

Fuzzy Linear Regression of Rainfall-Altitude Relationship [†]

Christos Tzimopoulos, Christos Evangelides, Christos Vrekos and Nikiforos Samarinas *

Department of Rural and Surveying Engineering, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; ctzimop@gmail.com (C.T.); evan@vergina.eng.auth.gr (C.E.); chvrekos@yahoo.gr (C.V.)

* Correspondence: samnikiforos1@gmail.com; Tel.: +30-6977-351770

[†] Presented at the 3rd EWaS International Conference on “Insights on the Water-Energy-Food Nexus”, Lefkada Island, Greece, 27–30 June 2018.

Published: 27 August 2018

Abstract: Classical linear regression has been used to measure the relationship between rainfall data and altitude in different meteorological stations, in order to evaluate a linear relation. The values of rainfall are supposed as dependent variables and the values of elevation of each station as independent variables. It has long been known that a classical statistical relationship exists between annual rainfall and the station elevation which in many cases is linear as the one examined in this article. However classical linear regression makes rigid assumptions about the statistical properties of the model, accepting the error terms as random variables, and the violation of this assumption could affect the validity of the classical linear regression. Fuzzy regression assumes ambiguous and imprecise parameters and data. For this reason it may be more effective than classical regression. In this paper we evaluate the relationship between annual rainfall data and the elevation of each station in Thessaly’s meteorological stations, using fuzzy linear regression with trapezoidal membership functions. In this possibilistic model the dependent measured elevations are crisp, and the independent observed rainfall values as well as the parameters of the model are fuzzy.

Keywords: fuzzy regression; trapezoidal parameters; fuzzy linear programming; possibilistic models

1. Introduction

The scope of constructing in engineering a model is always to attempt to maximize its usefulness. This aim is closely connected with the relationship among three key characteristics of every systems model: *complexity*, *credibility*, and *uncertainty*. Our purpose in systems modelling is to estimate an optimal level of uncertainty for each modelling problem. In 1965 Zadeh in [1], introduced the theory of fuzzy sets, in which the fuzziness of a system incorporated all the types of uncertainty, in contrast with probability theory, which was capable of representing only one of several distinct types of uncertainty.

In Hydrology, rainfall measurement models have been extensively used in the design process of water resource projects such as hydrological prediction, spillway design, climatic change studies, rainfall and runoff correlation etc. Rainfall measurements in a specific area are commonly displayed in the form of time series, where recorded values can be either continuous or discrete. In many instances, there is a correlation between rainfall data that belong to different stations and comprise measurements with differing range and stations elevations. Correlation analysis is used to depict the relation between the independent variable (usually the meteorological station elevations) and the dependent variables (meteorological stations rainfall data).

In classical linear regression, the difference between measurement values and estimated values, is a random variable with normal distribution and is considered to be caused by measurement errors. Upper and lower bounds of the estimated value are calculated and the probability that the estimated value will lie between them represents the estimation confidence. According to this, classical regression is considered to be probabilistic and has many uses, but can be rendered problematic if the data set is small, if it's hard to prove that error distribution is normal, if there is fuzziness between dependent and independent variables or if linearity acceptance is not proper [2].

Nowadays, new regression models have been introduced, based on fuzzy logic [3–13]. In fuzzy regression the difference between measurement values and estimated values is attributed to the inherent fuzziness of the system as well as to the fuzziness of input and output data. In contrast with classical regression analysis, fuzzy regression analysis uses fuzzy functions for the regression factors and usually meets one of the three cases: (a) Crisp input values x_{ij} and crisp output values y_j (CICO); (b) Crisp input values x_{ij} and fuzzy output values \tilde{y}_j (CIFO); and (c) Fuzzy input values \tilde{x}_{ij} and fuzzy output values \tilde{y}_j (FIFO). In all of these three cases, estimated values \tilde{Y}_j are fuzzy.

Most of the cited above writers have used tridiagonal fuzzy functions for the formulation of the problem. However the need to use trapezoidal functions, results of the following reasons [14]: (a) we need to optimize the fuzziness of the model and (b) we need to restrict experimental data inside the estimated value range. Using trapezoidal membership functions for estimated values [14–18] allows us to achieve inclusion for output data with triangular membership functions \tilde{y}_j and estimated values with trapezoidal membership functions \tilde{Y}_j , for confidence level $h = 1$, for which the kernel is not minimized in a point: $[\tilde{y}_j]_{h=1} \subseteq [\tilde{Y}_j]_{h=1}$. In addition, for a level of confidence $h = 0$, we can achieve inclusion: $[\tilde{y}_j]_{h=0} \subseteq [\tilde{Y}_j]_{h=0}$. Due to the linearity of the membership function, inclusion for those levels of confidence, allows us to ensure that inclusion is possible for every level of confidence: $[\tilde{y}_j]_h \subseteq [\tilde{Y}_j]_h, \forall h \in [0,1]$. Consequently, trapezoidal fuzzy parameters have higher capability to model varieties of fuzziness than triangular fuzzy parameters.

Trapezoidal membership function models have been used by [14,15,19–21,22–24]. Charfeddine in [19] extended Tanaka's method for the case of trapezoidal membership functions with crisp measured input and output values. She used a fuzzy level function \tilde{f} , with four crisp level functions f_a, f_b, f_c, f_d . For f_a, f_b , she used Tanaka's method [8], whereas for f_b, f_c , she used classical linear regression, which led to function \tilde{f} and standard deviation σ of the model. Based on this f_b, f_c , become: $f_b = \tilde{f} - \lambda\sigma, f_c = \tilde{f} + \lambda\sigma$ with λ being an adjustment factor. Ganesan and Veeramani in [22], described a fuzzy linear programming with symmetric trapezoidal fuzzy parameters. They proved fuzzy analogues of some important theorems of linear programming and they gave a numerical example, leading to a solution of fuzzy linear programming problems. Fung et al. in [25], proposed a new model for asymmetric trapezoidal fuzzy parameters. Kumar and Kaur in [24] presented a new method, called Mehar's method, suitable for solving fuzzy linear programming problems with trapezoidal fuzzy parameters. They proved that this method is easy to apply to fuzzy linear programming problems, with respect to existing methods. Kheirfam and Verdegay in [23] extended the dual simplex method to fuzzy linear programming, with symmetric trapezoidal fuzzy parameters. They studied the variation of values to a certain limit, so that the fuzzy optimal solution remains invariant.

In this article, a possibilistic model (Tzimopoulos et al. model) is described, where membership functions are trapezoidal, measured input values are crisp and measured output values are fuzzy triangular [13]. In this model, a rather simple two-phase method was used, with one crisp measured input value and one fuzzy measured output value. During step 1, measured input and output values were considered crisp, while parameters were considered fuzzy and supports were estimated using Tanaka's method. In step 2, estimated supports were considered to be the known kernel of the trapezoidal membership functions and in the inclusion they were transferred to the known terms. Supports for the trapezoidal membership functions of the estimation were calculated. Triangular membership functions were used for measured output values.

Further, we present one application of the above model, concerning a hydrological problem in the region of Central Thessaly (Greece), where twenty rainfall measurement stations of the region with rainfall data and their elevation have been used.

2. Mathematical Model

2.1. Model of Tzimopoulos et al. (2006)

2.1.1. Generalities

Measured input values x_{ij} are considered crisp and measured output values are considered symmetrical triangular fuzzy numbers $\tilde{y}_j = (y_j^-, y_j^k, y_j^+)$ (Figure 1), while estimated values are considered trapezoidal fuzzy numbers $\tilde{Y}_j = (S_j^-, K_j^-, K_j^+, S_j^+)$ (Figure 2). The above problem can be divided into the following two steps:

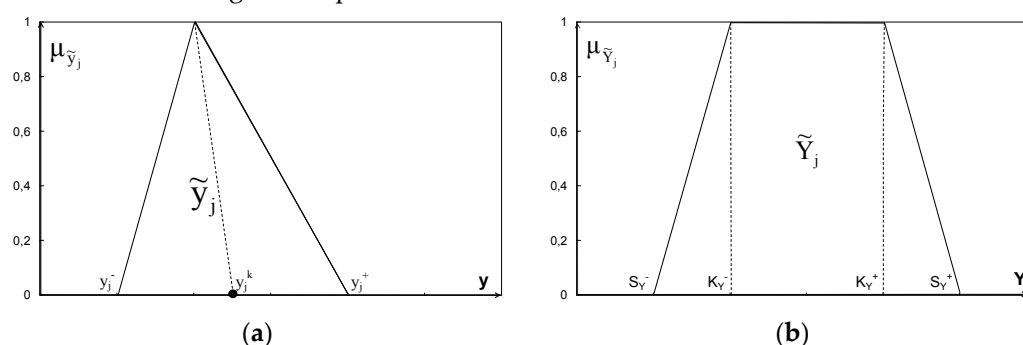


Figure 1. (a) Triangular membership function, (b) Trapezoidal membership function.

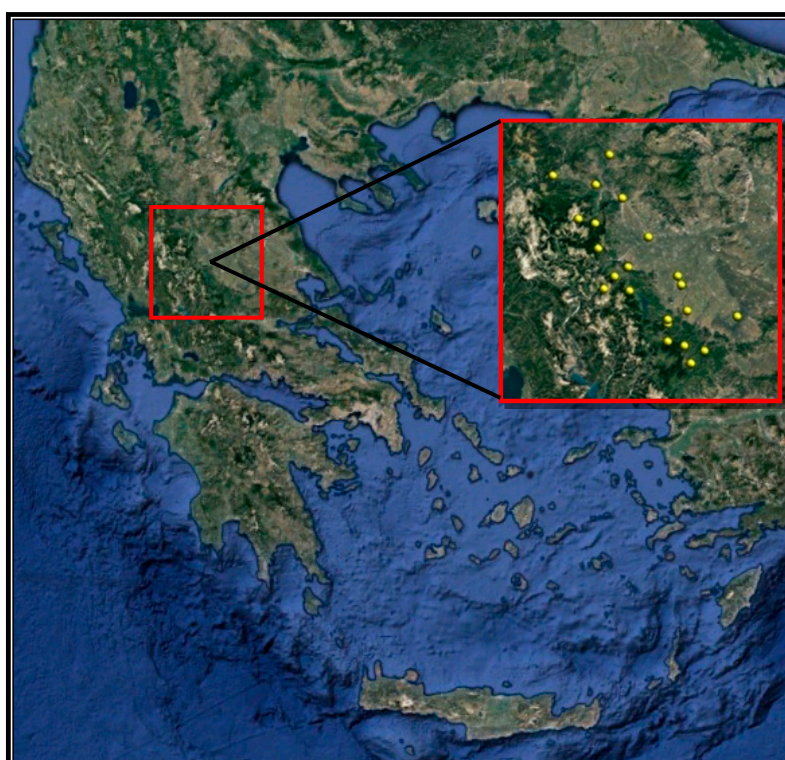


Figure 2. Central Thessaly meteorological stations.

2.1.2. Step 1

According to the kernel inclusion constraint mentioned above, we have:

$$y_j^k \in [Y_j^-, Y_j^+]^{h=1} \quad (1)$$

Namely, the kernel of experimental output values is within the kernel of estimated values. According to Shapiro et al. (2009), for the case of crisp output values (only the kernels of measured values), if we apply Tanaka's method (Tanaka 1987), the range $[Y_j^-, Y_j^+]^{h=1} = [K_j^-, K_j^+]^{h=1}$ encircles the kernels of experimental values. In this stage, only triangular functions $\tilde{A}_0(r_0, c_0), \tilde{A}_1(r_1, c_1)$ are applied for the method of Tanaka and the possibilistic model used is: $\tilde{Y}_j = \tilde{A}_0 + \tilde{A}_1 x_{1j}$. Thus, the problem of determining the estimated values in this step is:

$$\min(c) = M c_0 + c_1 \sum_{i=1}^M x_{1j} \quad (2)$$

$$Y_j^+ = K_j^+ = \sum_{i=0}^1 r_i x_{ij} + \sum_{i=0}^1 c_i x_{ij} \geq y_j^k \quad (3a)$$

$$Y_j^- = K_j^- = \sum_{i=0}^1 r_i x_{ij} - \sum_{i=0}^1 c_i x_{ij} \leq y_j^k, x_{0j} = 1 \quad (3b)$$

$$c_0, c_i \geq 0 \quad (3c)$$

Through the solution of this system we get the surroundings $[Y_j^-, Y_j^+]^{h=1} = [K_j^-, K_j^+]^{h=1}$ as follows:

$$Y_j^+ = (r_0 + c_0) + (r_1 + c_1)x_j \quad (4a)$$

$$Y_j^- = (r_0 - c_0) + (r_1 - c_1)x_j \quad (4b)$$

As long as they meet the same constraints, these surroundings coincide with the kernel of trapezoidal functions, resulting in the relations below:

$$\begin{aligned} Y_j^- = K_j^- &= K_{\tilde{A}_0}^- + K_{\tilde{A}_1}^- x_{1j}, & j = 1, \dots, M \\ Y_j^+ = K_j^+ &= K_{\tilde{A}_0}^+ + K_{\tilde{A}_1}^+ x_{1j}, & j = 1, \dots, M \end{aligned} \quad (5)$$

2.1.3. Step 2

We apply now support inclusion:

$$[S_{\tilde{y}_j}^-, S_{\tilde{y}_j}^+] \subseteq [S_{\tilde{Y}_j}^-, S_{\tilde{Y}_j}^+] \quad (6)$$

where space $[S_{\tilde{y}_j}^-, S_{\tilde{y}_j}^+]$ is given as follows:

$$[S_{\tilde{y}_j}^- = y_j^- = y_j - e_j, \quad S_{\tilde{y}_j}^+ = y_j^+ = y_j + e_j] \quad (7)$$

Based on relations (3a, 3b, 3c) and (4a, 4b), we get:

$$K_{\tilde{Y}_j}^- - \sum_{i=1}^2 c_i^l x_{ij} \leq y_j^-, \quad x_{1j} = 1 \quad (8)$$

$$K_{\tilde{Y}_j}^+ + \sum_{i=1}^2 c_i^r x_{ij} \geq y_j^+, \quad x_{1j} = 1, \quad j = 1, 2, \dots, M \quad (9)$$

where $K_{\tilde{Y}_j}^+$ and $K_{\tilde{Y}_j}^-$ are known since they have been calculated during step 1.

The problem turns now into the following:

$$\min(c) = M(c_0^\ell + c_0^r) + (c_1^\ell + c_1^r) \sum_{j=1}^M x_{1j} \quad (10)$$

s.t

$$\begin{aligned} K_{\tilde{y}_j}^- - \sum_{i=1}^2 c_i^r x_{ij} &\leq \tilde{y}_j^- - e_j, \quad x_{0j} = 1 \\ K_{\tilde{y}_j}^+ + \sum_{i=1}^2 c_i^r x_{ij} &\geq \tilde{y}_j^+ + e_j, \quad x_{0j} = 1, \quad j = 1, 2, \dots, M \\ c_0^\ell, c_0^r, c_1^\ell, c_1^r &\geq 0 \end{aligned} \quad (11)$$

3. Applications

3.1. Generalities

Given the following rainfall and elevation data of twenty meteorological stations of central Thessaly (Greece) (Tables 1 and 2), with crisp input data and symmetric output fuzzy data, we want to assess the two step model of Tzimopoulos et al., using the fuzzy regression method, considering trapezoidal membership functions for the assessment.

3.2. Step 1

We apply Tanaka's model, considering that output data are crisp. For, the model is written as follows:

$$\min(20c_0 + 10926c_1) \quad (12)$$

s.t

$$\begin{aligned} r_0 - c_0 + 90r_1 - 90c_1 &\leq 740.7 \\ r_0 - c_0 + 95r_1 - 95c_1 &\leq 611.1 \\ &\dots \dots \dots \\ r_0 - c_0 + 992r_1 - 992c_1 &\leq 1594.1 \\ r_0 - c_0 + 1000r_1 - 1000c_1 &\leq 1384.9 \\ r_0 + c_0 + 90r_1 + 90c_1 &\geq 740.7 \\ r_0 + c_0 + 95r_1 + 95c_1 &\geq 611.1 \\ &\dots \dots \dots \\ r_0 + c_0 + 992r_1 + 992c_1 &\geq 1594.1 \\ r_0 + c_0 + 1000r_1 + 1000c_1 &\geq 1384.9 \end{aligned} \quad (13)$$

Solving this problem we get kernel equations:

$$[Y_j^-, Y_j^+] = [y_\ell, y_u], \text{ or} \quad (14a)$$

$$y_\ell = 529.493 + 0.42706x, \quad y_u = 800.989 + 0.79956x \quad (3b)$$

Table 1. Elevation and Rainfall data from meteorological stations of Central Thessaly.

M. Stations	Kallifonio	Grammatiko	Karditsomag.	Karditsa	Trikala	Kalamapaka	Mouza. Ypex	Rahoula	Meg_Ker_Ypex	Agiofyllo
X (m)	90	95	95	103	110	202	226	330	500	581
Y (mm)	741	611	694	573	707	867	679	1065	864	823
e = 0.15Y	111,11071	91,66502743	104,089286	86,02214	106,06393	130,10625	101,859934	159,726	129,6461372	123,495

Table 2. Elevation and Rainfall data from meteorological stations of Central Thessaly.

M. Stations	Loutr._Ypex	Amarantos	Moloha	Bathylakkos	Malakasio	Brontero	Rentina	Chrisomilia	Argithea	Kerasia
X (m)	730	744	790	800	842	853	903	940	992	1000
Y (mm)	841	1172	1348	1118	1045	1475	1112	1294	1594	1384,9
e = 0.15Y	126,18743	175,8080864	202,244286	167,6779	156,731302	221,27336	166,768142	194,169	239,1214286	207,7294

Notation: e means the spread of the rainfall data, which are symmetrical triangular fuzzy numbers.

3.3. Step 2

In phase 2, the kernel is considered to be known and the model is written as follows:

$$\min(20c_0^\ell + 10926c_1^\ell + 20c_0^u + 10926c_1^u) \quad (4)$$

$$\text{s.t}$$

$$-c_0^\ell - 90c_1^\ell \leq 61.70$$

.....

$$-c_0^\ell - 1000c_1^\ell \leq 220.5 \quad (5)$$

$$c_0^u + 90c_1^u \geq -21.10$$

.....

$$c_0^u + 1000c_1^u \geq -8.00$$

In Figure 3 calculation results are shown, while support equations are:

$$[S_{\tilde{y}_j}^-, S_{\tilde{y}_j}^+] = [y_\ell, y_u], \text{ or} \quad (6a)$$

$$y_\ell = 450.069 + 0.3629x, \quad y_u = 887.100 + 1.20226x \quad (17b)$$

In the Figure 4a illustrates the estimated output data, Figure 4b illustrates the measured and predicted values at Megali_Kerasia_Ypex meteorological station, while Figure 3a, b illustrate the predicted parameters \tilde{A}_0 and \tilde{A}_1 respectively.

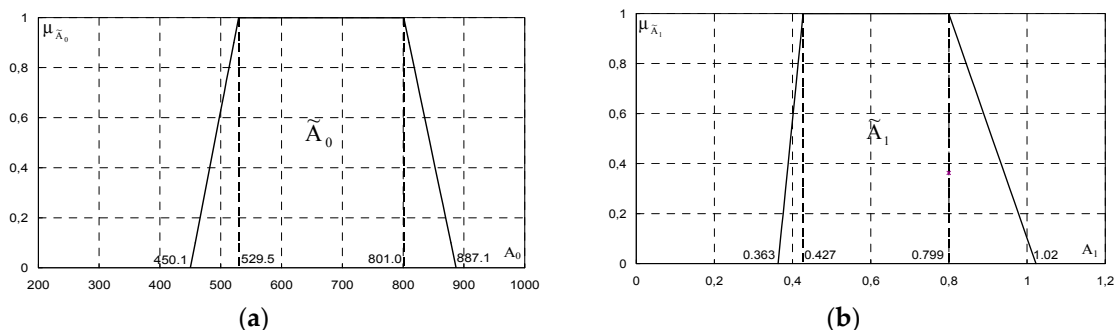


Figure 3. (a) Predicted parameters \tilde{A}_0 , (b) Predicted parameters \tilde{A}_1 .

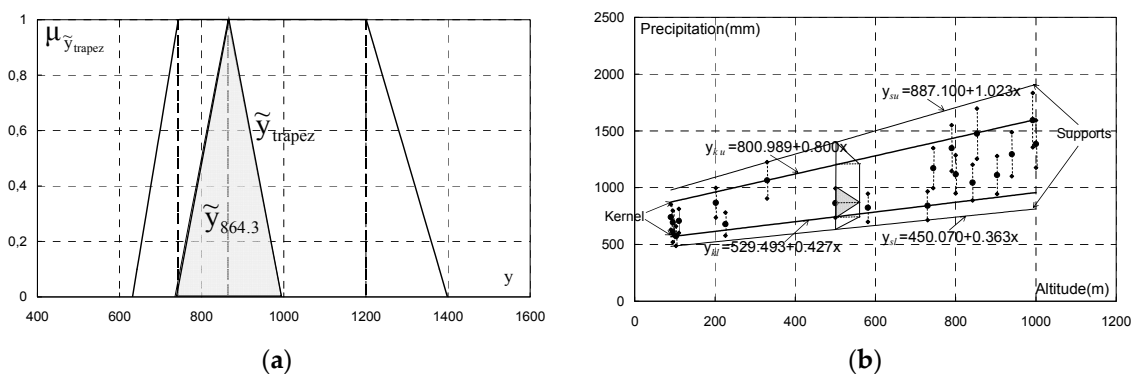


Figure 4. (a) Estimated fuzzy output data, (b) Measured and predicted output at Meg_Ker_Ypex station.

4. Comments-Conclusions

The two-step model has the advantage of using only four unknown quantities during each phase, in contrast with Bissierier and Fung models, which use eight.

In all the above cases, the presented model can be applied for all confidence levels and have the following property: data kernels are included into estimated kernels and data supports are also included into estimated supports.

In the case of rainfall measurement observations, station association is achieved, even for small samples and we can extend the shorter time series, due to fuzzy correlation for more rainfall stations.

References

1. Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353.
2. Savic, D.A.; Pedrycz, W. Evaluation of fuzzy linear regression models. *Fuzzy Sets Syst.* **1991**, *39*, 51–63.
3. Papadopoulos, B.; Sirpi, M. Similarities in Fuzzy Regression Models. *J. Optim. Theory Appl.* **1999**, *102*, 373–383.
4. Papadopoulos, B.; Sirpi, M. Similarities and Distances in Fuzzy Regression Models. *Soft Comput.* **2004**, *8*, 556–561.
5. Redden, D.T.; Woodall, W.H. Further examination of fuzzy linear regression. *Fuzzy Sets Syst.* **1996**, *79*, 203–211.
6. Shapiro, A.F.; Berry-Stölzle, T.R.; Koissi, M.-C. The Fuzziness in Regression Models. ARC2009_Shapiro_06 PDF. 2009; p. 6.
7. Tanaka, H.; Uejima, S.; Asai, K. Linear regression analysis with fuzzy models. *IEEE Trans. Syst. Man Cybern.* **1982**, *12*, 903–907.
8. Tanaka, H. Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets Syst.* **1987**, *24*, 363–375.
9. Tanaka, H.; Watada, J. Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets Syst.* **1988**, *27*, 275–289.
10. Tanaka, H.; Hayashi, I.; Watada, J. Possibilistic linear regression analysis for fuzzy data. *EJOR* **1989**, *40*, 389–396.
11. Tanaka, H.; Ishibuchi, H. Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters. *Fuzzy Sets Syst.* **1991**, *41*, 145–160.
12. Tzimopoulos, C.; Papadopoulos, B. *Fuzzy Logic with Application in Engineering*; Ziti: Thessaloniki, Greece, 2013; 627p. (In Greek)
13. Tzimopoulos, C.; Papadopoulos, K.; Papadopoulos, B. Fuzzy Regression with Applications in Hydrology. *IJEIT* **2016**, *5*, 69–75.
14. Bissierier, A. Une approche Paramétrique de la Régression Linéaire Floue-Formalisation par Intervalles. Ph.D. Thesis, Université de Savoie, Chambéry, France, 2010.
15. Bissierier, A.; Boukezzoula, R.; Galichet, S. Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals. In Proceedings of the IPMU'08, Torremolinos, Spain, 22–27 June 2008; pp. 181–188.
16. Bissierier, A.; Boukezzoula, R.; Galichet, S. Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals. *JUS* **2010**, *4*, 59–72.
17. Bissierier, A.; Boukezzoula, R.; Galichet, S. A revisited approach to linear fuzzy regression using trapezoidal fuzzy intervals. *Inf. Sci.* **2010**, *180*, 3653–3673.
18. Bissierier, A.; Boukezzoula, R.; Galichet, S. Linear Fuzzy Regression Using Trapezoidal Fuzzy Intervals. In Foundations of Reasoning under Uncertainty; Bouchon-Meunier, B., Magdalena, L., Ojeda-Aciego, M., Verdegay, J.L., Yager, R., Eds.; Springer: Berlin, Germany, 2010; Volume 249, pp. 2–22.
19. Charfeddine, S. Optimisation de l'offre d'une Compagnie Aérienne en Environnement Incertain. Thèse de Docteur; Université de Toulouse: Toulouse, France, 2004.
20. Charfeddine, S.; Mora-Camino, F.; De Coligny, M. Fuzzy linear regression: Application to the estimation of air transport demand. In Proceedings of the International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, Saint-Petersburg, Russia, 17–20 June 2004; pp. 350–359.
21. Charfeddine, S.; Zbidi, K.; Mora-Camino, F. Fuzzy Regression Analysis using Trapezoidal Fuzzy Numbers. In Proceedings of the EUSFLAT—LFA 2005, Barcelona, Spain, 7–9 September 2005; pp. 1213–1218.
22. Ganesan, K.; Veeramani, P. Fuzzy Linear Programs with Trapezoidal Fuzzy Numbers. *Ann. Oper. Res.* **2006**, *143*, 305–315.

23. Kheirfam, B.; Verdegay, J.L. The dual simplex method and sensitivity analysis for fuzzy linear programming with symmetric trapezoidal numbers. *Fuzzy Optim. Decis. Mak.* **2013**, *12*, 171–189.
24. Kumar, A.; Kaur, J. A New Method for Solving Fuzzy Linear Programs with Trapezoidal Fuzzy Numbers. *J. Fuzzy Set Valued Anal.* **2011**, *3*, 103–118.
25. Fung, R.Y.K.; Chen, Y.; Tang, J. Estimating the functional relationships for quality function deployment under uncertainties. *Fuzzy Sets Syst.* **2006**, *157*, 98–120.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).