Estimation of the Muskingum routing coefficients by using fuzzy regression

M. Spiliotis¹* and L. Garrote²
¹ Democritus University of Thrace, School of Engineering, Department of Civil Engineering, Section of Hydraulics, 12 Vasilissis Sofias Str., 67100 Xanthi, Greece
² Department of Civil Engineering: Hydraulics, Energy and Environment, Technical University of Madrid, ETSI Caminos. C/ Prof. Aranguren s/n 28040 Madrid, Spain
*e-mail: m.spiliotis@gmail.com

Abstract: The Muskingum model is a popular method to analyse the flood routing. Most of the methodologies for estimating the parameters of the routing are based on the distance between observed outflows and estimated outflows. In this work, for the first time, the parameters for a linear Muskingum model are determined by following a different principle. The coefficients are proposed to be determined based on the fuzzy linear regression principles. Therefore the scope of the fuzzy calibration is to assess the Muskingum routing parameters as symmetric fuzzy triangular numbers so that all the observed data must be included in the produced fuzzy band aiming at its minimum width. This formulation leads to a constrained optimization problem with unknown variables the centres and the widths of the Muskingum routing coefficients. Instead of the identification of the parameters K, x, the problem is oriented to the determination of the coefficients C1 and C2. Based on the problem itself the constant term of the usual fuzzy regression model has been eliminated. In order to reduce the fuzzy band a proposed mixed goal programming fuzzy regression analysis is also used. Thus, the produced fuzzy band can be reduced by enabling some observed data to be near the produced fuzzy curve. In any case the maximum observed outflow must be included within the produced fuzzy band.

Key words: Fuzzy Regression, Fuzzy sets and logic, Goal programming, Parameter estimation, Muskingum routing equation

1. INTRODUCTION

The Muskingum model is a popular method to analyse flood routing based on the continuity equation and a simplified version of the dynamics equation. Most of the methodologies for estimating the routing parameters are based on the comparison between observed and estimated outflows. Several techniques have been developed to achieve the calibration of the Muskingum routing parameters by using either traditional optimization techniques (e.g. Stephenson, 1979), or more advanced optimization techniques by using heuristics optimization processes (e.g. Ouyang et al., 2014).

In this work, for the first time, the parameters for a linear Muskingum model are determined by following a different principle. We propose to determine the coefficients through a fuzzy linear programming. Therefore the scope of the fuzzy calibration is to assess the Muskingum routing parameters as fuzzy numbers so that the observed data must be included in the produced fuzzy band aiming at its minimum width.

2. METHODOLOGY

The methods applied in this paper are presented in this section. The widely used Fuzzy Linear Regression Analysis is firstly presented. Next the Muskingum routing coefficients are calculated by using the fuzzy linear regression model of Tanaka. Finally, a novel method based on fuzzy regression enhanced by goal programming is proposed in this work in order to determine the Muskingum routing coefficients.
2.1 Fuzzy Linear Regression analysis

Fuzzy models are a useful approach to investigate complex physical phenomena such as the interconnection between adjacent watersheds (Tsakiris et al., 2006), the rainfall-runoff process, etc. The Fuzzy linear regression has been shown to be a useful tool to express functional relationships between variables, especially when data are not sufficient in number (Ganoulis, 1994). In contrast to the statistical regression, the fuzzy regression analysis has no error term since the uncertainty is incorporated in the model by means of fuzzy numbers (Spiliotis and Bellos, 2016).

Fuzzy regression analysis assumes a fuzzy functional relationship between the dependent and independent variables (Papadopoulos and Sirpi, 1999). The input data may be crisp or fuzzy. In this work the input data are considered to be crisp numbers. Firstly the fuzzy linear regression method, which was developed by Tanaka (1987), is examined. According to this approach, the problem of fuzzy linear regression is finally formulated as a constrained optimization problem. Additionally, in case of crisp data and fuzzy numbers, the problem is transformed into a linear programming problem (Tanaka et al., 1989; Tsakiris et al., 2006; Kitsikoudis et al., 2016; Spiliotis and Bellos, 2016). An ambiguous point in the formulation is that according to Tanaka et al. (1987) methodology, all the data must be included in the regression model.

The fuzzy linear regression model proposed by Tanaka (1987) and Tanaka et al. (1989) has the following form:

\[
\hat{Y}_j = \hat{A}_0 + \hat{A}_1 x_{ij} + ... + \hat{A}_n x_{nj},
\]

where \( n \) is the number of independent variables, \( m \) is the number of data, and \( \hat{A}_i = (a_i, w_i)_{TR} \) are symmetric fuzzy triangular numbers selected as coefficients (Fig. 1), which have the following membership function:

\[
\mu_{\hat{A}_i}(\alpha_i) = \begin{cases} 
1 - \frac{|a_i - \alpha_i|}{w_i}, & \text{if } a_i - w_i \leq \alpha_i \leq a_i + w_i \\
0, & \text{otherwise}
\end{cases}
\]

where \( a_i \) and \( w_i \) are the centres and the widths of the fuzzy coefficients, respectively. The general definition of fuzzy numbers can be found in Klir and Yuan (1995).

However, the constant term in our problem has no physical meaning and hence, it is omitted:

\[
\hat{Y}_j = \hat{A}_1 x_{ij} + ... + \hat{A}_n x_{nj}
\]

The model of fuzzy linear regression produces a fuzzy band which is calculated based on the extension principle of fuzzy sets and logic. In general, the extension principle enables us to define the crisp functions on a fuzzy domain and consequently the extension principle can be used in order to define the algebraic operations between fuzzy sets (e.g. Klir and Yuan, 1995). If the input data are crisp numbers, then the model of fuzzy linear regression can be interpreted mathematically by multiplying fuzzy numbers and crisp numbers as well as adding the fuzzy numbers. In case that the coefficients are fuzzy triangular numbers, the linearity remains also in the total regression output.

According to fuzzy arithmetic, the function \( Y_j \) has the following centre \((y_{a,j})\) and semi-width \((y_{c,j})\) (Tanaka, 1987):

Centre: \( y_{a,j} = a_1 x_{ij} + ... + a_n x_{nj} \)

(4.a)
Semi-width: \[ y_{c,j} = w_{c_1} x_{c,j} + \ldots + w_{c_l} x_{c,j} + \ldots + w_{c_l} x_{c,j} \]  

Figure 1. Fuzzy triangular symmetrical number.

The \( \alpha \) – cut set, which is a crisp set, of a fuzzy number \( A \) represented by \( [A]_\alpha \), is defined for \( 0 < \alpha \leq 1 \) as follows (Klir and Yuan, 1995):

\[
[A]_\alpha = \{ x | \mu_A(x) \geq \alpha, x \in \mathbb{R} \} 
\]

In case of \( \alpha = 0 \), the definition of the \( \alpha \)-cut (Eq. 5) is slightly changed, without the equality, to include the zero-cut which can be named also as the support set (Kitsikoudis et al., 2016; Buckley and Eslami, 2002; Spiliotis et al., 2016).

According to the conventional fuzzy linear regression model, the centres and the widths of the fuzzy coefficients could be determined by solving a linear programming problem with an objective function which can minimize the total spread of the fuzzy outputs, on condition that all the data given from the historical sample outputs \( y_j \) should be included in the estimated function \( Y_j \), that is:

\[
\mu_{Y_j} (y_j) \geq \alpha, j = 1, \ldots, m, 
\]

where \( \mu_Y \) is the membership function of \( Y_j \) (Papadopoulos and Sirpi, 1999; Kitsikoudis et al., 2016).

The above constraints arise from the concept of inclusion, the definition of which is provided below (e.g. Spiliotis and Bellos, 2016):

Definition: The inclusion of a fuzzy set \( A \) to the fuzzy set \( B \) with the associated degree \( 0 \leq \alpha \leq 1 \) is defined as follows:

\[
[A]_\alpha \subseteq [B]_\alpha 
\]

The above definition can also be used in order to apply fuzzy regression analysis in cases where the data are not crisp (as in this text), but fuzzy.

Based on the previous description, the fuzzy linear regression analysis is reduced to the estimation of the fuzzy coefficients \( A_j = (a_i, w_{c_i})_L, i=1,\ldots,n \) that minimizes the total spread of the fuzzy output subject to the constraints:

\[
J = \min \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} w_{c_i} |x_{c,j}| \right\} 
\]

As mentioned before, the constant term of the fuzzy regression is eliminated due to the fact that in the examined problem, it has no physical meaning. Hence, by selecting the \( \alpha \)-cut to check for the inclusion property and by using are symmetric fuzzy triangular numbers, the problem of selecting
the coefficients of the regression model leads to the following linear optimization problem:

\[
J = \min \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} |x_{ij}| \right\}
\quad \text{(9.a)}
\]

subjects to:

\[
y_j \geq \sum_{i=1}^{n} a_i x_{ij} - (1 - \alpha) \sum_{i=1}^{n} w_{ij} |x_{ij}|
\quad \text{(9.b)}
\]

\[
y_j \leq \sum_{i=1}^{n} a_i x_{ij} + (1 - \alpha) \sum_{i=1}^{n} w_{ij} |x_{ij}|
\quad \text{(9.c)}
\]

\[
w_{ij} \geq 0
\quad \text{(9.d)}
\]

where \(i=1,\ldots,n\) and \(j=1,\ldots,m\). Where \(n\) is the number of independent variables and \(m\) is the number of available data.

### 2.2 Estimation of the Muskingum routing coefficients by using the fuzzy regression model of Tanaka (1987)

In case that the relationship between storage and flow through a reach is linear, the Muskingum storage relationship can be written as:

\[
S = K \left[ xI + (1 - x)Q \right]
\quad \text{(10)}
\]

where \(I\) is the inflow rate to the reach, \(Q\) is the outflow rate from the reach, \(K\) the storage time constant for the reach and \(x\) a weighting factor that varies between 0 and 0.5 (Viessman and Lewis, 1996).

The mass balance equation can be expressed in discrete form as follows:

\[
\frac{(I_1 + I_2)}{2} \cdot \Delta t - \frac{(Q_1 + Q_2)}{2} \cdot \Delta t = (S_2 - S_1) \quad \text{(11)}
\]

By combining Eqs. 10 and 11 the Muskingum routing equation is achieved:

\[
Q_2 = C_0 I_2 + C_1 I_1 + C_2 Q_1 \quad \text{(12)}
\]

where

\[
C_0 = \frac{-(Kx - 0.5\Delta t)}{K - Kx + 0.5\Delta t},
\]

\[
C_1 = \frac{(Kx + 0.5\Delta t)}{K - Kx + 0.5\Delta t},
\]

\[
C_2 = \frac{(K - Kx - 0.5\Delta t)}{K - Kx + 0.5\Delta t}
\]

It is obvious that
A direct method of calibrating the Muskingum routing parameters can be achieved based on the fuzzy linear regression. Instead of the identification of the parameters $K$, $x$, the problem is oriented to the determination of $C_1$ and $C_2$ parameters. Therefore the scope of the fuzzy calibration is to assess the parameters $C_1$ and $C_2$ (as fuzzy numbers) so that all the data must be included in the produced fuzzy band aiming at its minimum width. The parameters $C_1$ and $C_2$ are considered to be fuzzy symmetrical triangular numbers, instead of crisp numbers.

After rearranging the Muskingum routing equation at time interval $j$, the examined fuzzy regression and by taking into account that $C_1 I_{j+1} + C_2 Q_j = Q_{j+1} - I_{j+1}$ will be the following:

$$C_1 I_{j+1} + C_2 Q_j = Q_{j+1} - I_{j+1}$$

This form, but without fuzziness, was investigated by Stephenson, 1979. The fuzzy linear regression-based model of Tanaka can be used to estimate the coefficients. Consequently, the fuzzy coefficients can be determined by solving the following constrained optimization problem:

$$J = \min \left\{ \sum_{j=1}^{m} \left( w_{c1} |I_j - I_{j+1}| + w_{c2} |Q_j - I_{j+1}| \right) \right\}$$

s.t.

$$a_{c1} (I_j - I_{j+1}) + a_{c2} (Q_j - I_{j+1}) + (1-a) \left( w_{c1} |I_j - I_{j+1}| + w_{c2} |Q_j - I_{j+1}| \right) \geq (Q_{j+1} - I_{j+1})$$

$$a_{c1} (I_j - I_{j+1}) + a_{c2} (Q_j - I_{j+1}) - (1-a) \left( w_{c1} |I_j - I_{j+1}| + w_{c2} |Q_j - I_{j+1}| \right) \leq (Q_{j+1} - I_{j+1})$$

where $a_{c1}$, $a_{c2}$ and $w_{c1}$, $w_{c2}$ the central values and the semi-widths of the fuzzy coefficients $C_1$ and $C_2$. In addition where $m$ is the number of data, whilst $n=2$ (number of independent variables). The term $(Q_{j+1} - I_{j+1})$ is considered as dependent variable. Thus, the produced fuzzy band with respect of the auxiliary independent variables $(I_j - I_{j+1})$, $(Q_j - I_{j+1})$ must contain all the available data.

2.3 Estimation of the Muskingum routing coefficients by using the fuzzy regression model enhanced by goal programming

Fuzzy linear regression is heavily influenced by the presence of outliers, which can extremely increase the uncertainty of the model (e.g., Peters 1994). Subsequently, several researchers attempted to modify the rigid inclusion constraints of the widely used fuzzy linear regression of Tanaka (1987) (Kitsikoudis et al., 2016). Here, we propose a modification of the fuzzy linear regression model enhanced with goal programming which is proposed by Kitsikoudis et al. (2016), based on the nature of the examined problem itself. The main idea is the use of the distance (error), $d^+$, which is associated with the over-estimation of the left-hand boundary and another distance measure (error), $d^-$, which is associated with the under-estimation of the right-hand boundary. As in goal programming extension (Appendix I), the proposed method enables the overstepping of the fuzzy boundaries at some points. Obviously, the model aims at minimizing the overstepping as well as the sum of the fuzzy spread. Therefore the extreme sensitivity of the original fuzzy regression model of Tanaka due to some "outliers" is reduced significantly by following the proposed methodology.
The basic model remains the same:
\[
\tilde{C}_1(I_j - I_{j+1}) + \tilde{C}_2(Q_j - I_{j+1}) = (Q_{j+1} - I_{j+1})
\]

However, a divergence (error) is permitted for each inequality constraint (A.II):
\[
\begin{align*}
& a_{i_1} (I_j - I_{j+1}) + a_{i_2} (Q_j - I_{j+1}) + (1 - a) \left[ w_{i_1} I_j - I_{j+1} + w_{i_2} Q_j - I_{j+1} \right] + d^+ \leq (Q_{j+1} - I_{j+1}) \\
& a_{i_1} (I_j - I_{j+1}) + a_{i_2} (Q_j - I_{j+1}) - (1 - a) \left[ w_{i_1} I_j - I_{j+1} + w_{i_2} Q_j - I_{j+1} \right] - d^- \leq (Q_{j+1} - I_{j+1}) \\
\end{align*}
\] (16.a)

\[ w_{i_1}, w_{i_2} \geq 0 \]

Finally, the scope of the final optimization problem is to minimize the total spread of components of the unit hydrograph as well as the sum of the error (distance) terms \( d^+_j, d^-_j \):
\[
J = \min \left\{ w_1 \cdot \sum_{j=1}^{m} (d^+_j + d^-_j) + w_2 \cdot \left( \sum_{j=1}^{m} (w_{i_1} I_j - I_{j+1} + w_{i_2} Q_j - I_{j+1}) \right) \right\} 
\]
\[ \text{in which}, \ d^+_j, d^-_j, w_1, w_2 \geq 0 \] (16.b)

By using the weight \( w_j \), the divergence from the fuzzy band is penalized, while, by using the weight \( w_2 \), the spread of the produced fuzzy band is penalized (Kitsikoudis et al., 2016).

3. CASE STUDY

The case study is based on the inflow and outflow rate hydrographs exhibiting linear characteristics (Viessman and Lewis, 2003; Al-Humoud and Esen, 2006). The data can be found at Al-Humoud and Esen (2006). By following the traditional approach of Tanaka (1987) with \( \alpha = 0 \) (that is, the zero-cut), the following curve is produced:
\[
\left( 0.0133, 0.7964 \right)_T \cdot \left( I_j - I_{j+1} \right) + \left( 0.5174, 0.0624 \right)_T \cdot \left( Q_j - I_{j+1} \right) = \left( Q_{j+1} - I_{j+1} \right) 
\] (17)

where the first number in the bracket shows the central value and second value denotes the semi-width of the fuzzy symmetrical triangular number. Unfortunately, the calculation of the \( K, x \) parameters based on the central values of \( C_1 \) and \( C_2 \) can not be achieved. It is easy to find the value of \( Q_{j+1} \) from Eq.17 because \( I_{j+1} \) is a crisp number. Figure 2 presents the outflow rate and the predicted fuzzy band for the same variable based on Eq. 17. Indeed, all the observed data are included in the produced fuzzy band. However, the fuzzy band is rather wide for some time steps and hence, the model can be characterised as not so functional.

Subsequently, the fuzzy regression method enhanced by goal programming is implemented for \( \alpha = 0 \) and the following curve is produced:
\[
\left( 0.4723, 0.2977 \right)_T \cdot \left( I_j - I_{j+1} \right) + \left( 0.4523, 0.06802 \right)_T \cdot \left( Q_j - I_{j+1} \right) = \left( Q_{j+1} - I_{j+1} \right) 
\] (18)

As it can be seen from Figure 3, even if, a small region does not belong to the produced fuzzy band, the maximum observed outflow belongs to the produced fuzzy band and in general, the predicted fuzzy band simulates the observed outflow. The spread of the fuzzy band is reduced almost to one half. Another interesting point of view, is that based on the central values of the parameters \( C_1 \) and \( C_2 \), \( K \) and \( x \) can be calculated (\( K = 1.69 \) and \( x = 0.21 \)). These values are not far from these from the least square method.
4. CONCLUSIONS

The consideration of a fuzzy version for the Muskingum routing model, instead of the usual precise (crisp) model, is closer to our uncertain understanding of nature. Thus, based on the Tanaka, 1987 method, the fuzzy linear regression is implemented to determine the Muskingum routing parameters. Based on the Tanaka model, the Muskingum routing parameters are determined as symmetric fuzzy triangular numbers, so that all the observed data must be included in the produced fuzzy band aiming to minimize its width.

However, the implementation of the Tanaka, 1987 method creates several problems, the most prominent among those being the large width of the produced fuzzy band. We have shown with our work that the determined fuzzy band can be significantly reduced, without a general cancelling of the inclusion property of the fuzzy regression, by using a goal programming based hybrid fuzzy regression approach.

Figure 2. Inflow and outflow hydrographs (data) and the predicted fuzzy band based on the Tanaka fuzzy linear regression model.

Figure 3. Inflow and outflow hydrographs (data) and the predicted fuzzy band based on our proposed method of fuzzy regression enhanced by goal programming.
REFERENCES


APPENDIX

In general terms, one of the most popular conventional forms of goal programming is the following:

\[
\max \sum_{j} (d_j^- + d_j^+) \\
\{(AX)\}_j + d_j^- - d_j^+ = b_j \text{ (to express the" = ")}
\]
other constrains

In which the term \((AX)_j + d_j^- - d_j^+ = b_j\), expresses the \(j^{th}\) soft equality constraints. The divergence from the equity constraint is measured by the corresponding distances \(d_j^-\) and \(d_j^+\).

However, we can extend the cases that we can use the goal programming when there are inequalities which are enabled to be partially (softly) satisfied. In this case, we can also use the goal programming approach by using only one distance of divergence for each equality constraint as follows (Kitsikoudis et al., 2016):

\[
\{(AX)_j^+ - d_j^+ \leq b_j \text{ (to express the" \leq ")}
\]
\[
\{(AX)_j^+ + d_j^- \geq b_j \text{ (to express the" \geq ")}
\]

In the same way with the conventional goal programming, the objective function is the sum of the inequality diversions.