

Παλινδρόμηση με Ασαφή Μοντέλα

Χ. Τζιμόπουλος

Ομότιμος Καθηγητής, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, 54124, Θεσσαλονίκη

Περίληψη: Η ανάλυση συσχέτισης παρατηρήσεων φυσικών μεγεθών χρησιμοποιείται για να απεικονίσει τη σχέση μεταξύ μιας εξαρτημένης μεταβλητής (συνήθως μια ομάδα Y με τα ίδια χαρακτηριστικά καταγραφής δεδομένων) και των ανεξαρτήτων μεταβλητών (ομάδες X_1, X_2, \dots, X_n με διαφορετικά χαρακτηριστικά καταγραφής). Σε αντίθεση με την κλασική παλινδρόμηση, όπου η διαφορά μεταξύ των παρατηρηθεισών τιμών και των εκτιμηθεισών τιμών ϵ_i , θεωρείται ότι οφείλεται σε λάθη παρατήρησης (τυχαία μεταβλητή), έχουν αναπτυχθεί σήμερα νέα μοντέλα παλινδρόμησης με τη βοήθεια της Ασαφούς Λογικής, όπου η παραπάνω διαφορά θεωρείται ότι οφείλεται στην ενυπάρχουσα ασάφεια του συστήματος, αλλά επίσης και στην ασάφεια των δεδομένων εισόδου και εξόδου. Στην παρούσα εργασία επιχειρείται η συσχέτιση μεταξύ των βροχομετρικών παρατηρήσεων των μετεωρολογικών σταθμών Αγγίστρου και Άνω Βροντούς, στην περιοχή Σερρών, και η ασάφεια του συστήματος εξετάζεται α) με ένα δυνατοτικό μοντέλο γραμμικής παλινδρόμησης, στο οποίο οι εισοδοί (παρατηρηθείσες τιμές) θεωρούνται σαφείς, ενώ οι παράμετροι του μοντέλου και η έξοδος (παρατηρηθείσα τιμή) είναι ασαφείς και β) με ένα μοντέλο ελαχίστων τετραγώνων.

Λέξεις κλειδιά: Ασαφής γραμμική παλινδρόμηση, βροχομετρικές παρατηρήσεις, επεξεργασία αυτών, δυνατοτικό μοντέλο, μοντέλο ελαχίστων τετραγώνων.

Regression using Fuzzy Models

C. Tzimopoulos

Aristotle University of Thessaloniki, Thessaloniki 54124

Abstract: Classical regression has been used to measure the relationship between experimental data in different groups, in order to evaluate a relation and to predict the values of one group (dependent variables), from the values of the other groups (independent variables). Classical 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 regression. Fuzzy regression assumes ambiguous and imprecise parameters and data and may be more effective than classical regression. In this paper we evaluate the relation between rainfall data in Aggistron and Ano Vrontou

meteorological stations, using a) fuzzy linear regression with a possibilistic model where the dependent values Y (observed rainfall values) are crisp, and the independent values (observed rainfall values) as well as the parameters of the model are fuzzy and b) a least squares model.

Key words: Fuzzy linear regression, rainfall data, application, possibilistic model, least squares model.

1. Εισαγωγή

Οι βροχομετρικές παρατηρήσεις μιας περιοχής εμφανίζονται συνήθως με μορφή χρονοσειρών και οι καταγραφόμενες τιμές είναι άλλοτε συνεχείς και άλλοτε διακριτές σε ίσα χρονικά διαστήματα. Οι διακριτές μεταβλητές σχετίζονται με τη συχνότητα με την οποία η βροχόπτωση εμφανίζεται στη διάρκεια κάποιας συγκεκριμένης περιόδου, ενώ οι συνεχείς μεταβλητές καταγράφονται με τρόπο ώστε να παίρνουν οποιαδήποτε τιμή σε κάποια συνεχή κλίμακα. Κατά τη διάρκεια μιας χρονικής περιόδου σε ένα συγκεκριμένο βροχομετρικό σταθμό υπάρχει περίπτωση να χρειάζεται συμπλήρωση μιας τιμής σε κάποια χρονική στιγμή που για οποιοδήποτε λόγο δεν έχει καταγραφεί και η συμπλήρωση αυτή μπορεί να γίνει με τη βοήθεια προηγούμενων τιμών ή και επόμενων τιμών. Δηλαδή έχοντας διαθέσιμες τιμές μια χρονοσειράς βροχοπτώσεων X μέχρι τη χρονική στιγμή t_k , υπολογίζεται η τιμή της βροχόπτωσης στην επόμενη χρονική στιγμή t_{k+1} . Συχνά δύο χρονοσειρές βροχοπτώσεων που ανήκουν σε διαφορετικούς σταθμούς βρίσκονται σε κάποια συσχέτιση και αποτελούνται από διαφορετικό εύρος μετρήσεων. Π.χ. στον σταθμό Α υπάρχει μια διαθέσιμη χρονοσειρά 15 ετών, ενώ στον σταθμό Β υπάρχει μια διαθέσιμη χρονοσειρά 30 ετών. Εξ αιτίας της υπάρχουσας συσχέτισης συμπληρώνονται οι ελλείπουσες τιμές, ούτως ώστε να επιμηκυνθεί η μικρότερη διαθέσιμη χρονοσειρά. Βέβαια για τη συσχέτιση των δύο σταθμών υπάρχει η έννοια του μετεωρολογικού συστήματος και θα πρέπει οι συσχετιζόμενοι σταθμοί να θεωρούνται γειτονικοί (ΥΔΡΟΓΝΩΜΩΝ, 2010).

Η ανάλυση συσχέτισης χρησιμοποιείται για να απεικονίσει τη σχέση μεταξύ της εξαρτημένης μεταβλητής (συνήθως ο βροχομετρικός σταθμός με τη μικρότερη περίοδο καταγραφής δεδομένων) και των ανεξαρτήτων μεταβλητών (γειτονικοί σταθμοί με μεγάλη περίοδο καταγραφής). Οι μέθοδοι συσχέτισης που έχουν χρησιμοποιηθεί περιγράφονται ως: α) Πολυωνυμική, β) Απλή Γραμμική παλινδρόμηση, γ) Πολλαπλή Γραμμική παλινδρόμηση, δ) Εκθετική, ε) Λογαριθμική, στ) Κυβικές σπλήνες (Σκριμιζέας, 2014). Συνήθως χρησιμοποιείται ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης:

$$Y_i = a_0 + \sum_{j=1}^n a_j X_{ij} + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \sigma^2(\varepsilon_i) = \sigma^2, \quad \text{cov}(\varepsilon_i, \varepsilon_j) = 0.$$

Στην κλασσική γραμμική παλινδρόμηση η διαφορά μεταξύ των παρατηρηθεισών τιμών και των εκτιμηθεισών τιμών από το μοντέλο, θεωρείται ότι οφείλεται σε λάθη παρατήρησης και η διαφορά αυτή ε_i αποτελεί τυχαία μεταβλητή με κανονική κατανομή. Υπολογίζονται άνω και κάτω όρια της εκτιμηθείσας τιμής και η πιθανότητα ότι η εκτιμηθείσα τιμή θα κείται μεταξύ των δύο ορίων παριστάνει την εμπιστοσύνη της εκτίμησης. Η κλασσική λοιπόν παλινδρόμηση θεωρείται **πιθανοτική** (probabilistic), έχει πολλές εφαρμογές αλλά είναι προβληματική εάν το σύνολο των δεδομένων είναι μικρό, ή εάν είναι δύσκολο να αποδειχθεί ότι το σφάλμα έχει κανονική κατανομή, ή εάν υπάρχει ασάφεια μεταξύ ανεξαρτήτων και εξαρτημένων μεταβλητών ή εάν η παραδοχή γραμμικότητας είναι ακατάλληλη (Savic and Pedrycz, 1991).

Η ασαφής λογική (Zadeh, 1965) δίνει απαντήσεις στα παραπάνω προβλήματα και σήμερα έχουν αναπτυχθεί νέα μοντέλα παλινδρόμησης με τη βοήθειά της (Tanaka et al, 1982, Tanaka, 1987, Tanaka and Watada, 1988, Tanaka and Hayashi, 1989, Tanaka and Ishibuchi, 1991, Papadopoulos and Sirpi, 1999, 2004, Τζιμόπουλος και Παπαδόπουλος, 2013). Στην ασαφή παλινδρόμηση η παραπάνω διαφορά θεωρείται ότι οφείλεται στην ενυπάρχουσα ασάφεια του συστήματος, αλλά επίσης και στην ασάφεια των δεδομένων εισόδου και εξόδου. Η απάντηση για μια ειδική τιμή εισόδου βρίσκεται σε μια περιοχή δυνατών τιμών, δηλαδή η απάντηση μπορεί να πάρει οποιαδήποτε από τις δυνατές αυτές τιμές. Η ασαφής παλινδρομική ανάλυση χρησιμοποιεί ασαφείς συναρτήσεις για τους συντελεστές της παλινδρόμησης σε αντίθεση με την κλασσική παλινδρομική ανάλυση. Το παραπάνω πρόβλημα (Tanaka and Hayashi, 1989, Redden and Woodall, 1996) συνήθως έχει τις ακόλουθες περιπτώσεις: a) Σαφείς τιμές εισόδου x_{ij} και εξόδου y_j , b) Σαφείς τιμές εισόδου x_{ij} , ασαφείς τιμές εξόδου \tilde{y}_j και εκτίμησης \tilde{Y}_i , c) Ασαφείς τιμές εισόδου \tilde{x}_{ij} και εξόδου \tilde{y}_j , ασαφείς τιμές εκτίμησης \tilde{Y}_i . Σε όλες τις παραπάνω περιπτώσεις οι παράμετροι του μοντέλου είναι ασαφείς

Η προσαρμογή ενός μοντέλου ασαφούς παλινδρόμησης γίνεται με δύο γενικές μεθόδους:

I) Το δυνατοτικό μοντέλο (Tanaka et al, 1982, Tanaka, 1987, Redden and Woodall, 1996, Savic and Pedrycz, 1991 κ. ά.). Η ασαφής παλινδρόμηση είναι δυνατοτική (possibilistic, Γουάλλες, 2005) και η συνάρτηση συμμετοχής $\mu_{\tilde{F}}$ ενός ασαφούς αριθμού \tilde{F} θεωρείται ίση με την συνάρτηση κατανομής της δυνατότητας (possibility distribution function) $\pi_x(x)$. Ελαχιστοποιείται η ασάφεια του μοντέλου, λαμβάνοντας το ελάχιστο της ολικής απόκλισης των ασαφών παραμέτρων από τη μέση τιμή τους, υπό τον όρο ότι, οι πειραματικές τιμές κάθε δείγματος θα περιλαμβάνονται μέσα σε ένα ορισμένο διάστημα δυνατών τιμών.

II) Το μοντέλο των ελαχίστων τετραγώνων (Diamond, 1988, Chang and Lee, 1996, Yang and Liu, 2003). Ελαχιστοποιείται η απόσταση μεταξύ της εκτιμηθείσας τιμής εξόδου του μοντέλου \tilde{Y}_i και της παρατηρηθείσας τιμής εξόδου \tilde{y}_j . Η μέθοδος αυτή του Diamond θεωρείται ως επέκταση της κλασσικής προσέγγισης της γραμμικής παλινδρόμησης, στηριγμένη στην ιδέα της έρευνας της καλύτερης

επάρκειας του μοντέλου στα δεδομένα.

Γενικότητες για τους ασαφείς αριθμούς

Ένας ασαφής αριθμός αποτελεί ένα ασαφές σύνολο \tilde{A} , $\mu_{\tilde{A}}: \mathbb{R} \rightarrow [0,1]$, το οποίο είναι άνω ημι-συνεχές, κανονικοποιημένο και κυρτό, δηλαδή:

1. $[\tilde{A}]^h = \{x \mid \mu_{\tilde{A}}(x) \geq h\}$ αποτελεί ένα κλειστό διάστημα,
2. $\exists x, \mu_{\tilde{A}}(x) = 1$,
3. $\mu_{\tilde{A}}(\lambda x_1 + (1-\lambda)x_2) \geq \mu_{\tilde{A}}(x_1) \wedge \mu_{\tilde{A}}(x_2)$, $\lambda \in [0,1]$.

2. Δυνατοτικό μοντέλο

Έστω μια ασαφής εξαρτημένη μεταβλητή \tilde{Y}_j και x_{ij} οι ανεξάρτητες μεταβλητές που επηρεάζουν τη μεταβλητή \tilde{Y}_j . Το αποτέλεσμα της ασαφούς γραμμικής παλινδρόμησης θα είναι μια εξίσωση της μορφής: (Tanaka, 1987, Tanaka and Watada, 1988, Tanaka and Hayashi, 1989, Tanaka and Ishibuchi, 1991, κ.ά.),

$$\tilde{Y}_j = \tilde{A}_0 + \tilde{A}_1 x_{1j} + \tilde{A}_2 x_{2j} + \dots + \tilde{A}_n x_{nj} = \sum_{i=0}^n \tilde{A}_i x_{ij}, \quad x_{0j} = 1, \quad (1)$$

όπου οι μετρηθείσες τιμές εισόδου x_{ij} , αποτελούν κλασσικούς σαφείς αριθμούς και οι μετρηθείσες τιμές εξόδου y_j , αποτελούν κλασσικούς σαφείς αριθμούς ή και ασαφείς αριθμούς \tilde{y}_j . Οι παράμετροι $\tilde{A}_i = (r_i, c_i)$ θεωρούνται ασαφείς συμμετρικοί τριγωνικοί αριθμοί, με συνάρτηση συμμετοχής που δίνεται ως ακολούθως:

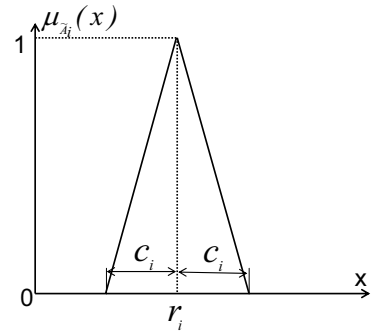
$$\mu_{\tilde{A}_i}(x) = \begin{cases} 0 & x < r_i - c_i \\ 1 + \frac{|x - r_i|}{c_i} & r_i - c_i \leq x \leq r_i + c_i \\ 0 & x > r_i + c_i \end{cases} \quad (2)$$

και παρίσταται γραφικά στο σχ. 1. Οι παράμετροι r_i, c_i , αποτελούν αντίστοιχα το μέσον και το εύρος (spread) της παραμέτρου \tilde{A}_i .

Οι συναρτήσεις συμμετοχής γράφονται συνήθως με τη βοήθεια των L, R , τύπων ως ακολούθως:

$$\mu_{\tilde{A}_i}(x) = \begin{cases} L\left(\frac{r_i - x}{c_i}\right) & r_i - c_i \leq x \leq r_i, \\ R\left(\frac{x - r_i}{c_i}\right) & r_i \leq x \leq r_i + c_i, \\ 0 & \text{αλλού,} \end{cases}$$

οι δε συναρτήσεις L, R , έχουν την ιδιότητα:



Σχήμα 1.

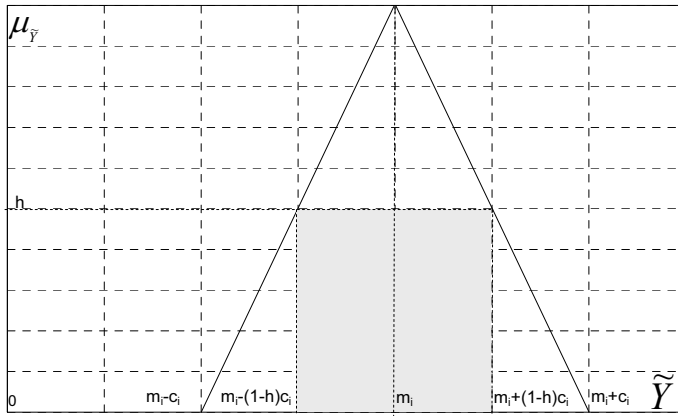
$L(0)=R(0)=1$, και $L(1)=R(1)=0$, και $L^{-1}(h)=R^{-1}(h)=1-h$.

Οι h -Τομές εμπιστοσύνης (α -Τομές), (Chang and Lee, 1996) της παραμέτρου \tilde{A}_i δίνονται ως ακολούθως:

$$[\tilde{A}_i]^h = [r_i - L^{-1}(h)c_i, r_i + L^{-1}(h)c_i], \quad (3)$$

ενώ οι h -Τομές εμπιστοσύνης της εξαρτημένης ασαφούς μεταβλητής \tilde{Y}_j δίνονται ως εξής:

$$[\tilde{Y}_j]^h = \left[\sum_{i=0}^n r_i x_{ij} - L^{-1}(h) \sum_{i=0}^n c_i x_{ij}, \sum_{i=0}^n r_i x_{ij} + L^{-1}(h) \sum_{i=0}^n c_i x_{ij} \right], \quad x_{0j} = 1, \quad (4)$$



Σχήμα 2. Περίπτωση για σαφή τιμή εξόδου y_i .

και το μέσον m_i και το εύρος c_i του ασαφούς αριθμού \tilde{Y}_j δίνονται ως ακολούθως:

$$m_i = \sum_{i=0}^n r_i x_{ij}, \quad c_i(h) = L^{-1}(h) \sum_{i=0}^n c_i x_{ij}. \quad (5)$$

Η παρατηρηθείσα τιμή y_i (για συμβατικούς αριθμούς) για βαθμό εμπιστοσύνης h , πρέπει να περιλαμβάνεται μέσα στην εκτιμηθείσα τιμή $[\tilde{Y}_j]^h$ (Moskowitz and Kim, 1993, Charfeddine et al, 2004). Επομένως θα ισχύει:

$$y_j \in [Y_L^h, Y_R^h] \quad \acute{\eta}$$

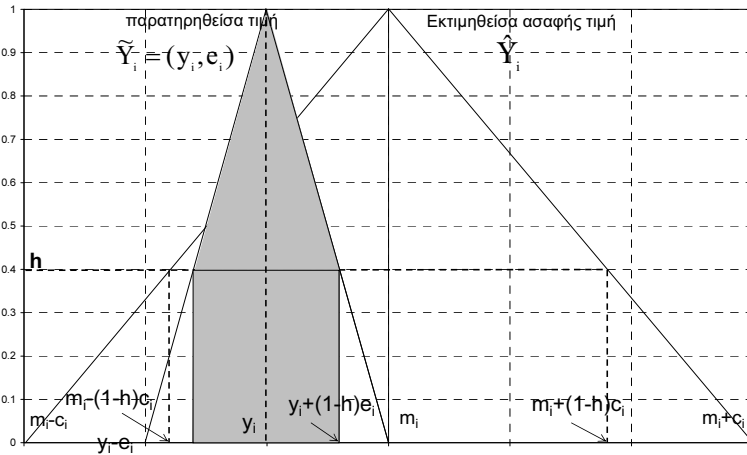
$$\sum_{i=0}^n r_i x_{ij} - L^{-1}(h) \sum_{i=0}^n c_i x_{ij} \leq y_j \leq \sum_{i=0}^n r_i x_{ij} + L^{-1}(h) \sum_{i=0}^n c_i x_{ij}, \quad x_{0j} = 1 \quad (6)$$

Επιπλέον θα πρέπει να ελαχιστοποιείται η ολική ασάφεια του μοντέλου με ελαχιστοποίηση του συνολικού εύρους (spread) των συντελεστών, και σύμφωνα με τους

Tanaka, (1987), Bisserier, (2010) κ.ά. το πρόβλημα αυτό καλείται ελάχιστο πρόβλημα (min. Problem):

$$\text{Min}(c) = \sum_{j=1}^m c_0 + \sum_{j=1}^m \sum_{i=1}^n c_i x_{ij} = mc_0 + \sum_{j=1}^m \sum_{i=1}^n c_i x_{ij}, c_i \geq 0. \quad (7)$$

Για τη γενική περίπτωση όπου η παρατηρηθείσα τιμή εξόδου \tilde{y}_i είναι ασαφής με μέση τιμή y_i και εύρος e_i (Σχήμα 3), θα ισχύει: $[\tilde{y}_j]^h \subseteq [Y_L^h, Y_R^h]$, και το παραπάνω πρόβλημα μορφοποιείται ως ένα πρόβλημα Γραμμικού Προγραμματισμού:



Σχήμα 3. Περίπτωση για ασαφή μετρηθείσα τιμή εξόδου \tilde{y}_i .

$$\min(c) = mc_0 + \sum_{j=1}^m \sum_{i=1}^n c_i x_{ij}, c_i \geq 0. \quad (8)$$

s.t.

$$\begin{aligned} \sum_{i=0}^n r_i x_{ij} + L^{-1}(h) \sum_{i=0}^n c_i x_{ij} &\geq y_j + (1-h)e_j, \\ \forall j=1,2,\dots,m \rightarrow (\bar{r}^T + L^{-1}(h)\bar{c}^T) \cdot X &\geq \bar{y} + L^{-1}(h)\bar{e}, \\ \sum_{i=0}^n r_i x_{ij} - L^{-1}(h) \sum_{i=0}^n c_i x_{ij} &\leq y_j - (1-h)e_j, \\ \forall j=1,2,\dots,m \rightarrow (\bar{r}^T - L^{-1}(h)\bar{c}^T) \cdot X &\leq \bar{y} - L^{-1}(h)\bar{e}, \end{aligned} \quad (9)$$

$$\begin{aligned} \bar{r}^T &= [r_0, r_1, \dots, r_n], \bar{c}^T = [c_0, c_1, \dots, c_n], \bar{y}^T = [y_0, y_1, \dots, y_n], \\ \bar{e}^T &= [e_0, e_1, \dots, e_n] \end{aligned}$$

$$X = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & \dots & x_{nm} \end{bmatrix}$$

Για την περίπτωση που έχουμε σαφείς τιμές εξόδου y_j , στις παραπάνω εξισώσεις τίθεται $e_i = 0$.

3. Μέθοδος ελαχίστων τετραγώνων

3.1. Ασαφείς συμμετρικοί τριγωνικοί αριθμοί

3.1.1 Απόσταση μεταξύ ασαφών αριθμών

Στα Μαθηματικά μια **μετρική** $d(x, y)$ (ή συνάρτηση απόστασης) αποτελεί μια συνάρτηση που ορίζει μία απόσταση μεταξύ των στοιχείων ενός συνόλου X (Giles, 1987). Το ζεύγος $(X, d(x, y))$ καλείται **μετρικός χώρος**. Μία μετρική προσδίδει μια τοπολογία σε ένα σύνολο, αλλά όλες οι τοπολογίες δεν μπορούν να δημιουργηθούν από μία μετρική. Ένας τοπολογικός χώρος έχοντας μία τοπολογία που περιγράφεται από μία μετρική καλείται μετρήσιμος. Μία μετρική στο σύνολο των συμβατικών αριθμών X είναι μία συνάρτηση $d: X \times X \rightarrow [0, +\infty]$, για την οποία ισχύουν οι ακόλουθες ιδιότητες:

1. $\forall x \in X, \quad d(x, x) = 0$
 2. \forall διακεκριμένα $x, y \in X, \quad d(x, y) > 0$
 3. (συμμετρία) $\forall x, y \in X, \quad d(x, y) = d(y, x)$
 4. (Τριγωνική ανισότητα) $\forall x, y, z \in X, \quad d(x, y) \leq d(x, z) + d(z, y)$
- (10)

Η δεύτερη ιδιότητα συνεπάγεται από τις άλλες. Πράγματι, εάν θέσουμε στην τριγωνική ανισότητα: $z=y, y=x$, η τριγωνική ανισότητα γίνεται:

$$d(x, x) = 0 \leq d(x, y) + d(y, x) \rightarrow 2d(x, y) \geq 0 \rightarrow d(x, y) > 0$$

3.1.2. Η μετρική του Hausdorff

Η μετρική του Hausdorff (Dubuisson and Jain, 1994) αποτελεί μια γενίκευση της απόστασης μεταξύ δύο σημείων σένα μετρικό χώρο από δύο συμπαγή πλήρη υποσύνολα του χώρου. Εάν A, B αποτελούν δύο συμπαγή πλήρη υποσύνολα των συμβατικών αριθμών, τότε η απόσταση του Hausdorff ορίζεται ως εξής:

$$d_H(A, B) = \max \{ \sup \inf d(A, B), \sup \inf d(B, A) \}$$
(11)

Η παραπάνω σχέση ερμηνεύεται ως ακολούθως:

- a) Για κάθε σημείο a του υποσυνόλου A βρίσκουμε την ελάχιστη απόστασή του από τα σημεία b του υποσυνόλου B .
- b) Μεταξύ όλων αυτών των ελαχίστων αποστάσεων βρίσκουμε την μέγιστη. Έτσι ορίζουμε τη συνάρτηση $h(A,B)=\sup \inf d(A,B)$.
- c) Επαναλαμβάνουμε την ίδια διαδικασία κατά την αντίθετη όμως φορά από το B στο A και βρίσκουμε τη συνάρτηση $h(B,A)=\sup \inf d(B,A)$.
- d) η απόσταση του Hausdorff ορίζεται πλέον ως εξής:

$$d_H(A,B) = \max\{h(A,B), h(B,A)\}. \quad (12)$$

όπου $A_\alpha = [A^L_\alpha, A^R_\alpha]$ είναι το α -Τομής σύνολο του ασαφούς αριθμού \tilde{A} , ενώ οι Zwick et al., (1987) προτείνουν την ακόλουθη μετρική: $d_1(\tilde{A}, \tilde{B}) = d(A_{\alpha=1}, B_{\alpha=1})$.

Ο Diamond (1988) πρότεινε την διαδικασία των ελαχίστων τετραγώνων για τον προσδιορισμό των ασαφών παραμέτρων, ορίζοντας μια μετρική μεταξύ δύο ασαφών αριθμών. Στο χώρο των τριγωνικών ασαφών αριθμών:

$$\tilde{X}_1^{\alpha=0} = (r_1 - c_1^\ell, r_1, r_1 + c_1^r), \tilde{X}_2^{\alpha=0} = (r_2 - c_2^\ell, r_2, r_2 + c_2^r), \quad (13)$$

ο Diamond ορίζει την ακόλουθη απόσταση d :

$$d(\tilde{X}_1, \tilde{X}_2)^2 = (r_2 - r_1)^2 + ((r_2 - c_2^\ell) - (r_1 - c_1^\ell))^2 + ((r_2 + c_2^r) - (r_1 + c_1^r))^2, \quad (14)$$

και όταν οι ασαφείς αριθμοί είναι συμμετρικοί: $\tilde{X}_1 = (r_1, c_1), \tilde{X}_2 = (r_2, c_2)$, ορίζει μια ισοδύναμη μετρική d_S ως ακολούθως:

$$d_S(\tilde{X}_1, \tilde{X}_2)^2 = (X_2^\ell - X_1^\ell)^2 + (X_2^r - X_1^r)^2 = ((r_2 - c_2) - (r_1 - c_1))^2 + ((r_2 + c_2) - (r_1 + c_1))^2, \quad (15)$$

Οι Yang and Ko(1997), Yang and Liu (2003) όρισαν μια μετρική ανάλογη με αυτή του Diamond ως ακολούθως:

$$d(\tilde{X}_1, \tilde{X}_2)^2 = (r_2 - r_1)^2 + ((r_2 - l_1 c_2^\ell) - (r_1 - l_1 c_1^\ell))^2 + ((r_2 + l_2 c_2^r) - (r_1 + l_2 c_1^r))^2, \quad (16)$$

όπου $l_1 = \int_0^1 L^{-1}(h)dh$, $l_2 = \int_0^1 R^{-1}(h)dh$.

Οι Ming et al (1997), Bargiela et al (2007), και οι Popescu and Giuclea(2007) για δύο αριθμούς \tilde{X}_1, \tilde{X}_2 , με α -Τομή, όρισαν την ακόλουθη μετρική:

$$d(\tilde{X}_1, \tilde{X}_2)^2 = \int_0^1 (X_1^\ell(\alpha) - X_2^\ell(\alpha))^2 d\alpha + \int_0^1 (X_1^r(\alpha) - X_2^r(\alpha))^2 d\alpha. \quad (17)$$

Οι Tran and Ducstein(2002) όρισαν την ακόλουθη μετρική:

$$d(\tilde{X}_1, \tilde{X}_2)^2 = \left[\frac{(r_1 - c_1^\ell) + (r_1 + c_1^r)}{2} - \frac{(r_2 - c_2^\ell) + (r_2 + c_2^r)}{2} \right]^2 + \frac{1}{3} \left[\frac{(r_1 + c_1^\ell) - (r_1 - c_1^r)}{2} + \frac{(r_2 + c_2^\ell) - (r_2 - c_2^r)}{2} \right]^2 \quad (18)$$

Οι Torfi et al (2011) όρισαν μια μετρική ανάλογη με αυτή του Diamond ως ακολούθως:

$$d(\tilde{X}_1, \tilde{X}_2)^2 = \frac{1}{3} \{ (r_2 - r_1)^2 + ((r_2 - c_2^\ell) - (r_1 - c_1^\ell))^2 + ((r_2 + c_2^\ell) - (r_1 + c_1^\ell))^2 \}, \quad (19)$$

Οι Kim et al (2008) και ο Nong (2012), όρισαν την ακόλουθη μετρική:

$$d(\tilde{X}_1, \tilde{X}_2)^2 = \left\{ \frac{(r_2 - c_2^\ell) + r_2 + (r_2 + c_2^\ell)}{3} - \frac{(r_1 - c_1^\ell) + r_1 + (r_1 + c_1^\ell)}{3} \right\}^2, \quad (20)$$

Οι παραπάνω μετρικές είναι ανάλογες της αρχικής μετρικής του Diamond και για το λόγο αυτό εδώ επιλέγεται η μετρική του Diamond (1988) για τους συμμετρικούς ασαφείς αριθμούς d_S , θεωρώντας τον συμμετρικό αριθμό εκτίμησης:

$$\begin{aligned} \tilde{Y}_i &= \tilde{A}_0 + \tilde{A}_1 x_i, \quad \tilde{A}_0(r_0, c_0), \quad \tilde{A}_1 = (r_1, c_1) \quad \eta \\ [\tilde{Y}_{i=0}]^{\alpha=0} &= [r_0 - c_0 + (r_1 - c_1)x_i, r_0 + c_0 + (r_1 + c_1)x_i]^{\alpha=0}. \end{aligned} \quad (21)$$

3.1.3 Μαθηματικό μοντέλο για ασαφείς συμμετρικούς αριθμούς

Θεωρούμε τώρα για τους δύο ασαφείς συμμετρικούς αριθμούς $(\tilde{Y}_i^{\alpha=0}, \tilde{y}_i^{\alpha=0})$ ένα πρόβλημα ελαχιστοποίησης ως ακολούθως:

$$\begin{aligned} \text{Minimize } S &= \sum d_S (\tilde{Y}_i^{\alpha=0}, \tilde{y}_i^{\alpha=0})^2 = \\ &= \sum_{i=1}^n \{ (y_i^\ell - (r_0 - c_0 + (r_1 - c_1)x_i))^2 + (y_i^r - (r_0 + c_0 + (r_1 + c_1)x_i))^2 \}. \end{aligned} \quad (22)$$

Η ελαχιστοποίηση της ποσότητας S επιτυγχάνεται εξισώνοντας τις παραγώγους της ποσότητας αυτής, ως προς τις παραμέτρους $(r_0 - c_0)$, $(r_0 + c_0)$, $(r_1 - c_1)$, $(r_1 + c_1)$, με το 0.

$$\begin{aligned} \frac{\partial S}{\partial (r_0 - c_0)} &= -2 \sum_{i=1}^n \{ (y_i^\ell - (r_0 - c_0 + (r_1 - c_1)x_i)) = 0 \\ &\rightarrow \bar{y}^\ell - (r_0 - c_0) - (r_1 - c_1)\bar{x} = 0, \\ \frac{\partial S}{\partial (r_1 - c_1)} &= -2 \sum_{i=1}^n x_i \{ (y_i^\ell - (r_0 - c_0 + (r_1 - c_1)x_i)) = 0 \\ &\rightarrow \sum_{i=1}^n x_i y_i^\ell - n\bar{x}(r_0 - c_0) - (r_1 - c_1) \sum_{i=1}^n x_i^2 = 0 \end{aligned} \quad (23)$$

$$r_1 - c_1 = \frac{\sum_{i=1}^n x_i y_i^\ell - n\bar{x}\bar{y}^\ell}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \quad r_0 - c_0 = \bar{y}^\ell - (r_1 - c_1)\bar{x} \quad (24)$$

$$\begin{aligned} \frac{\partial S}{\partial(r_0 + c_0)} &= -2 \sum_{i=1}^n \{(y_i^r - (r_0 + c_0 + (r_1 + c_1)x_i)) = 0 \\ &\rightarrow \bar{y}^r - (r_0 + c_0) - (r_1 + c_1)\bar{x} = 0, \\ \frac{\partial S}{\partial(r_1 + c_1)} &= -2 \sum_{i=1}^n x_i \{(y_i^r - (r_0 + c_0 + (r_1 + c_1)x_i)) = 0 \\ &\rightarrow \sum_{i=1}^n x_i y_i^r - n\bar{x}(r_0 + c_0) - (r_1 + c_1) \sum_{i=1}^n x_i^2 = 0 \end{aligned} \quad (25)$$

$$r_1 + c_1 = \frac{\sum_{i=1}^n x_i y_i^r - n\bar{x}\bar{y}^r}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \quad r_0 + c_0 = \bar{y}^r - (r_1 + c_1)\bar{x} \quad (26)$$

Στα παραπάνω οι ποσότητες $\bar{y}^\ell, \bar{x}, \bar{y}^r$ δίνονται ως ακολούθως:

$$\bar{y}^\ell = \frac{1}{n} \sum_{i=1}^n y_i^\ell, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y}^r = \frac{1}{n} \sum_{i=1}^n y_i^r. \quad y_i^\ell = y_i - e_i, y_i^r = y_i + e_i, \quad (27)$$

θεωρώντας ότι ο ασαφής αριθμός $\tilde{y}_i = (y_i, e_i)$ είναι συμμετρικός. Από τις παραπάνω εκφράσεις προκύπτουν οι ποσότητες r_0, c_0, r_1, c_1 και επομένως οι τιμές των παραμέτρων $\tilde{A}_0 = (r_0, c_0), \tilde{A}_1 = (r_1, c_1)$:

$$\begin{aligned} r_1 &= \frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i y_i^\ell - n\bar{x}\bar{y}^\ell}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} + \frac{\sum_{i=1}^n x_i y_i^r - n\bar{x}\bar{y}^r}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right\} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \\ c_1 &= \frac{1}{2} \left\{ \frac{\sum_{i=1}^n x_i y_i^r - n\bar{x}\bar{y}^r}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} - \frac{\sum_{i=1}^n x_i y_i^\ell - n\bar{x}\bar{y}^\ell}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right\} = \frac{\sum_{i=1}^n x_i e_i - n\bar{x}\bar{e}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \end{aligned}$$

$$r_0 = \frac{\bar{y}^l + \bar{y}^r - 2r_1\bar{x}}{2} = \bar{y} - r_1\bar{x},$$

$$c_0 = \frac{\bar{y}^r - \bar{y}^l - 2c_1\bar{x}}{2} = \bar{e} - c_1\bar{x}, \quad \bar{e} = \frac{1}{n} \sum_{i=1}^n e_i \quad (28)$$

Για την περίπτωση που έχουμε σαφή πειραματικά δεδομένα εξόδου ισχύει:

$$\tilde{y}_i \equiv y_i, \quad y_i^l = y_i^r = y_i, e_i = 0. \quad (29)$$

Από τα παραπάνω παίρνουμε:

$$r_1 = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, \quad c_1 = 0, \quad r_0 = \bar{y} - 2r_1\bar{x}, \quad c_0 = 0, \quad \tilde{A}_0 = (r_0, 0), \quad \tilde{A}_1 = (r_1, 0). \quad (30)$$

Σημείωση: Επομένως για την περίπτωση σαφών πειραματικών δεδομένων εξόδου η μέθοδος των ελαχίστων τετραγώνων δίνει σαφείς αριθμούς για τις παραμέτρους \tilde{A}_0, \tilde{A}_1 , ενώ η δυνατοτική μέθοδος δίνει ασαφείς αριθμούς.

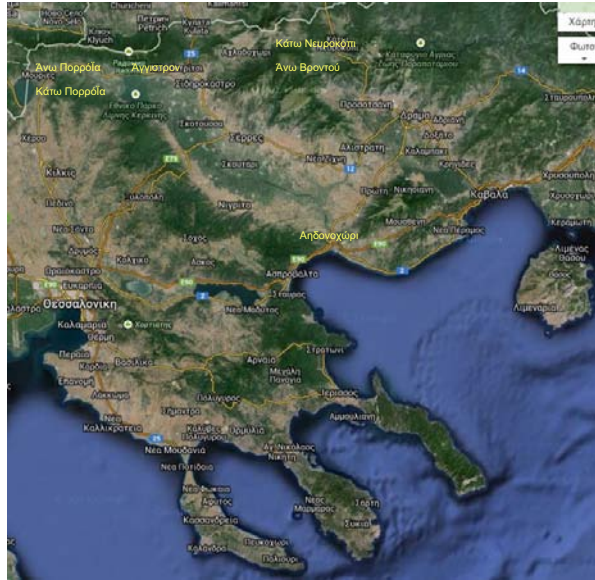
4. Εφαρμογή

Στα πλαίσια μιας έρευνας (Ορφανίδου, 2005), επιλέχτηκε μια περιοχή στη πεδιάδα των Σερρών με 6 Βροχομετρικούς Σταθμούς με κοινά Βροχομετρικά δεδομένα από 1929 έως 1941. Οι 6 Βροχομετρικοί Σταθμοί είναι οι ακόλουθοι: 1) Άνω Πορρόια, 2) Κάτω Πορρόια, 3) Άγγιστρον, 4) Κάτω Νευροκόπι, 5) Άνω Βροντού, 6) Αηδονοχώρι. (Σχήμα 4)

Τα γεωγραφικά πλάτη, γεωγραφικά μήκη και υψόμετρα των σταθμών αυτών δίνονται στον Πίνακα 1

Πίνακας 1

α/α	Σταθμοί	Γεωγραφικό Μήκος	Γεωγραφικό Πλάτος	Υψόμετρο	Υδρολογική Λεκάνη	Περίοδος	Φορέας
1	Αηδονοχώρι	23 ^ο 44'	40 ^ο 50'	212	Στρυμόνα	1929-1941	Δ.Ε.Β.Υ.Π.
2	Άνω Πορρόια	23 ^ο 02'	41 ^ο 18'	395	Στρυμόνα	1929-1941	Δ.Ε.Β.Υ.Π.
3	Κάτω Πορρόια	23 ^ο 33''	41 ^ο 17'	184	Στρυμόνα	1929-1941	Δ.Ε.Β.Υ.Π.
4	Άνω Βροντού	23 ^ο 41'	41 ^ο 18'	1040	Στρυμόνα	1929-1941	Δ.Ε.Β.Υ.Π.
5	Άγγιστρο	23 ^ο 26'	41 ^ο 22'	289	Στρυμόνα	1929-1045	Δ.Ε.Β.Υ.Π.
6	Κάτω Νευροκόπι	23 ^ο 45'	41 ^ο 18'	555	Στρυμόνα	1929-1045	Δ.Ε.Β.Υ.Π.



Σχήμα 4.

Οι συντελεστές συσχέτισης των παραπάνω χρονικών περιόδων δίδονται στον παρακάτω Πίνακα 2. Από τον Πίνακα αυτόν λαμβάνονται μόνο τα ζεύγη Σταθμών με συντελεστή Συσχέτισης μεγαλύτερο από 0.77. Οι αριθμοί αυτοί φαίνονται στον Πίνακα με έντονο μαύρο χρώμα.

Πίνακας 2

	Άγγιτρο	Άνω Βροντούς	Αηδονοχώρι	Άνω Πορρόια	Κ. Νευροκοπίου	Κάτω Πορρόια
Άγγιτρο	1.00	0.84	0.70	0.35	0.73	0.86
Άνω Βροντούς		1.00	0.78	0.44	0.84	0.77
Αηδονοχώρι			1.00	0.34	0.76	0.78
Άνω Πορρόια				1.00	0.18	0.48
Κ. Νευροκοπίου					1.00	0.67
Κάτω Πορρόια						1.00

Για λόγους οικονομίας στον Πίνακα 3 δίνονται τα μέσα μηνιαία ύψη βροχής(mm) μόνο για τους δύο σταθμούς Άγγιτρο (X) και Άνω Βροντούς (Y) για τις 12 χρονικές περιόδους, οι οποίοι και εξετάζονται στη συνέχεια.

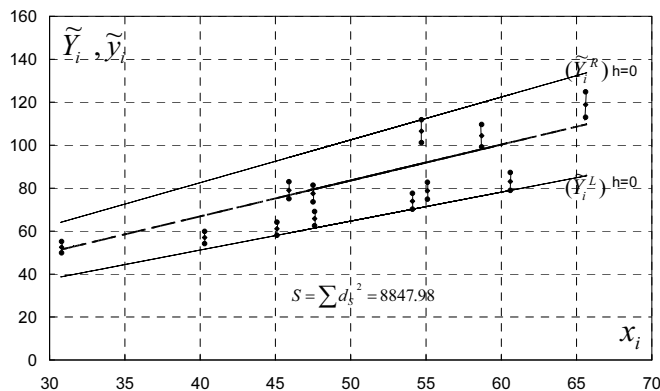
Πίνακας 3

T	1929-30	1930-31	1931-32	1932-33	1933-34	1934-35	1935-36	1936-37	1937-38	1938-39	1939-40	1940-41
Άγγιτρο-X	47,6	65,6	40,3	30,8	45,1	54,1	58,7	60,6	55,1	45,9	54,7	47,5
Άνω Βροντού-Y	65,8	118,9	57	52,5	61,1	73,9	104,4	83,1	78,8	79	106,5	77,5
e	3,29	5,945	2,85	2,625	3,055	3,695	5,22	4,155	3,94	3,95	5,325	3,875

4.1. Δυνατοτικό μοντέλο

Στην εξίσωση (1) έχουμε $\tilde{Y}_j = \tilde{A}_0 + \tilde{A}_1 x_{1j}$ και τα δεδομένα του σταθμού Άνω Βρο-

ντού θεωρούνται ασαφή συμμετρικά $\tilde{Y} = (y_i, e_i)$. Στο παρακάτω Σχήμα 5 εφαρμόστηκε η θεωρία του Tanaka με $h=0$. Οι τιμές των r_0, r_1, c_0, c_1 προέκυψαν ίσες προς: $r_0=0, r_1=1.67107, c_0=2.767268, c_1=0.32268$.



Σχήμα 5. Μέθοδος του Tanaka.

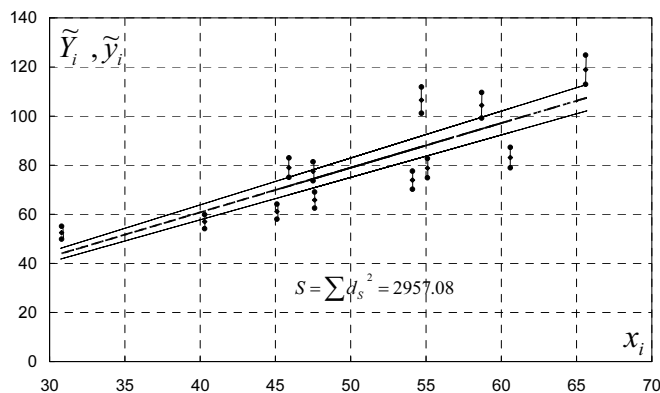
Τελικά η λύση είναι:

$$\tilde{Y}_i = (0, 2.767265) + (1.67107, 0.32268)x_i$$

4.2. Μοντέλο ελαχίστων τετραγώνων

Εφαρμόστηκε η θεωρία των ελαχίστων τετραγώνων του Diamond. Οι τιμές των r_0, r_1 , προέκυψαν ίσες προς $r_0 = -11.9413, r_1 = 1.818145$, δηλαδή εν προκειμένω ισχύει: $r_0 < 0, r_1 > 0$, και επομένως η λύση προκύπτει ίση με $c_0 = -0.59707, c_1 = 0.090907$.

Τελικά η λύση είναι:



Σχήμα 6. Μέθοδος των ελαχίστων τετραγώνων.

$$\tilde{Y} = (-11.9413, -0.57707) + (1.818145, 0.090907)x_i$$

Στο Σχήμα 6 φαίνεται η λύση των ελαχίστων τετραγώνων και οι μετρηθείσες τιμές. Επίσης στο σχήμα αυτό φαίνεται και η ποσότητα $S_{LS} = \sum d_s^2 = 2957.08$. Στο Σχήμα 5 η ίδια ποσότητα είναι ίση με $S_T = \sum d_s^2 = 8847.98$.

Ως συμπέρασμα προκύπτει το ακόλουθο:

- Η μέθοδος των ελαχίστων τετραγώνων πλεονεκτεί ως προς την δυνατοτική μέθοδο του Ταπακα σε ότι αφορά τις αποστάσεις μεταξύ μετρημένων και υπολογισμένων τιμών ($S_{LS} < S_T$).
- Η μέθοδος του Ταπακα καλύπτει όλες τις ακραίες μετρημένες ασαφείς τιμές, δηλαδή ισχύει ($[\tilde{y}_i] \subseteq [\tilde{Y}_i], \forall i$), όπου \tilde{y}_i =μετρηθείσα τιμή, και \tilde{Y}_i =εκτιμηθείσα τιμή, ενώ δεν ισχύει αυτό στη μέθοδο των ελαχίστων τετραγώνων, όπου πολλές μετρηθείσες τιμές βρίσκονται εκτός των οριακών εκτιμηθεισών τιμών.

5. Βιβλιογραφία

Ελληνική

- Γουάλλες Ε., 2005**, «Ευφυή συστήματα βασισμένα στη γνώση σε αβέβαια περιβάλλοντα», Διδακτορική Διατριβή, Ε.Μ.Π., σελ. 149
- Ορφανίδου Ε., 2005**, «Το Υδατικό ισοζύγιο του Στρυμόνα στον Ελληνικό χώρο», Διπλωματική Εργασία, Εργαστήριο Υδραυλικών Έργων και Διαχείρισης Περιβάλλοντος, Α.Π.Θ., Θεσσαλονίκη.
- Σκριμιζέας Π., 2014**, «Εφαρμογές χωρικής ανάλυσης κλιματικών δεδομένων- Χαρτογράφηση βροχομετρικών δεδομένων στον Ελληνικό χώρο», Μεταπτυχιακή Εργασία, Χαρτοκόπιο Πανεπιστήμιο, Τμήμα Γεωγραφίας, σελ. 224.
- Τζιμόπουλος Χ. και Παπαδόπουλος Β. , 2013**, «Ασαφής Λογική με εφαρμογές στις επιστήμες του μηχανικού», Εκδόσεις Ζήτη, σελ. 627.
- ΥΔΡΟΓΝΩΜΩΝ, 2010**, «Θεωρητική τεκμηρίωση για το λογισμικό: ΥΔΡΟΓΝΩΜΩΝ- Έκδοση4», Ε.Μ.Π., σελ. 173.

Ξενόγλωσση

- Bargiela A., Pedrycz W., Nakashima T., 2007**, “Multiple regression with fuzzy data”, Fuzzy Sets and Systems, 158, 2169-2188.
- Bisserier A., 2010**, “Une approche paramétrique de la régression linéaire floue- Formalisation par intervalles”, Thèse de Docteur, Université de Savoie, France, pp. 172.
- Chang P-T and Lee E.S., 1996**, “A generalized fuzzy least-squares regression”, Fuzzy Sets and Systems, 82, 289-298.

- Charfeddine S., Mora-Camino F., De Coligny M., 2004**, “Fuzzy linear regression : application to the estimation of air transport demand”, International Conference on Fuzzy Sets and Soft Computing in Economics and Finance, Saint-Petersburg, Russia.
- Diamond P.**, 1988, “Fuzzy least squares”, 46, pp. 141-157
- Dubuisson M. P. and Jain A. K., 1994**, “A Modified Hausdorff Distance for Object Matching”, Proc. International Conference on Pattern Recognition, Israel, pp. 566-568
- Giles, J.R., 1987**, “Introduction to the Analysis of Metric Spaces”, Australian Mathematical Lecture Series 3, Cambridge University Press, Cambridge.
- Kim, H.K., Yoon, J.H., Li, Y., 2008**, “Asymptotic properties of least squares estimation with fuzzy observations”, Information Science 178, 439-451
- Ming M., Friedman M., Kandel A., 1997**, “General fuzzy least squares ”, Fuzzy Sets and Systems, 88, 107-118
- Moskowitz H. and Kim K., 1993**, “On assessing the H value in fuzzy linear regression”, Fuzzy Sets and Systems, 58, 303-327.
- Nong X., 2012**, “A new fuzzy linear regression model for least squares estimate”, X. Qu and Y, Yang(Eds): IBI2011. Part II, CCIS 268, 709-715.
- Papadopoulos B. and M. Sirpi, 1999**, “Similarities in Fuzzy Regression Models”, Journal of Optimization Theory and Applications, Vol. 102, No. 2, pp. 373-383.
- Papadopoulos B. and M. Sirpi, 2004**, “Similarities and Distances in Fuzzy Regression Models”, Soft Computing, Vol. 8, No. 8, pp. 556-561.
- Popescu C., Giuclea M., 2007**, “A model of multiple linear regression”, Proceedings of the Romanian Academy, Series A, Vol. 8(2), 1-8.
- Redden D. T. and Woodall W. H., 1996**, “Further examination of fuzzy linear regression”, Fuzzy Sets and Systems, 79, 203-211
- Savic D.A., Pedrycz W., 1991**, “Evaluation of fuzzy linear regression models”, Fuzzy Sets and Systems, 39, 51-63.
- Tanaka, H., S. Uejima, K. Asai, 1982**, “Linear regression analysis with fuzzy models”, IEEE TRANS. SYS. MAN AND CYBER., 12(6):903-907.
- Tanaka, H., 1987**, “Fuzzy data analysis by possibilistic linear models”, Fuzzy Sets and Systems, 24, 363-375.
- Tanaka, H. and Watada J., 1988**, “Possibilistic linear systems and their application to the linear regression model”, Fuzzy Sets and Systems, 27, 275-289.
- Tanaka, H. and Hayashi I., 1989**, “Possibilistic linear regression analysis for fuzzy data”, European Journal of Operational Research, 40, 389-396.
- Tanaka, H. and Ishibuchi H., 1991**, “Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters”, Fuzzy Sets and Systems, 41, 145-160.
- Torfi F., Farahani R.Z., Mahdavi I., 2011**, “Fuzzy least-squares linear regression approach to ascertain stochastic demand in the vehicle routing problem”, Applied Mathematics, 2, 64-73.

- Tran L., Ducstein L., 2002**, “Comparison of fuzzy numbers using a fuzzy distance measure”, *Fuzzy Sets and Systems*, 130, 331-341.
- Yang, M.S., Ko, C.H.1 1997**, “On cluster-wise fuzzy regression analysis”, *IEEE Trans. Systems Man. Cybernet.-Part B: Cybernet.* 27(1), 1-13.
- Yang M.-S., Liu H.-H., 2003**, “Fuzzy least squares algorithms for interactive fuzzy linear regression models”, *Fuzzy Sets and Systems*, 135, 305-316.
- Zadeh L.A., 1965**, “Fuzzy Sets”, *Inform. and Control* 8, 338-353.
- Zwick R., Carlstein E., Budescu D.V., 1987**, “Measures of similarity among fuzzy concepts: A comparative analysis”, *Int. J. of Approx. Reasoning*, 1, pp.221-242.