Fuzzy Linear Regression for the Time Series Data which is Fuzzified with SMRGT Method

Seçil YALAZ[*1], Arife ATAY[1]

[1] Dicle Üniversitesi, Fen Fakültesi, Matematik Bölümü, 21280 Diyarbakır

Abstract: Our work on regression and classification provides a new contribution to the analysis of time series used in many areas for years. Owing to the fact that convergence could not obtained with the methods used in autocorrelation fixing process faced with time series regression application, success is not met or fall into obligation of changing the models' degree. Changing the models' degree may not be desirable in every situation. In our study, recommended for these situations, time series data was fuzzified by using the simple membership function and fuzzy rule generation technique (SMRGT) and to estimate future an equation has created by applying fuzzy least square regression (FLSR) method which is a simple linear regression method to this data. Although SMRGT has success in determining the flow discharge in open channels and can be used confidently for flow discharge modeling in open canals, as well as in pipe flow with some modifications, there is no clue about that this technique is successful in fuzzy linear regression modeling. Therefore, in order to address the luck of such a modeling, a new hybrid model has been described within this study. In conclusion, to demonstrate our methods' efficiency, classical linear regression for time series data and linear regression for fuzzy time series data were applied to two different data sets, and these two approaches performances were compared by using different measures.

SMGRT Yöntemi İle Bulanıklaştırılmış Zaman Serileri İçin Bulanık Doğrusal Regresyon

Özet: Regresyon ve sınıflandırma üzerine yaptığımız bu çalışma, yıllardır birçok alanda kullanılan zaman serileri analizine yeni bir katkı sağlamaktadır. Zaman serileri için regresyon uygulamasında karşılaşılan otokorelasyonun kaldırılması aşamasında çoğu kez ya uyum sağlanamadığından başarıya ulaşılamamakta ya da modelin derecesinin değiştirilmesi zorunluluğuyla karşı karşıya kalınmaktadır. Modelin derecesinin değiştirilmesi ise her zaman istenilen bir durum olmayabilir. Böyle durumlarda kullanılmak üzere önerilen çalışmamızda, zamana bağlı veriler basit üyelik fonksiyonu ve bulanıklık kuralı üretim tekniği (SMRGT) ile bulanıklaştırılmış ve elde edilen değişkenler için bulanık en küçük kareler (Bulanık EKK) modeli ile basit doğrusal regresyon yöntemi uygulanarak geleceğe yönelik tahmine ilişkin bir denklem oluşturulmuştur. SMRGT açık kanallarda debi akışını belirlemede başarılı olmasına ve açık kanallarda debi akışını modellemede güvenle kullanılabilmesine rağmen bu tekniğin bulanık doğrusal regresyon modellemesinde de başarılı olacağı hakkında hiçbir ip ucu yoktur. Bu nedenle bu tür bir modellemenin eksikliği adres gösterilerek yeni bir hibrit model bu çalışma kapsamında tarif edilmiştir. Sonuç olarak yöntemin geçerliliğinin ölçülebilmesi bakımından zaman serileri için doğrusal regresyon ve bulanık zaman serileri için doğrusal regresyon iki ayrı veri setine uygulanmış ve bu iki yaklaşımın performansları çeşitli ölçüm kriterleri kullanılarak karşılaştırılmıştır.

1. Introduction

While the modeling of some systems that human estimation is effective, it can be encountered with a fuzzy structure. These fuzzy structure parameters can be presented as a fuzzy linear function obtained from fuzzy sets. Fuzzy linear functions are defined by the Zadehs' expansion theory [1].

The fuzzy time series forecasting problem has been studied by several authors at last decade. Song and Chissom introduced time-variant and time-invariant fuzzy time series models, gave properties of them and discussed procedures to develop fuzzy time series models [2]. Chen (1996) proposed a method which is more efficient than the one presented in Song and Chissom due to the fact that his method uses simplified arithmetic operations rather than the complicated max-min composition operations presented in Song and Chissom [3]. Huarng determined effective lengths of intervals, distribution and average based length, which has not been touched in previous studies [4]. Chen (2002) presented a new method, to deal with the forecasting problems based on high-order fuzzy time series and genetic algorithms, which achieves a higher forecasting accuracy rate than the existing methods [5]. Tsaur et. al obtained a fuzzy relation matrix represents a time-invariant relation to measure the degrees of fuzziness [6]. Singh proposed a new method of fuzzy time series forecasting based on difference parameters for the accuracy in the forecasted values [7].

All these studies are related directly with fuzzy time series. In present work the aim is to introduce the method to fuzzify variables for using fuzzy least squares method.

When a fuzzy linear function is considered for modeling the fuzziness of the system, a fuzzy linear regression analysis is formulated. Owing to the fact that convergence could not obtained with the methods used in autocorrelation fixing process faced with time series regression application, success is not met or fall into obligation of changing the models' degree. Changing the models' degree may not be desirable in every situation. In these situations a new regulation methodology is needed.

Toprak proposed SMRGT for open canal flow modeling [8]. Toprak et. al also used this method on determination of losses in water-networks [9,10]. Coşkun created the automated fuzzy model generation based on SMRGT [11]. Yalaz (Toprak) et. al used SMRGT to fuzzify variables to use least squares method to predict linear parameters [12,13].

In this study a new hybrid model is proposed based on the basic concepts of SMRGT to fuzzify variables and fuzzy regression models approaches to time-series forecasting [14]. In the proposed model, the SMRGT is used to preprocess raw data and to provide the necessary background to apply a fuzzy regression model. Success of SMRGT is only demonstrated in determining the flow discharge in open channels. Because of this reason the new hybrid model has been described with some changes in SMRGT to show success of this technique in fuzzy linear regression modeling.

In order to highlight its appropriateness and effectiveness, our proposed method is applied to two different data sets and their performance is compared with linear regression model for time series.

2. Material and Method

2.1. Time series

Time series, known as the series obtained as a function of time. Time series which is the most important way to make predictions for the future is based on time dependency on successive observations. Incoming observations can be predicted with a function which is adapted to the series. Economic time series is in the first place in time series application subjects.

To perform time series analysis, trend, cyclical variations, seasonal variations and irregular variations which lead to errors in parameter estimation, in other words cause stationary took into account, are necessary to control. A time series is stationary if its mean and variance do not vary systematically over time [15]. In nonstationary series using indexes calculated for the effects except time, relevant effects are destroyed. Time series analysis is applicable after this process [16].

In most of the time series, consecutive observations are interconnected. In this case the method that does not use dependency advantage is not suitable. Instead of this method using *ARIMA* models which use the dependence structure of time series in a very efficient manner and known as Box-Jenkins (BJ) forecasting models [17] will be more appropriate.

BJ forecasting models are successful methods used to predict future estimates of univariate time series. Because these models determine the structure of a time series, use the dependency of observations and include the statistical tests in model determination stages, compared to other estimation methods they are the superior models to make short-term prediction [18]. BJ forecasting models can be divided into four models;

  i. *AR* (Autoregressive) Models,
 ii. *MA* (Moving Average) Models,
iii. *ARMA* (Autoregressive Moving Average) Models,
 iv. *ARIMA* (Autoregressive Integrated Moving Average) Models.

Because we will only deal with the AR model, this model will be introduced here. Autoregressive (AR) model describes observation of a time series at any time during, depending on the error term and a certain number of past period observations of the series. AR model is called according to the number of past period observation that they contain. Generally, AR model of order p denoted by $AR(p)$ and defined

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + \varepsilon_t \tag{1}$$

where X_t, X_{t-1}, $X_{t-2}, \ldots$, X_{t-p} are the independent variables and $\varphi_1, \varphi_2, \ldots, \varphi_p$ are the model parameters. It is assumed white noise process has normal distribution with zero mean and σ_ε^2 variance. The model contains $p + 2$ unknown parameters $(\mu, \sigma_\varepsilon^2)$ which have to be estimated. In practice commonly used AR models are the first and second degree AR models which are demonstrated as $AR(1)$ and $AR(2)$. AR models can be written in the form of difference models as follows,

$$BX_t = X_{t-1}, B^2 X_t = X_{t-2}, \cdots, B^p X_t = X_{t-p} \tag{2}$$

Using this equation the p-th degree of AR model is written as
$$X_t = \left(\varphi_1 B + \cdots + \varphi_p B^p\right) X_t + \varepsilon_t$$
or,
$$\varepsilon_t = \left(1 - \varphi_1 B - \cdots - \varphi_p B^p\right) X_t \tag{3}$$

where B is called backward shift operator. Stationary condition in AR models is possible in terms of remaining the roots found by equalizing polynomial to zero outside the unit circle. If the roots are outside the unit circle, $AR(p)$ can be used for stationary time series [17]. $AR(1)$ model descibed by backward shift operator is $\varepsilon_t = (1 - \varphi_1 B) X_t. |\varphi_1| < 1$ condition should be satisfied for the stationary model. $AR(2)$ model descibed by backward shift operator is $\varepsilon_t = (1 - \varphi_1 B - \varphi_2 B^2) X_t$. $\varphi_1 + \varphi_2 < 1$, $\varphi_2 - \varphi_1 < 1$ and $|\varphi_1| < 1$ equations should be satisfied for the stationary model.

Selection of appropriate models for time series, is performed by BJ Models. The steps are described below:

i. Model group is determined by examining the series forming observations. At this stage, it is decided to which model of the group would be appropriate. The first work which has to be done in determining the appropriate model is to determine the stationary. The tools for examining the stationary are autocorrelation and correlograms of functions.

ii. It is decided from the group agreed, which model type would be appropriate for the relevant series. While analysing stationary time series and at forecasting stages one of three types of models (AR, MA, $ARMA$) is used. Model selection is performed with using autocorrelation and partial autocorrelation functions. The degree of parameters is also determined for the model selected in this way.

iii. The model parameters judged to appropriate are estimated. Estimation of the parameters phase is extremely complex and requires the time-consuming process. These processes differ in the type and each model is performed with the use of statistical packages.

iv. Using the standard error of the estimated autocorrelations as a measure cannot be clearly demonstrated the importance of autocorrelations differs from zero, calculated in low degree delay [19]. Therefore, adequacy of the model is tested with Box-Pierce statistics. If the model is adequate it is used for the estimation. Otherwise, it should be returned to the first step.

2.1.1. Linear regression for time series

When equation (1) demonstrates autoregressive $AR(p)$ model considered for $t = 2, \ldots, n$; it can be written as;

$$AR(p) = \begin{cases} X_2 = \varphi_1 X_1 + e_2 \\ X_3 = \varphi_1 X_2 + \varphi_2 X_1 + e_3 \\ \vdots \\ X_n = \varphi_1 X_{n-1} + \cdots + \varphi_{n-1} X_1 + e_n \end{cases} \tag{4}$$

This equation is not defined for $t = 1$. Because there is no X_0. When $\mathbf{y} = (X_2, \ldots, X_n)^T$ the matrix form of the equations (4) can be shown as

$$\begin{pmatrix} X_2 \\ X_3 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ X_2 & X_1 & \cdots & 0 \\ \vdots & \vdots & X_1 & \vdots \\ X_{n-1} & X_{n-2} & \cdots & X_1 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{n-1} \end{pmatrix} + \begin{pmatrix} e_2 \\ e_3 \\ \vdots \\ e_n \end{pmatrix}$$

$$\mathbf{y}_{((n-1)x1)} = \mathbf{X}_{((n-1)x(n-1))} \boldsymbol{\varphi}_{((n-1)x1)} + \boldsymbol{\varepsilon}_{((n-1)x1)}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\varphi} + \boldsymbol{\varepsilon} \tag{5}$$

In linear regression method, putting $\boldsymbol{\varphi}$ parameter in the regression equation gives you ability to make predictions about future [20].

Equation (5) which is constructed dependent on $AR(p)$ time series model is similar to linear regression equation. However, to be able to apply regression method to this equation, $E(\varepsilon_i, \varepsilon_{i-1}) = 0$ assumption which is not coincided in time series, should be implemented [21]. Thus, DW test is applied to this model to measure autocorrelation [22,23]. If, there are autocorrelations between consecutive errors, Prais-Winsten approach [24] can be used to fix autocorrelation. In this method, autocorrelation between consecutive errors can be fixed gradually [25]. Thus, $\boldsymbol{\varphi}$ parameters can be estimated via linear least square regression estimation method.

Lets fix the autocorrelations between consecutive errors via Prais-Winsten method, considering $AR(1)$ model $X_t = \varphi_1 X_{t-1} + \varepsilon_t$; $t = 2, \ldots, n$,. Stationary condition is defined as $|p| < 1$. For this correction, in the first step, first derivative of the residual sum of squares RSS should be equalized to zero as follows:

$$RSS = \sum_{t=2}^{n} \varepsilon_t^2 = \sum_{t=2}^{n} [X_t - \hat{\varphi}_1 X_{t-1}]^2,$$

$$\frac{\partial RSSS}{\partial \hat{\varphi}_1} = \sum_{t=2}^{n} (-2) X_{t-1} [X_t - \hat{\varphi}_1 X_{t-1}] = 0,$$

$$\sum_{t=2}^{n} X_{t-1} X_t - \hat{\varphi}_1 \sum_{t=2}^{n} X_{t-1}^2 = 0. \tag{6}$$

When the equation above is solved, estimated value of φ_1 can be found as $\hat{\varphi}_1 = \frac{\sum_{t=2}^{n} X_{t-1} X_t}{\sum_{t=2}^{n} X_{t-1}^2}$. Using with estimated $\hat{\varphi}_1$ in the equation, for $t = 2, \ldots, n$;

$$\hat{\varepsilon}_t = X_t - \hat{\varphi}_1 X_{t-1} = X_t - \left(\frac{\sum_{t=2}^{n} X_{t-1} X_t}{\sum_{t=2}^{n} X_{t-1}^2} \right) X_{t-1} \tag{7}$$

values can be calculated. Thus, correlation coefficient between t_th and $(t-1)$_th errors can be calculated as using $\hat{\rho}_{pw}^1 = \frac{\sum \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_{t-1}^2}$ for $t = 3, \ldots, n$

$$\hat{\rho}_{pw}^1 = \frac{\sum \left[X_t - \frac{\sum_{t=2}^{n} X_{t-1} X_t}{\sum_{t=2}^{n} X_{t-1}^2} X_{t-1} \right] \left[X_{t-1} - \frac{\sum_{t=3}^{n} X_{t-2} X_{t-1}}{\sum_{t=3}^{n} X_{t-2}^2} X_{t-2} \right]}{\sum \left[X_{t-1} - \frac{\sum_{t=3}^{n} X_{t-2} X_{t-1}}{\sum_{t=3}^{n} X_{t-2}^2} X_{t-2} \right]^2} \tag{8}$$

If this value is put to the 3_th step equation of Prais-Winsten method, for $t = 3, \ldots, n$

$$\left(X_t - \hat{\rho}_{pw}^1 X_{t-1} \right) = b \left(X_{t-1} - \hat{\rho}_{pw}^1 X_{t-2} \right) + v_t \tag{9}$$

is acquired. In the second step, using $X_{t-1}^* \to \left(X_{t-1} - \hat{\rho}_{pw}^1 X_{t-2} \right)$, $X_t^* \to \left(X_t - \hat{\rho}_{pw}^1 X_{t-1} \right)$ a new equation can be found as $X_t^* = b X_{t-1}^* + v_t$ and returned to the first step. In this manner, this cycle should be run till acquire $\hat{\rho}_{pw}^i \to 0$, $(i = 1, \ldots, n)$.

The last equation which has no autocorrelation will provide the regression assumptions stated above and will be ready to be applied regression methods. The equation which has no autocorrelations also can be found for $AR(2), \ldots, AR(p)$ models. After the time series equation gets ready to be applied linear regression with Prais-Winsten method, the least square estimation of $\boldsymbol{\varphi}$ can be obtained as

$$\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_2, \hat{\varphi}_2, \ldots, \hat{\varphi}_{n-1})^T;$$

$$\hat{\boldsymbol{\varphi}} = min_\varphi \|\boldsymbol{\varepsilon}\|^2 = min_\varphi \|\boldsymbol{y} - \boldsymbol{X\varphi}\|^2$$
$$= min_\varphi (\boldsymbol{y} - \boldsymbol{X\varphi})^T (\boldsymbol{y} - \boldsymbol{X\varphi}) \tag{10}$$

This minimization problem, acquired with the solutions of normal equations in linear regression defined as

$$\frac{\partial (\boldsymbol{y} - \boldsymbol{X\varphi})^T (\boldsymbol{y} - \boldsymbol{X\varphi})}{\partial \boldsymbol{\varphi}} = 0$$

$$\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}} = (\mathbf{X}^\mathrm{T} \mathbf{X})^{-1} \mathbf{X}^\mathrm{T} \mathbf{y} \tag{11}$$

If $\mathbf{X}^\mathrm{T} \mathbf{X}$ is singular, the system is inconsistence and the equation above can not be obtained. In this case, using Tikhonov (Ridge estimator) regularization the problem can be solved with singular value decomposition [26].

However at the end of the Prais-Winsten procedure, fixes autocorrelation gradually, autocorrelation could not removed or fall into obligation of changing the models' degree. In these undesirable situations a new hybrid model is proposed based on the basic concepts of SMRGT to fuzzify variables and fuzzy regression models approaches to time-series forecasting.

3. Fuzzy Forecasting Models

3.1. Simple membership function and fuzzy rule generation technique (SMRGT)

As mentioned previously, the main question in any given fuzzy system is how construct the membership functions (MFs) and fuzzy rules (FRs), such that the system yields the best results. In this study, therefore, a simple technique is proposed to help those who have difficulties in deciding on the number, the shape, and the logic of the MFs and FRs in any fuzzy system. In Toprak, 2009, a new fuzzy technique was introduced for open canal flow modeling depends only on some key numbers for all MFs of input and output variables. The key numbers were selected according to the MF shape (triangular, trapezoid, etc.) and the defuzzification method (centroid, maximum membership degree, etc.).

The proposed method used the following algorithm with relevant steps, which should be applied for successful results [8]:
i. Decide on the independent variables which affect the dependent variable for the event at hand. The independent variables are inputs, whereas the dependent variable(s) is the output of the fuzzy system.
ii. Determine the maximum and the minimum values (variation domain) for each variable.
iii. Decide on the MF shape (i.e., triangular, trapezoidal, etc.).
iv. Decide on the number of the MF for each independent variable (a minimum number of 3MF is required).
v. Determine the width and the core of the MFs with their key values for each independent

variable. Note that the number of key values will be equal to the number of the MFs for each independent variable (Figure 1).
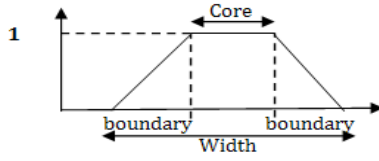

Figure 1. Parts of fuzzy set

vi. These key values are the inputs of the fuzzy model.
vii. The fuzzy model is valid for the data distributed between the key values of the first and the last MFs for each variable.
viii. Tables of fuzzy rules are prepared (Table1).

Although this technique is successful in determining the flow discharge in open channels and can be used confidently for flow discharge modeling in open canals, as well as in pipe flow with some modifications, there is no research paper which demonstrates that this technique is successful in fuzzy linear regression modeling. Therefore, in order to address the luck of such a modeling, a new model has been described within this study.

3.2. Fuzzy least squares

A clear path to move fuzzy regression model, in line with the statistical regression, is modeling fuzzy regression along the same lines [27]. Let began with the standard linear regression model described in equation (5).

In contrast, the fuzzy regression model may take the following form:

$$\tilde{y} = \tilde{X}\varphi + \tilde{\varepsilon} \qquad (12)$$

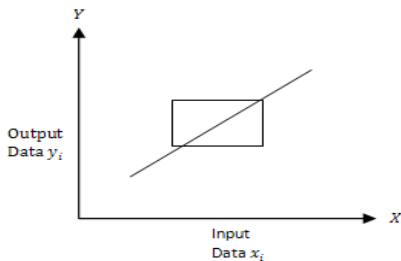Conceptually the i-th fuzzy response and explanatory variables as shown in the figure 2.

Figure 2. Relationships between variables in fuzzy linear regression

If the equation is rearranged, $\tilde{\varepsilon} = \tilde{y} - \tilde{X}\varphi$ is obtained. The problem turns into the following form with least squares perspective:

$$\min \sum \left(\tilde{y} - \tilde{X}\varphi \right)^2 \qquad (13)$$

4. Formulation of the Proposed Model

For a given fuzzy linear regression how can the least squares approach be optimally designed? SMRGT can be used with the well-known fuzzy least squares equation which is shown in equation (12) for fuzzification of the variables. With this method, the limitation of autocorrelation fixing process in linear regression process is lifted through investing on the advantages of the fuzzy regression models.

The main difference of the proposed method from the SMRGT is that output is the multiple of inputs in SMRGT, but in our method φ parameter is estimated using input and output. Because of this situation the number of MFs (n_{MFs}) for dependent variable is chosen square of n_{MFs} for independent variable.

Let fuzzify the independent variable $X = (X_1, \dots, X_n)^T$ and dependent variable $y = (X_2, \dots, X_n)^T$. The problem can be modeled in light of the SMRGT algorithm given in the section 3.1.

The maximum and minimum of the independent and dependent variable X_{max}, X_{min} were decided.

For this model the MF shapes adopted were triangles, which are the simplest form of the MF, for each variable.

Because of the difficulty in running the program from hard or removable drives n_{MFs} for each independent variable is limited to five. With this decision n_{MFs} for dependent variable is limited to twenty five.

The width (variation range) X_R of the variables can be obtained by:

$$X_R = X_{max} - X_{min} \qquad (14)$$

The key values of the intermediated MFs, K_i is determined. For the intermediated MFs one can accept the unit width of the MF as the width of a right angled triangle. Therefore, the unit width, UW, is given as:

$$UW = \frac{X_R}{n_u} \qquad (15)$$

where n_u was the number of units.

On the other hand, the neighboring MFs should overlap with each other. Therefore, for each unit of MF it was necessary to have extended unit width, EUW, symmetrically as:

$$EUW = \frac{3UW}{2} \qquad (16)$$

For the center MF:

$$K_i = \frac{X_R}{2} + X_{min} \quad (17)$$

For the previous MF:

$$K_{i-1} = \frac{K_i - X_{min}}{2} + X_{min} \quad (18)$$

For the next MF:

$$K_{i+1} = X_{max} - \left(\frac{X_{max} - K_i}{2}\right) \quad (19)$$

The key values of the first, K_1, and the last, $K_{n_{MFs}}$, MFs corresponded to the centroid of these MFs, which can be calculated as:

$$K_1 = X_{min} + \frac{EUW}{3} \text{ and } K_{n_{MFs}} = X_{max} - \frac{EUW}{3} \quad (20)$$

Using this rules the key values fuzzy dependent $\tilde{y} = \tilde{X}_\tau$ and independent variable $\tilde{X} = [\tilde{X}_1\ \tilde{X}_2 \dots \tilde{X}_p] = [\tilde{X}_{\tau-1}\ \tilde{X}_{\tau-2} \dots \tilde{X}_{\tau-p}]$ can be calculated. These key values are used to create table of fuzzy rules (Table 1). Data of this table are the inputs and outputs of the fuzzy least squares method for fuzzy linear regression model

$$\tilde{\varphi} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \quad (21)$$

In the current modeling we proposed to use triangular MFs and the centroid method for

defuzzification. The procedure of the proposed model is described in figure 3.
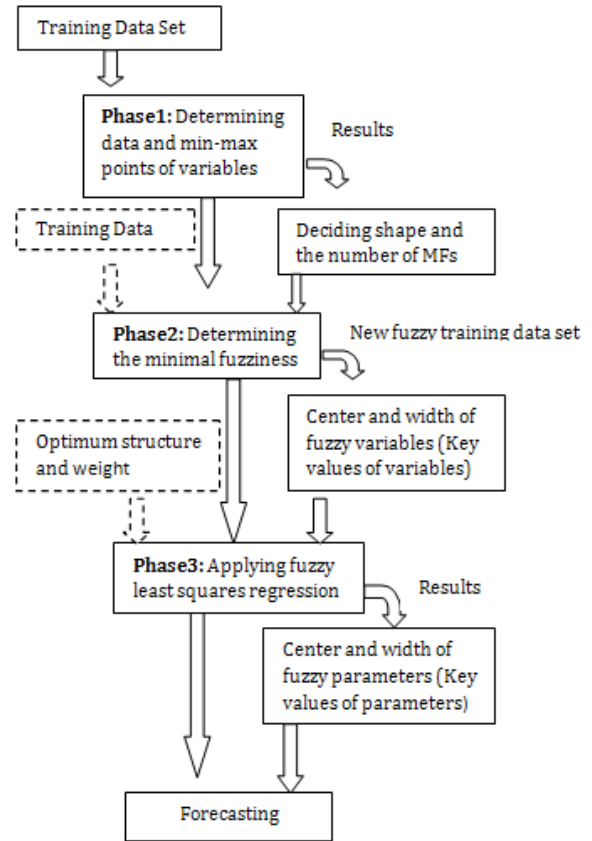

Figure 3. Procedure in the proposed model

Table 1. Fuzzy rules for $n_{MFs} = 5, p = 2$

Fuzzy Rules Number	Independent Variable1		Independent Variable2		Dependent Variable	
	Numerical	Linguistic	Numerical	Linguistic	Numerical	Linguistic
1	K_1^1	Very Low	K_1^2	Very Low	K_1^y	1
2	K_1^1	Very Low	K_2^2	Low	K_2^y	2
3	K_1^1	Very Low	K_3^2	Medium	K_3^y	3
4	K_1^1	Very Low	K_4^2	High	K_4^y	4
5	K_1^1	Very Low	K_5^2	Very High	K_5^y	5
6	K_2^1	Low	K_1^2	Very Low	K_6^y	6
7	K_2^1	Low	K_2^2	Low	K_7^y	7
8	K_2^1	Low	K_3^2	Medium	K_8^y	8
9	K_2^1	Low	K_4^2	High	K_9^y	9
10	K_2^1	Low	K_5^2	Very High	K_{10}^y	10
11	K_3^1	Medium	K_1^2	Very Low	K_{11}^y	11
12	K_3^1	Medium	K_2^2	Low	K_{12}^y	12
13	K_3^1	Medium	K_3^2	Medium	K_{13}^y	13
14	K_3^1	Medium	K_4^2	High	K_{14}^y	14
15	K_3^1	Medium	K_5^2	Very High	K_{15}^y	15
16	K_4^1	High	K_1^2	Very Low	K_{16}^y	16
17	K_4^1	High	K_2^2	Low	K_{17}^y	17
18	K_4^1	High	K_3^2	Medium	K_{18}^y	18
19	K_4^1	High	K_4^2	High	K_{19}^y	19
20	K_4^1	High	K_5^2	Very High	K_{20}^y	20
21	K_5^1	Very High	K_1^2	Very Low	K_{21}^y	21
22	K_5^1	Very High	K_2^2	Low	K_{22}^y	22
23	K_5^1	Very High	K_3^2	Medium	K_{23}^y	23
24	K_5^1	Very High	K_4^2	High	K_{24}^y	24
25	K_5^1	Very High	K_5^2	Very High	K_{25}^y	25

5. Applications

5.1. Description of data sets used in applications

Two data sets are used in the applications. First data set, Chemical Process Concentration Readings (CPCR), denotes a chemical processes concentration readings measured at intervals of two hours. It contains 197 observations. Second data set, Chemical Process Viscosity Readings (CPVR), denotes a chemical processes viscosity values of an hour. It contains 310 observations. Data sets scatterplots are shown in the Appendix. For detailed information about CPCR and CPVR see Box-Jenkins (1976).

5.2. Validation approach and comparison measures

To evaluate the performances of regression for time series and SMRGT several measures can be used. The performance measures that we used in our applications are Adjusted R^2 (Adj-R^2), R^2, Mean Absolute Error (MAE), Mean Absolute Percentage Error ($MAPE$), Mean Square Error (MSE), Root Mean Square Error ($RMSE$), Akaike's Information Criteria (AIC) and Correlation Coefficient (rho).

5.3. Construction of models

5.3.1. Construction of linear regression model

While regression model is comprised, firstly, time series models should equip with regression assumptions. In time series there is autocorrelation between consecutive errors and this situation can be fixed with some autocorrelation correction methods. In this study, autocorrelation is measured with Darwin Watson Test [22,23], and it is fixed with Prais-Winsten Procedure using STATA [28]. In both data sets convergence is satisfied for $p = 2$ and so, $AR(2)$ time series model is used.

5.3.2. Construction of SMRGT model

In this case, the five MFs, of which the first and the second of them are right angled triangles, consisted of eight units. Because of this reason for independent variables $n_{MFs} = 5, n_u = 8$ and for dependent variable $n_{MFs} = 25, n_u = 48$.

Application of the method for CPCR data

$$X_R = 18.5 - 16 = 2.5$$
$$UW = \frac{2.5}{8} = 0.3125$$
$$EUW = \frac{3 * 0.3125}{2} = 0.46875$$
$$K_1 = 16 + \frac{0.46875}{3} = 16.15625$$
$$K_5 = 18.5 - \frac{0.46875}{3} = 18.34375$$
$$K_i = \frac{2.5}{2} + 16 = 17.25$$
$$K_{i+1} = 18.5 - \left(\frac{18.5 - 17.25}{2}\right) = 17.875$$
$$K_{i-1} = \frac{17.25 - 16}{2} + 16 = 16.625$$

Application of the method for CPVR data

$$X_R = 10.5 - 7 = 3.5$$
$$UW = \frac{3.5}{8} = 0.4375$$
$$EUW = \frac{3 * 0.4375}{2} = 0.65625$$
$$K_1 = 7 + \frac{0.65625}{3} = 7.21875$$
$$K_5 = 10.5 - \frac{0.65625}{3} = 10.28125$$
$$K_i = \frac{3.5}{2} + 7 = 8.75$$
$$K_{i+1} = 10.5 - \left(\frac{10.5 - 8.75}{2}\right) = 9.625$$
$$K_{i-1} = \frac{8.75 - 7}{2} + 7 = 7.875$$

According to our method, for $X_{\tau-1}$ and $X_{\tau-2}$ in CPCR data new key values are found as 16.15625, 16.625, 17.25, 17.875, 18.34375. Also for $X_{\tau-1}$ and $X_{\tau-2}$ in CPVR data new key values are found as 7.21875, 7.875, 8.75, 9.625, 10.28125. We can also obtain key values for X_τ in a similar way. Now, we can construct the tables (Table 2 and 3) contain fuzzy rules for located key values of both data.
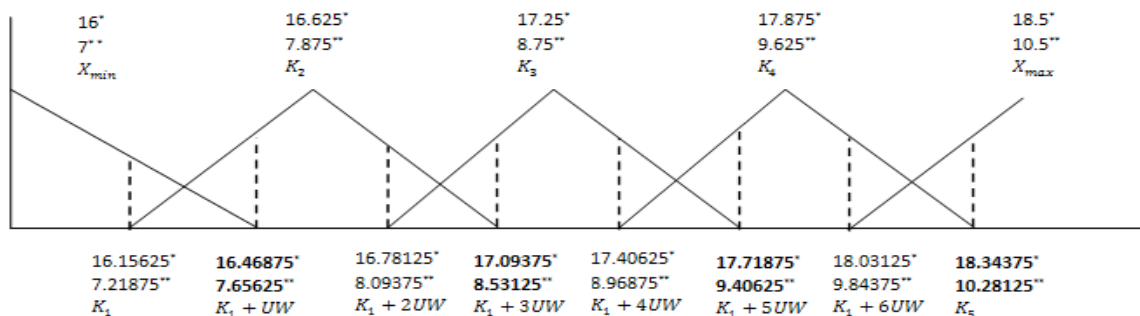
Figure 4. Construction of triangular membership functions and key values to apply centroid deffuzzyfication method for the CPCR* and CPVR** data

5.4. Comparison of models

MATLAB [29] codes are written for performance measures which are used to measure the performances of regression for time series and SMRGT methods and the following results are obtained.

Table 4. Performance Measures For CPCR Data

	LINEAR REGRESSION	SMRGT
Adj-R^2	0.3607	*0.7916*
.R^2	0.3674	*0.8090*
MAE	0.2422	0.2759
MAPE	1.4175	1.5950
MSE	0.1013	0.1065
RMSE	0.3207	0.3479
AIC	-447.06	-437.20
rho	0.6076	*0.9819*

Table 5. Performance measures for CPVR data

	LINEAR REGRESSION	SMRGT
Adj-R^2	0.7284	*0.8767*
.R^2	0.7302	*0.8870*
MAE	0.2432	0.2899
MAPE	2.6937	3.5185
MSE	0.0960	0.1214
RMSE	0.3114	0.3714
AIC	-722.46	-649.69
rho	0.8588	*0.9962*

6. Discussion and Conclusion

Owing to the fact that convergence could not obtained with the methods used in autocorrelation fixing process faced with time series regression application, success is not met or fall into obligation of changing the models' degree. Changing the models' degree may not be desirable in every situation. In these situations a new regulation methodology is needed. With this study, it is proposed to use a new method SMRGT to fuzzify the variables for using fuzzy OLS models and the steps should be followed are shown. In this method, the limitation of autocorrelation fixing process in linear regression process is lifted through investing on the advantages of the fuzzy regression models.

Although SMRGT is successful in determining the flow discharge in open channels and can be used confidently for flow discharge modeling in open canals, as well as in pipe flow with some modifications, there is no research paper which demonstrates that this technique is successful in fuzzy linear regression modeling. Therefore, in order to address the luck of such a modeling, the new hybrid model has been described with some changes in SMRGT to show success of this technique in fuzzy linear regression modeling.

This study also include the several performance measurement criteria results independent from methods to measure how effective the proposed technique. When given 8 performance criteria used for the two sets of data are considered, on three of these criteria SMRGT model showed a better performance than the linear regression model (Table 4 and Table 5). MAE, MAPE, MSE, RMSE and AIC values calculated for SMRGT are close to the values obtained for linear regression. According to this study when working with fuzzy numbers fuzzified with SMRGT it can be said that our method is an alternative method to linear regression model for time series data.
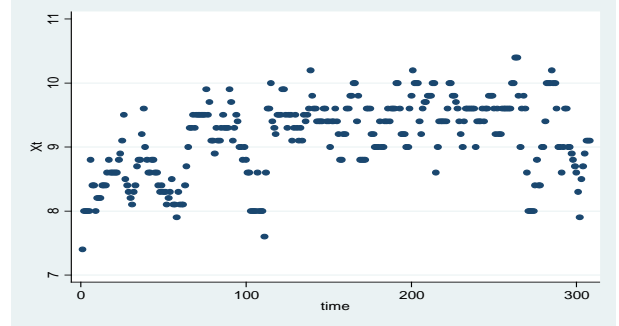
References

[1] Zadeh, L.A., 1965. Fuzzy Sets, Inf. Control 8 (3) (1965), 338–353.

[2] Q. Song and B. S. Chissom, "Fuzzy time series and its models," Fuzzy Sets and Systems, vol. 54, no. 3, pp. 269–277, 1993.

[3] S. M. Chen, "Forecasting enrollments based on fuzzy time series,"Fuzzy Sets and Systems, vol. 81, no. 3, pp. 311–319, 1996.

[4] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," Fuzzy Sets and Systems, vol. 123, no. 3, pp. 387–394, 2001.

[5] S. Chen, "Forecasting enrollments based on high-order fuzzy time series," Cybernetics and Systems, vol. 33, no. 1, pp. 1–16, 2002.

[6] R.-C. Tsaur, J. C. O. Yang, and H.-F. Wang, "Fuzzy relation analysis in fuzzy time series model," Computers & Mathematics with Applications, vol. 49, no. 4, pp. 539–548, 2005.

[7] S. R. Singh, "A simple method of forecasting based on fuzzy time series," Applied Mathematics and Computation, vol. 186, no. 1, pp. 330–339, 2007.

[8] Toprak, Z.F., 2009. Flow Discharge Modeling in Open Canals Using a New Fuzzy Modeling Technique (SMRGT), CLEAN-Soil. Air. Water 37 (9) (2009), 742–752.

[9] Toprak ZF, Songur M, Hamidi N, and Gulsever H, (2012), Determination of Losses in Water-Networks Using a New Fuzzy Technique (SMRGT), AWERProcedia Information Technology & Computer Science, Vol 03 (2013) 833-840.

[10] Toprak ZF, Songur M, Hamidi N, and Gulsever H, (2012), Determination of Losses in Water-Networks Using a New Fuzzy Technique (SMRGT), 3rd World Conference on Information Technology (WCIT 2012), 14-16 November 2012, Barcelona-Spain.

[11] Coşkun, C. (2014), Automated Fuzzy Model Generation and an Analysis of the Proposed

Method, Int. J. Open Problems Compt. Math., Vol. 7, No. 2, June 2014 ISSN 1998-6262

[12] Yalaz S, A Atay, Toprak ZF (2015), SMRGT yöntemi ile bulanıklaştırılmış veriler için bulanık doğrusal regresyon, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 31(3), 152-158, ISSN: 1012-2354

[13] Toprak S, Atay A, Toprak ZF (2013), SMRGT yöntemi ile bulanıklaştırılmış veriler için bulanık doğrusal regresyon, XXVI. Ulusal Matematik Sempozyumu, Dicle Üniversitesi Kongre Merkezi, Diyarbakır.

[14] Toprak S, Atay A, Toprak ZF (2013), SMRGT yöntemi ile bulanıklaştırılmış zamana bağlı veriler için bulanık doğrusal regresyon, XXVI. Ulusal Matematik Sempozyumu, Dicle Üniversitesi Kongre Merkezi, Diyarbakır.

[15] Gujarati, D.N., 1995. Basic Econometrics, third edit. MC-Graw-Hill.Inc., USA, (1995), pp. 712-713.

[16] Toprak, S., 2011. Çok Değişkenli Uyarlamalı Regresyon Eğrileri ve Konik Programlama ile Zaman Serilerinin Modellenmesi, Dicle University, Science Institue, Master Dissertation, 107, Diyarbakır.

[17] Box, G.E.P., Jenkins, G.M., 1976. Time series analysis forecasting and control, California.

[18] Özmen, A., 1986. Zaman Serisi Analizinde Box-Jenkins Yöntemi ve Banka Mevduat Tahmininde Uygulama Denemesi, Anadolu University Press, 201, Eskişehir.

[19] Kayım, H., 1985. İstatistiksel Ön Tahmin Yöntemleri, Hacettepe University Faculty of Economics and Administrative Sciences Press, 160, Ankara.

[20] Tekin, V. N., 2006. SPSS Uygulamalı İstatistik Teknikleri, 296, Ankara.

[21] Department of Political Science and International Relations Posc/Uapp 816 2010. Time Series Regression. http://www.udel.edu/htr/Statistics/Notes816/class20.PDF, (23.10.2010).

[22] Evans, W., 2010. Durbin-Watson Significance Tables, http://www.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf, (24.10.2010).

[23] Hamit, A., 2010. Korelasyon Analizi, http://maden.karaelmas.edu.tr/hocalar/hamitaydin/Sunu_6ci_hafta.pdf, (18.05.2010).

[24] Kmenta, J., 1986. Elements of econometrics, second edition, Macmillan, 655, New York.

[25] People Emich, 2010, Autocorrelation, people.emich.edu/jthornton/text−files/Econ415_autocorrelation.doc, (13.03.2010).

[26] Aster, R., Borchers, B., Thurber, C., 2004. Parameter Estimation and Inverse Problems, Academic Press, 355.

[27] Kao, C., Chyu, C.L., 2003. Least-squares estimates in fuzzy regression analysis, Europen Journal of Operational Research, 148 (2003), 426-435.

[28] STATA, http://www.stata.com/, (28.12.2010).

[29] MATLAB, http://www.mathworks.com/products/matlab/tryit.html, (29.12.2010).

Appendix

CPCR Data

CPVR Data