-613.

[18] A. Safari, H. Shayeghi, and H. A. Shayanfar, Robust state feedback controller design of STATCOM using chaotic optimization algorithm, *Serbian Journal of Electrical Engineering* **7**(2) (2010) 253–268.

[19] H.M. Jiang, C.K. Kwong, W.H. Ip and Z.Q. Chen, Chaos-based fuzzy regression approach to modelling customer satisfaction for product design, *IEEE Transactions on Fuzzy Systems* **21**(5) (2013) 926-936.

[20] J.O. Rawlings, Applied Regression Analysis: A Research Tool. Wadsworth and Brooks, 1988.

[21] E.J. Vladislavleva, G.F. Smits and D.D. Hertog, Order of nonlinearity as complexity measure for models generated by symbolic regression via pareto genetic programming, *IEEE Transactions on Evolutionary Computation* **13**(2) (2009) 333-349.

[22] K.Y. Chan, C.K. Kwong, T.S. Dillon, and K.Y. Fung, An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness, *Journal of Engineering Design* **22**(8) (2011), 523-542.

[23] Y.H.O. Chang, Hybrid fuzzy least-squares regression analysis and its reliability measures, *Fuzzy Sets and Systems* **119**(2) (2001) 225-246.

[24] T. Johnsson, A procedure for stepwise regression analysis, *Statistical Papers* **33** (1992) 21-29.

[25] A. Hong and A. Chen, Piecewise regression model construction with sample efficient regression tree (SERT) and applications to semiconductor yield analysis, *Journal of Process Control* **22**(7) (2006) 1307-1317.

[26] A. Soroush, A. Bahreininejad and J. Van Den Berg, A hybrid customer prediction system based on multiple forward stepwise logistic regression model, *Intelligent Data Analysis* **16**(2) (2012) 265-278.

[27] G. Mauris, L. Berrah, L. Foulloy and A. Haurat, Fuzzy handling of measurement errors in instrumentation, *IEEE Transactions on Instrumentation and Measurement* **49**(1) (2000) 89-93.

[28] C.K. Kwong, Y. Chen, K.Y. Chan, and H. Wong, The hybrid fizzy least-squares regression approach to modeling manufacturing processes, *IEEE Transactions on Fuzzy Systems* **16**(3) (2008) 644-651.

[29] H. Tanaka and J. Watada, Possibilistic linear systems and their application to the linear regression model, *Fuzzy Sets and Systems* **27**(3) (1988) 275-289.

[30] G. Peters, Fuzzy linear regression with fuzzy intervals, *Fuzzy Sets and Systems* **63**(1) (1994) 45-55.

[31] J. Madar, J. Abonyi and F. Szeifert, Genetic programming for the identification of nonlinear input – output models, *Industrial and Engineering Chemistry Research* **44** (2005) 3178-3186.

Table 1
Experimental ranges for the five engineering characteristics

Engineering characteristic	*Reheating temperature (degrees Celsius)* x_1	*Number of drops in the first brewing cycle* x_2	*Dipping time (minutes)* x_3	*Number of drops in the second brewing cycle* x_4`	*Immersion time in the second brewing cycle (seconds)* x_5
Experimental ranges	**93-99**	**1-4**	**8.5-10**	**2-5**	**10-40**
Level 1	93	1	8.5	2	10
Level 2	95	2	9	3	20
Level 3	97	3	9.5	4	30
Level 4	99	4	10	5	40

Table 2
The orthogonal array, $L_{16}(4^5)$, used for the tea maker design

Experiments	*Reheating temperature (degrees Celsius)* x_1	*Number of drops in the first brewing cycle* x_2	*Dipping time (minutes)* x_3	*Number of drops in the second brewing cycle* x_4`	*Immersion time in the second brewing cycle (seconds)* x_5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	2	1	2	3	4
6	2	2	1	4	3
7	2	3	4	1	2
8	2	4	3	2	1
9	3	1	3	4	2
10	3	2	4	3	1
11	3	3	1	2	4
12	3	4	2	1	3
13	4	1	4	2	3
14	4	2	3	1	4
15	4	3	2	4	1
16	4	4	1	3	2

Table 3
Fuzzy observations for the catechin content, y_1, tea concentration, y_2, and tea temperature, y_3

The k-th fuzzy observation	**Catechin content** $\tilde{y}_1(k)$	**Tea concentration** $\tilde{y}_2(k)$	**Tea temperature** $\tilde{y}_3(k)$
1	(1.607,0.077,0.077)	(1.556,0.011,0.011)	(87.800,2.200,2.200)
2	(1.671,0.064,0.064)	(1.723,0.047,0.047)	(85.950,2.250,2.250)
3	(1.522,0.041,0.041)	(1.665,0.222,0.222)	(86.600,0.500,0.500)
4	(1.566,0.063,0.062)	(1.719,0.132,0.132)	(86.400,0.300,0.300)
5	(1.466,0.104,0.104)	(1.781,0.011,0.011)	(87.000,2.800,2.800)
6	(1.595,0.029,0.029)	(1.739,0.159,0.159)	(87.000,1.700,1.700)
7	(1.686,0.144,0.144)	(1.833,0.054,0.054)	(88.200,0.500,0.500)
8	(1.582,0.085,0.085)	(1.835,0.023,0.023)	(84.450,0.850,0.850)
9	(1.690,0.079,0.079)	(1.851,0.084,0.084)	(88.250,3.850,3.850)
10	(1.644,0.042,0.042)	(1.847,0.200,0.200)	(90.850,5.550,5.550)
11	(1.738,0.058,0.058)	(1.847,0.070,0.070)	(88.750,2.050,2.050)
12	(1.727,0.067,0.067)	(1.915,0.049,0.049)	(88.900,1.500,1.500)
13	(1.687,0.023,0.023)	(2.081,0.019,0.019)	(87.300,1.500,1.500)
14	(1.821,0.114,0.114)	(1.941,0.163,0.163)	(91.850,0.250,0.250)
15	(1.580,0.137,0.137)	(1.860,0.048,0.048)	(89.850,1.350,1.350)
16	(1.883,0.049,0.039)	(2.007,0.002,0.002)	(89.400,2.700,2.700)