Noname manuscript No.
(will be inserted by the editor)

Editorial
Fuzzy Data Analysis and Classification

Special Issue *in memoriam* of Professor Lotfi A. Zadeh, father of Fuzzy Logic

Pierpaolo D'Urso · María Ángeles Gil

Received: date / Accepted: date

In analyzing and classifying data from a statistical perspective, fuzzy sets and logic have become a valuable tool either to model and handle imprecise data or to establish flexible techniques to deal with precise data.

From the very beginning of his 52 years-old theory, Professor Zadeh highlighted that "Probability Theory/Statistics and Fuzzy Logic should be viewed as complementary rather than competitive," and he anticipated and encouraged the materialization of such a complementarity. Nowadays, this assertion is a reality, as shown by the many related papers, specialized conferences, special sessions and tracks in general conferences, and so on.

This special issue started in 2015, with the 50th anniversary of the seminal paper on fuzzy sets by Zadeh [156], aiming to collect a sample of research papers about the current trends on the combination of Fuzzy Sets/Logic and Data Analysis/Classification.

When this special issue was almost ready for publication, Zadeh unfortunately passed away at age 96 (February 4, 1921 - September 6, 2017). ***We wish this special issue to be dedicated to Professor Zadeh, as a modest part of the many tributes that he will receive, and intending to show that Fuzzy Sets/Logic and Data Analysis/Classification can certainly work in synergy.***

Pierpaolo D'Urso
Department of Social Sciences and Economics, Sapienza Università di Roma, P.za Aldo Moro, 5, Rome, Italy
E-mail: pierpaolo.durso@uniroma1.it

María Ángeles Gil
Department of Statistics, OR and TM, Universidad de Oviedo, C/ Federico García Lorca, 18, Oviedo, Spain
E-mail: magil@uniovi.es

1 Introduction

This issue means a contribution of a well-known rather young journal from a scientific field that, sometimes and especially at the beginning of the introduction of fuzzy theory, put it into question. However, Zadeh has been permanently trying to build bridges between Statistics/Probability and Fuzzy Logic. We have thought that this was a cordial and gentle way to combate criticisms, in accordance with his well-known friendly attitude, by demonstrating the potential benefits of collaboration between both fields. But we have been 'probably' wrong. Zadeh's fondness and affection towards Probability and Statistics is neither incidental nor temporary, but distant in time.

To illustrate the last assertion, in his 2015's paper [161] Professor Zadeh has taken a look back over his past in research and he says: "... The early years in my academic career coincided with the birth of the age of computers and information. It was an exciting period, spurred by competition between the United States and the Soviet Union. At Columbia, my research was focused on system theory and information systems. Probability theory had a position of centrality in my work. My first paper, published in 1949 in the Journal of Applied Physics, was entitled, *Probability criterion for the design of servomechanisms.* My second paper, published also in the Journal of Applied Physics in 1950, was entitled, *An extension of Wiener's theory of prediction.* I had a close relationship with the Department of Mathematical Statistics and its Chair, Herbert Robbins, a brilliant mathematician, who became my best friend..." Actually, many of Zadeh's first publications' titles involve unequivocally statistical terms such as *stochastic operators*, *correlation functions*, *prediction*, *time-series*, and so on. The same applies to several communications presented in meetings of the American Mathematical Society.

Anyway, as it often happens with new theories and approaches, there has been more than two decades in which probabilistic and statistical journals scarcely publish papers concerning fuzzy material, but maybe some eventual debates. In this respect, as outlined in Belohlavek *et al.* [7] and Ross *et al.* [129], one should at least mention the following:

- First, the debate along the Conference on the Calculus of Uncertainty in Artificial Intelligence and Expert Systems held in 1984. In this conference, Lotfi Zadeh, as supporter of fuzzy logic, Dennis Lindley, as supporter of subjective probability, Glenn Shafer, as supporter of evidence theory, and David Spiegelhalter, as supporter of applications of expert systems, acted as invited speakers. This debate was mostly gathered in the Issue 1 of Volume 2 of the journal *Statistical Science* [140], although Zadeh could not send his written presentation in due time but he often referred to what was discussed there in different papers.
- Secondly, the invited paper with discussions by Laviolette *et al.* [102] in the journal *Technometrics*, where Laviolette *et al.* reviewed some basic ideas in fuzzy theory and offer what they consider to be simpler alternatives based on traditional probability and statistical theory; the corresponding Zadeh's discussion [159] argued that probability theory by itself is not sufficient

for dealing with uncertainty and imprecision in real-world settings, but allowing them to coexist is much more effective and reasonable.

- Finally, another interesting debate is the one around the invited paper with discussions by Singpurwalla and Booker [138] in the *Journal of the American Statistical Association*, in which authors articulated that probability theory has a sufficiently rich structure for incorporating fuzzy sets within its framework, and that probability and fuzzy set theories can work in concert. Zadeh's discussion about this last paper [160] suggested to restructure probability theory by involving a shift in its foundations from bivalent to fuzzy logic.

As we have already commented, nowadays the controversy has been substantially diminishing and ideas of reconciliation and coexistence of fuzzy and probability/statistics theories and the development of hybrid models and methods have been thriving. Kruse *et al.* have pointed out [99] that "... We must differentiate between *fuzzy* data analysis and *fuzzy data* analysis. The former deals with the analysis of classical data using methods based on fuzzy set theory. These methods, e.g., fuzzy clustering or fuzzy regression analysis, have been used successfully in lots of industrial applications. The second approach tries to analyze fuzzy data by using statistical methods..." In other words, Fuzzy Data Analysis and Classification studies are mainly focussed

– either on developing concepts, results and methods to deal with classical (non-fuzzy) data, where fuzziness is involved in the construction of the analysis/classification procedures,
– or on developing/extending concepts, results and methods concerning data analysis and classification of fuzzy-valued data,
– or on both.

2 On the fuzzy analysis and the fuzzy classification of non-fuzzy/standard data

The development of fuzzy approaches to classify 'crisp' data started soon after the formalization of fuzzy sets by Zadeh [156]. In fact, Zadeh along with Bellman and Kalaba were the first in suggesting fuzzy sets as a theoretical basis to develop clustering algorithms [6]. Some of the most influential pioneer works on the subject are, among others, those by Ruspini [132,133], Tamura *et al.* [141], Dunn [59,60], Bezdek [9–11], and Bezdek *et al.* [13], which have inspired both applications and many further methodologies. At present, this is one of the most successful topics involving Fuzzy Sets and Statistical theories, and the number of research papers on it is unquestionably growing (among the most recent ones see, for instance, the approaches in Liu *et al.* [107], Gong *et al.* [78], Yamashita and Mayekawa [152], Ruan *et al.* [130], and Nguyen-Trangand and Vo-Van [118]), and it appears often either combined with or supporting other data analysis problems.

In more detail, useful references to the extensive literature on the fuzzy clustering (from both theoretical and applicative points of view) can be found in the chapter on the fuzzy clustering by D'Urso [32], the seminal monograph by Bezdek [12], the books by Jain and Dubes [93], De Oliveira and Pedrycz [52], Miyamoto *et al.* [113] and, e.g., the following journals: Fuzzy Sets and Systems, IEEE Transactions on Fuzzy Systems, Information Sciences, Pattern Recognition, Applied Soft Computing, Soft Computing, Advances in Data Analysis and Classification, Computational Statistics and Data Analysis, Chemometrics and Intelligent Laboratory Systems, Pattern Recognition Letters, etc.

As remarked by D'Urso [33], there are different uncertainty-based clustering methods that can be considered extensions, variants and alternatives of the fuzzy clustering for non-fuzzy/standard data, like

- possibilistic clustering (see, for instance, Krishnapuram and Keller [96]),
- shadowed clustering (see, for instance, Pedrycz [121]),
- rough sets-based clustering (see, for instance, Lingras and West [105]),
- intuitionistic fuzzy clustering (see, for instance, Hung *et al.* [90]),
- evidential clustering, credal clustering or belief clustering (see, for instance, Denoeux and Masson [55]),
- credibilistic clustering (see, for instance, Zhou *et al.* [162]),
- type-2 fuzzy clustering (see, for instance, Hwang and Rhee [91]),
- neutrosophic clustering (see, for instance, Shan *et al.* [134]),
- hesitant fuzzy clustering (see, for instance, Chen *et al.* [20]),
- interval-based fuzzy clustering (see, for instance, Silva *et al.* [135]),
- picture fuzzy clustering (see, for instance, Son [139]).

Fuzzy approaches to analyze crisp/standard data, have been not been carried out as exhaustively as fuzzy clustering ones for the same data. And they were developed several years after fuzzy sets were introduced. Among them, one can highlight

- the fuzzy linear regression ideas between non-fuzzy input and output data, by considering the problem as a linear programming one (see, for instance, the first formulation by Tanaka *et al.* [143] and Tanaka and Watada [144]),
- hypothesis fuzzy testing, testing of fuzzy hypotheses, and fuzzy estimation regarding non-fuzzy parameters on the basis of non-fuzzy data (see, for instance, Watanabe and Imaizumi [149], Arnold [2], Buckley [16], Hryniewicz [88], Parchami *et al.* [120]),
- fuzzy statistical quality control (see, for instance, Grzegorzewski and Hryniewicz [82]),
- statistical decision problems with fuzzy utilites/losses (see, for instance, Gil and Jain [71], Gil and López-Díaz [72]).

3 On the analysis and classification of fuzzy data

On the other hand, approaches to classify fuzzy-valued data are becoming a challenging topic. Among the first published approaches one should mention those by Esogbue [61], Hathaway *et al.* [85], and Pedrycz *et al.* [122] and, among the recent ones, see those by Coppi *et al.* [26], D'Urso and De Giovanni [35], D'Urso *et al.* [37], Ansari *et al.* [1] and Ferraro and Giordani [64].

The analysis of fuzzy-valued data is also a topic receiving an increasing attention along the years. Some of the developed methodologies aiming to analyze fuzzy data consider a descriptive view and do not refer to models associated with the probabilistic framework. Nevertheless, most of the methodologies are based on the modeling of the random mechanisms generating fuzzy data within a probabilistic setting. In this respect, we can mention, among some of these methodologies the following:

- The methodologies based on the notion of *fuzzy information system*, introduced by Okuda *et al.* [119], who consider the available information from a classical random experiment associated with a real-valued random variable to be fuzzy (i.e., they consider an epistemic viewpoint in accordance with the distinction made by Couso and Dubois [28]) and assume that this available information constitutes a fuzzy partition (in Ruspini's sense [133]) of the sample space of the variable, and probabilities are based on Zadeh's probabilistic definition of fuzzy events [157]. Some data analysis developments using this model can be seen, for instance, in Gil *et al.* [70], Gil [69] and, more recently, Denoeux [53].
- The methodologies based on the notion of *fuzzy random variable*, introduced by Kwakernaak [101] and later formalized by Kruse and Meyer [100]. As Kruse *et al.* pointed out in [99], this deep and wide data analysis with vague data was a fruit of the encouragement by Professors Lotfi Zadeh and Heinz Skala (editor of the Series Theory and Decision Library of the D. Reidel Publishing Co., see, e.g., [137]), which was mostly prompted by the development by Kruse and Meyer of useful fuzzy methods and a software tool for statistical applications for the Siemens AG. The model refers to the epistemic perspective and fuzzy random variable is viewed as the fuzzy perception of an original non-fuzzy random variable. Statistical developments with fuzzy data coming from the fuzzy perception of real-valued ones will be mainly based on propagating the associated imprecision to the distribution function, parameters, etc., through Zadeh's extension principle [158]. It should be remarked that, albeit based on fuzzy information, statistical conclusions with Kruse and Meyer's fuzzy random variables always concern the original random variable and its parameters. Among the studies based on Kruse and Meyer's fuzzy random variables one can refer, for instance, to Kruse [97,98], Grzegorzewski [81], Wang [148], and Wu [150].
- The methodologies based on the notion of *random fuzzy sets*, introduced by Féron [62], and to some extent anticipated by Fréchet [66], and later formalized by Puri and Ralescu [124,125]. The model, which was initially coined as fuzzy random variables, refers to a kind of ontic perspective, since a random fuzzy set (or random fuzzy number if values are fuzzy numbers) is viewed as a mapping associating experimental outcomes with fuzzy values in a Borel-measurable way, so that the induced distribution associated with the random fuzzy set is immediate, the stochastic independence between random fuzzy sets is also trivially induced, and so on. It should be remarked that, in contrast to the Kruse and Meyer's approach, statistical conclusions with Puri and Ralescu random fuzzy sets always concern the fuzzy-valued

random element and the parameters associated with its induced distribution. An interesting distinctive feature of the statistical methodology based on this approach to generate fuzzy data is that most of the classical ideas in data analysis can be immediately preserved without needing to either define or adapt them expressly. Among the statistical developments involving this approach one can refer, for instance, to Bandemer and Näther [5] and Näther [115,116] and, more recently, Blanco-Fernández *et al.* [14,15].

Other developments and approaches can be found in the literature (for instance, those by Viertl [147], Grzegorzewski and Szymanowiski [83], etc.). Among the scientific journals publishing papers on the topic, one can mention those indicated in Section 2 for the fuzzy clustering and analysis of standard data.

4 Additional related literature

It is pertinent to state at this point that some other procedures have been suggested in the literature to categorize non-standard data in a fuzzy manner, and some of them have been gathered in Table 1 (see also, D'Urso [32]).

Table 1 Some relevant references on fuzzy clustering of non-standard data

Typology of data	See, for instance,...
fuzzy data, symbolic data, interval-valued data	Section 3 and Table 2 of this editorial
categorical data	Huang and Ng [89], Lee and Pedrycz [103]
functional data	Tokushige *et al.* [146], Tan *et al.* [142]
textual data (text data)	Runkler and Bezdek [131]
time data	Coppi and D'Urso [22–24], D'Urso [30], Maharaj and D'Urso [112], D'Urso *et al.* [46,39,45]
spatial data	Pham [123]
spatial-time data	Coppi *et al.* [25], Disegna *et al.* [58]
three-way data	Giordani [76], Rocci and Vichi [128]
sequence data	D'Urso and Massari [47]
network data	Liu [106]
directional data	Yang and Pan [155], Kesemen *et al.* [94]
distributional data	Irpino *et al.* [92]
mixed data	Yang *et al.* [153]
outlier data	Davé [50], Krishnapuram and Keller [96], Frigui and Krishnapuram [67], Wu and Yang [151], D'Urso and Giordani [42], Fritz *et al.* [68], Ferraro and Vichi [65], Ferraro and Giordani [64], D'Urso *et al.* [37–40], D'Urso and Leski [44], Yang and Nataliani [154]
incomplete data	Hathaway and Bezdek [84]
data streams	Berlinger and Hüllermeier [8]
big data	Havens *et al.* [86]

In a similar way, Table 2 collects some of the most relevant references on the methodological statistical studies on the analysis of fuzzy data (see also D'Urso [34]).

Table 2 Some relevant references on the analysis of fuzzy data

Typology of the analysis	See, for instance,...
one-sample, two-sample, ANOVA tests about means and variances of random fuzzy sets	Körner [95], Montenegro *et al.* [114], González-Rodríguez *et al.* [80], Ramos-Guajardo *et al.* [126], Ramos-Guajardo and Lubiano [127], Lubiano *et al.* [110]
fuzzy estimates of location of random fuzzy numbers; robustness	Lubiano and Gil [109], Sinova *et al.* [136]
statistical comparison of fuzzy scale with other imprecise-valued scales	De la Rosa de Sáa *et al.* [51], Gil *et al.* [74], Lubiano *et al.* [108,111]
fuzzy inequality	Gil *et al.* [73]
discriminant analysis	Colubi *et al.* [21]
cluster analysis	Hathaway *et al.* [85], Pedrycz *et al.* [122], Auephanwiriyakul and Keller [4], D'Urso [31], Coppi *et al.* [26]
regression analysis	Celminš [19], Diamond [57], Näther and Albrecht [117], Coppi *et al.* [27], D'Urso and Gastaldi [43], D'Urso [29], González-Rodríguez *et al.* [79], D'Urso *et al.* [48]
principal component analysis	Denoeux and Masson [56], D'Urso and Giordani [41], Giordani and Kiers [77], Calcagnì *et al.* [17]
multidimensional scaling	Denoeux and Masson [54], Hébert *et al.* [87]
self-organizing maps	D'Urso *et al* [36]
clusterwise regression analysis	D'Urso and Santoro [49]
correspondence analysis	Theodorou *et al.* [145], Aşan and Greenacre [3]
regression trees	Cappelli *et al.* [18], Lertworaprachaya *et al.* [104]
three-way analysis	Coppi and D'Urso [23], Giordani [76]

5 On this special issue

The papers in this special issue are to be considered as a sample of recent advances in data analysis and classification involving fuzziness, illustrating the need of taking advantage of other topics like Fuzzy Logic to enrich and to widen statistical methodologies. Although small samples are not usually informative enough from a statistical perspective, and this special issue is certainly a very small sample, we trust that readers can get a flavour of some of the current trends about.

The first four papers in the issue concern *fuzzy* data analysis or classification:

- The paper "A fuzzy approach to robust regression clustering", by Dotto, Farcomeni, García-Escudero and Mayo-Iscar, proposes a fuzzy regression clustering method based on a maximum likelihood approach but in such a way that the method resists well to data contamination.
- The paper "A novel method for forecasting time series based on fuzzy logic and visibility graph", by Zhang, Ashuri and Deng, relates to a new suggestion to forecast time series, which is based on fuzzy logic, visibility graph and link prediction.
- The paper "Fuzzy rule based classification systems for big data with MapReduce: granularity analysis", by Fernández, del Río, Bawakid and Herrera, aims to discuss the effect of the granularity level and the number of selected Maps on the performance of the Chi-Fuzzy Rule Based Classification Systems with a MapReduce approach for big data.
- The paper "On ill-conceived initialization in archetypal analysis", by Suleman, addresses the problem of initialization and the performance of fuzzy clustering by means of an archetypal analysis.

The last two papers in the issue concern *fuzzy data* analysis or classification:

- The paper "Robust scale estimators for fuzzy data", by de la Rosa de Sáa, Lubiano, Sinova and Filzmoser, regards the introduction of some robust location-based scale measures/estimates for random fuzzy numbers, along with the analysis of their robustness.
- The paper "Parametric classification with soft labels using the evidential EM algorithm. Linear discriminant analysis *vs* logistic regression", by Denoeux, Quost and Li, analyzes the problem of partially supervised classification when learning instances are labeled by means of Dempster-Shafer mass functions (which include fuzzy sets as a particular case).

Acknowledgements We are deeply indebted to the ADAC Co-Editor Professor Maurizio Vichi because of having kindly and patiently supported this initiative, which is dedicated to Professor Lotfi A. Zadeh, the father of Fuzzy Logic.

References

1. Ansari ZA, Sattar SA, Babu AV (2017) A fuzzy neural network based framework to discover user access patterns from web log data. Adv Data Anal Classif 11(3):519–546
2. Arnold BF (1998) Testing fuzzy hypotheses with crisp data. Fuzzy Sets Syst 94(3):323–333
3. Aşan Z, Greenacre M (2011) Biplots of fuzzy coded data. Fuzzy Sets Syst 183:57–71
4. Auephanwiriyakul S, Keller JM (2002) Analysis and efficient implementation of a linguistic fuzzy c-means. IEEE Trans Fuzzy Syst 10:563–582
5. Bandemer H, Näther (1992) Fuzzy Data Analysis. Springer, Dordrecht
6. Bellman RE, Kalaba R, Zadeh LA (1966) Abstraction and pattern classification. J Math Anal Appl 13:1–7

7. Belohlavek R, Dauben JW, Klir GJ (2017) Fuzzy Logic and Mathematics. A Historical Perspective. Oxford University Press, New York
8. Berlinger J, Hüllermeier E (2007) Fuzzy clustering of parallel data streams. In: [52], pp 333–352
9. Bezdek JC (1973) Cluster validity with fuzzy sets. Cybern & Syst/J Cybern 3(3):58–73
10. Bezdek JC (1974) Numerical taxonomy with fuzzy sets. J Math Biol 1(1):57–71
11. Bezdek JC (1980) Convergence theorem for the fuzzy ISODATA clustering algorithms. IEEE Trans Patt Anal Mach Intel 2(1):1–8
12. Bezdek JC (1981) Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York
13. Bezdek JC, Ehrlich R, Full W (1984) FCM - The fuzzy c-means clustering-algorithm. Comp & Geosci 10(2–3):191–203
14. Blanco-Fernández A, Casals MR, Colubi A, Corral N, García-Bárzana M, Gil MA, González-Rodríguez G, López MT, Lubiano MA, Montenegro M, Ramos-Guajardo AB, de la Rosa de Sáa S, Sinova B (2014) A distance-based statistical analysis of fuzzy number-valued data. Int J Approx Reas 55(7):1487–1501
15. Blanco-Fernández A, Casals MR, Colubi A, Corral N, García-Bárzana M, Gil MA, González-Rodríguez G, López MT, Lubiano MA, Montenegro M, Ramos-Guajardo AB, de la Rosa de Sáa S, Sinova B (2014) Rejoinder on “A distance-based statistical analysis of fuzzy number-valued data”. Int J Approx Reas 55(7):1601–1605
16. Buckley JJ (2004) Fuzzy Statistics. Studies in Fuzziness and Soft Computing Series 149. Springer, Berlin
17. Calcagnì A, Lombardi L, Pascali E (2016) A dimension reduction technique for two-mode non-convex fuzzy data. Soft Comput 20:749–762
18. Cappelli C, D'Urso P, Di Iorio F (2013) Change point analysis for imprecise time series. Fuzzy Sets Syst 225:23–38
19. Celminš A (1987) Multidimensional least-squares fitting of fuzzy models. Math Model 9:669–690
20. Chen N, Xu Z, Xia M (2013) Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. Appl Math Mod 37:2197–2211
21. Colubi A, González-Rodríguez G, Gil MA, Trutschnig W (2011) Nonparametric criteria for supervised classification of fuzzy data. Int J Approx Reas 52:1272–1282
22. Coppi R, D'Urso P (2002) Fuzzy k-means clustering models for triangular fuzzy time trajectories. Statist Meth Appl 11(1):21–40
23. Coppi R, D'Urso P (2003) Three-way fuzzy clustering models for LR fuzzy time trajectories. Comput Stat Data Anal 43:149–177
24. Coppi R, D'Urso P (2006) Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization. Comput Stat Data Anal 50:1452–1477
25. Coppi R, D'Urso P, Giordani P (2010) A fuzzy clustering model for multivariate spatial time series. J Classif 27:54–88
26. Coppi R, D'Urso P, Giordani P (2012) Fuzzy and possibilistic clustering for fuzzy data. Comput Stat Data Anal 56(4):915–927
27. Coppi R, D'Urso P, Giordani P, Santoro A (2006) Least squares estimation of a linear regression model with LR fuzzy response. Comput Stat Data Anal 51:267–286
28. Couso I, Dubois D (2014) Statistical reasoning with set-valued information: Ontic *vs* epistemic views. Int J Approx Reas 55(7):1502–1518
29. D'Urso P (2003) Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput Stat Data Anal 42(1-2):47–72
30. D'Urso P (2005) Fuzzy clustering for data time array with inlier and outlier time trajectories. IEEE Trans Fuzzy Syst 13:583–604
31. D'Urso P (2007) Fuzzy clustering of fuzzy data. In: [52], pp 155–192
32. D'Urso P (2016) Fuzzy clustering. In: Hennig C, Meila M, Murtagh F, Rocci R (eds) Handbook of Cluster Analysis, Chapman & Hall, Boca Raton, pp 545–573
33. D'Urso P (2017) Informational Paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review. Inform Sci 400-401:30–62
34. D'Urso P (2017) Exploratory multivariate analysis for empirical information affected by uncertainty and modeled in a fuzzy manner: A review. Gran Comput 2:225–247
35. D'Urso P, De Giovanni L (2014) Robust clustering of imprecise data. Chem Intel Lab Syst 136:58–80

36. D'Urso P, De Giovanni L, Massari R (2014) Self-organizing maps for imprecise data. Fuzzy Sets Syst 237:63–89
37. D'Urso P, De Giovanni L, Massari R (2015) Trimmed fuzzy clustering for interval-valued data. Adv Data Anal Classif 8(1):21–40
38. D'Urso P, De Giovanni L, Massari R (2015) Time series clustering by a robust autoregressive metric with application to air pollution. Chemom Intel Lab Syst 141(15):107–124
39. D'Urso P, De Giovanni L, Massari R (2016) GARCH-based robust fuzzy clustering of time series. Fuzzy Sets Syst 305:1–28
40. D'Urso P, De Giovanni L, Massari R, Cappelli C (2017) Exponential distance-based fuzzy clustering for interval-valued data. Fuzzy Optim & Decis Mak 16:51–70
41. D'Urso P, Giordani P (2005) A possibilistic approach to latent component analysis for symmetric fuzzy data. Fuzzy Sets Syst 150:285–305
42. D'Urso P, Giordani P (2006) A robust fuzzy k-Means clustering model for interval valued data. Comput Stat 21:251–269
43. D'Urso P, Gastaldi T (2000) A least-squares approach to fuzzy linear regression analysis. Comput Stat Data Anal 34:427–440
44. D'Urso P, Leski J (2016) Fuzzy C-ordered medoids clustering of interval-valued data. Patt Recogn 58:49–67
45. D'Urso P, Maharaj EA, Alonso AM (2017) Fuzzy clustering of time series using extremes. Fuzzy Sets Syst 318:56–79
46. D'Urso P, Maharaj EA, Galagedera DUA (2010) Wavelets-based fuzzy clustering of time series. J Classif 27:231–275
47. D'Urso P, Massari R (2013) Fuzzy clustering of human activity patterns. Fuzzy Sets Syst 215:29–54
48. D'Urso P, Massari R Santoro A (2011) Robust fuzzy regression analysis. Inform Sci 181:4154–4174
49. D'Urso P, Santoro A (2006) Fuzzy clusterwise regression analysis with symmetrical fuzzy output variable. Comput Stat Data Anal 51:287–313
50. Davé RN (1991) Characterization and detection of noise in clustering. Patt Recogn Lett 12:657–664
51. De la Rosa de Sáa S, Gil MA, González-Rodríguez G, López MT, Lubiano MA (2016) Fuzzy rating scale-based questionnaires and their statistical analysis. IEEE Trans Fuzzy Syst 23(1):111–126
52. De Oliveira JV, Pedrycz W (2007) Advances in Fuzzy Clustering and Its Applications. John Wiley & Sons, Chichester
53. Denoeux T (2011) Maximum likelihood estimation from fuzzy data using the EM algorithm. Fuzzy Sets Syst 183:72–91
54. Denoeux T, Masson MH (2000) Multidimensional scaling of interval-valued dissimilarity data. Patt Recogn Lett 21:83–92
55. Denoeux T, Masson MH (2004) EVCLUS: Evidential clustering of Proximity Data. IEEE Trans Syst Man Cybern Part B-Cybern 34:95–109
56. Denoeux T, Masson MH (2004) Principal component analysis of fuzzy data using autoassociative neural networks. IEEE Trans Fuzzy Syst 12:336–349
57. Diamond P (1988) Fuzzy least squares. Inform Sci 46:141–157
58. Disegna M, D'Urso P, Durante F (2017) Copula-based fuzzy clustering of spatial time series. Spat Stat 21:209–225
59. Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Cybern & Syst/J Cybern 3(3):32–57
60. Dunn JC (1974) Well-separated clusters and optimal fuzzy partitions. Cybern & Syst/J Cybern 4(1):95–104
61. Esogbue AO (1986) Optimal clustering of fuzzy data via fuzzy dynamic programming. Fuzzy Sets Syst 18(3):283–298
62. Féron R (1979) Sur les notions de distance et d'ecart dans une structure floue et leurs applications aux ensembles aléatoires flous. CR Acad Sc Paris A 289:35–38
63. Ferraro MB, Colubi A, González-Rodríguez G, Coppi R (2011) A determination coefficient for a linear regression model with imprecise response. Environmetrics 22(4):516–529
64. Ferraro MB, Giordani P (2017) Possibilistic and fuzzy clustering methods for robust analysis of non-precise data. Int J Approx Reas 88:23–38

65. Ferraro MB, Vichi M (2015) Fuzzy double clustering: A robust proposal. In: Grzegorzewski P, Gagolewski M, Hryniewicz O, Gil MA (eds) Strengthening Links Between Data Analysis and Soft Computing. Springer, Cham, pp 225–232
66. Fréchet M (1948) Les éléments aléatoires de nature quelconque dans un espace distancié. Ann L'Inst H Poincaré 10:215–310
67. Frigui H, Krishnapuram R (1996) A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. Patt Recogn Lett 17:1223–1232
68. Fritz H, García-Escudero LA, Mayo-Iscar A (2013) Robust constrained fuzzy clustering. Inform Sci 245:38–52
69. Gil MA (1992) Sufficiency and fuzziness in random experiments. Ann Inst Stat Math 44(3):451–462
70. Gil MA, Corral N, Gil P (1988) The minimum inaccuracy estimates in χ^2 tests for goodness of fit with fuzzy observations. J Stat Plan Infer 19(1):95–115
71. Gil MA, Jain P (1992) Comparison of experiments in statistical decision problems with fuzzy utilities. IEEE Trans Syst Man Cybern 22(4):662–670
72. Gil MA, López-Díaz M (1996) Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. Int J Approx Reas 15(3): 203–224
73. Gil MA, López-Díaz M, López-García H (1998) The fuzzy hyperbolic inequality index associated with fuzzy random variables. Eur J Oper Res 110(2):377–391
74. Gil MA, Lubiano MA, de la Rosa de Sáa S, Sinova B (2015) Analyzing data from a fuzzy rating scale-based questionnaire. A case study. Psicothema 27(2):182–191
75. Gil MA, Montenegro M, González-Rodríguez G, Colubi A, Casals MR (2006) Bootstrap approach to the multi-sample test of means with imprecise data. Comput Stat Data Anal 51:148–162
76. Giordani P (2010) Three-way analysis of imprecise data. J Multiv Anal 101:568–582
77. Giordani P, Kiers HAL (2006) A comparison of three methods for principal component analysis of fuzzy interval data. Comput Stat Data Anal 51:379–397
78. Gong MG, Su LZ, Jia M, Chen WS (2014) Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images. IEEE Trans Fuzzy Syst 22(1):98–109
79. González-Rodríguez G, Blanco-Fernández A, Colubi A, Lubiano MA (2009) Estimation of a simple linear regression model for fuzzy random variables. Fuzzy Sets Syst 160(3):357–370
80. González-Rodríguez G, Colubi A, Gil MA (2012) Fuzzy data treated as functional data: A one-way ANOVA test approach. Comput Stat Data Anal 56(4):943–955
81. Grzegorzewski P (1998) Statistical inference about the median from vague data. Contr & Cybern 27(3):447–464
82. Grzegorzewski P, Hryniewicz O (2000) Soft methods in statistical quality control. Contr & Cybern 29(1):119–140
83. Grzegorzewski P, Szymanowski H (2014) Goodness-of-fit tests for fuzzy data. Inform Sci 288(1):374–386
84. Hathaway RJ, Bezdek JC (2001) Fuzzy c-means clustering of incomplete data. IEEE Trans Syst Man Cybern Part B-Cybern 31:735–744
85. Hathaway RJ, Bezdek JC, Pedrycz W (1996) A parametric model for fusing heterogeneous fuzzy data. IEEE Trans Fuzzy Syst 4(3):270–281
86. Havens TC, Bezdek JC, Leckie C, Hall LO, Palaniswami M (2012) Fuzzy c-means algorithms for very large data. IEEE Trans Fuzzy Syst 20:1130–1146
87. Hébert PA, Denoeux T, Masson MH (2006) Fuzzy multidimensional scaling. Comput Stat Data Anal 51:335–359
88. Hryniewicz O (2006) Possibilistic decisions and fuzzy statistical tests. Fuzzy Sets Syst 157(19):2665–2673
89. Huang Z, Ng MK (1999) A fuzzy k-modes algorithm for clustering categorical data. IEEE Transn Fuzzy Syst 7:446–452
90. Hung W-L, Lee J-S, Fuh C-D (2004) Fuzzy clustering based on intuitionistic fuzzy relations. Int J Unc Fuzz Know-Based Syst 12:513–529
91. Hwang C, Rhee F (2007) Uncertain fuzzy clustering: interval Type-2 fuzzy approach to C-means. IEEE Trans Fuzzy Syst 15:107–120
92. Irpino A, Verde R, de Carvalho FAT (2017) Fuzzy clustering of distributional data with automatic weighting of variable components. Inform Sci 406–407:248–268

93. Jain A, Dubes R (1988) Algorithms for Clustering Data. Prentice-Hall, Upper Saddle River
94. Kesemen O, Tezel Ö, Özkul E (2016) Fuzzy c-means clustering algorithm for directional data (FCM4DD). Exp Syst Appl 58:76–82
95. Körner R (2000) An asymptotic α-test for the expectation of random fuzzy variables. J Stat Plan Infer 83(2):331–346
96. Krishnapuram R, Keller J (1993) A possibilistic approach to clustering. IEEE Trans Fuzzy Syst 1:98–110
97. Kruse R (1984) Statistical estimation with linguistic data. Inform Sci 33(3):197–207
98. Kruse R (1987) On a software tool for statistics with linguistic data. Fuzzy Sets Syst 24(3):377–383
99. Kruse R, Held P, Moewes C (2013) On fuzzy data analysis. In: Seising R, Trillas E, Moraga C, Termini S (eds.) On Fuzziness – A Homage to Lotfi A. Zadeh, Volume 1. Series Studies in Fuzziness and Soft Computing, Vol. 298. Springer, Heidelberg, pp 343–347
100. Kruse R, Meyer KD (1987) Statistics with Vague Data. Mathematical and Statistical Methods. Series Theory and Decision Library B, Vol. 6. D Reidel Pub Co, Dordrecht
101. Kwakernaak H (1978-1979) Fuzzy random variables, Part I: Definitions and theorems. Inform Sci 15:1–15; Part II: Algorithms and examples for the discrete case. Inform Sci 17:253–278
102. Laviolette M, Seaman JW, Barrett JD, Woodall WH (1995) A probabilistic and statistical view of fuzzy methods. Technometrics 37(3):249–261
103. Lee M, Pedrycz W (2009) The fuzzy C-means algorithm with fuzzy P-mode prototypes for clustering objects having mixed features. Fuzzy Sets Syst 160:3590–3600
104. Lertworaprachaya Y, Yang Y, John R (2014) Interval-valued fuzzy decision trees with optimal neighbourhood perimeter. Appl Soft Comput 24:851–866
105. Lingras P, West C (2004) Interval set clustering of web users with rough k-means. J Intel Inform Syst 23:5–16
106. Liu J (2010) Detecting the fuzzy clusters of complex networks. Pattern Recog 43:1334–1345
107. Liu S, Matzavinos A, Sethuraman S (2013) Random walk distances in data clustering and applications. Adv Data Anal Classif 7(1):83–108
108. Lubiano MA, de la Rosa de Sáa S, Montenegro M, Sinova M, Gil MA (2016) Descriptive analysis of responses to items in questionnaires. Why not using a fuzzy rating scale? Inform Sci 360:131–148
109. Lubiano MA, Gil MA (1999) Estimating the expected value of fuzzy random variables in random samplings from finite populations. Stat Pap 40(3):277–295
110. Lubiano MA, Montenegro M, Sinova B, de la Rosa de Sáa S, Gil MA (2016) Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. European J Oper Res 251(3):918–929
111. Lubiano MA, Salas A, Carleos C, de la Rosa de Sáa S, Gil MA (2017) Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data. Int J Approx Reas 88:128–147
112. Maharaj EA, D'Urso P (2011) Fuzzy clustering of time series in the frequency domain. Inform Sci 181:1187–1211
113. Miyamoto S, Ichihashi H, Honda K (2008) Algorithms for Fuzzy Clustering - Methods in c-Means Clustering with Applications. Springer, Berlin
114. Montenegro M, Casals MR, Lubiano MA, Gil MA (2001) Two-sample hypothesis tests of means of a fuzzy random variable. Inform Sci 133(1-2):89–100
115. Näther W (1997) Linear statistical inference for random fuzzy data. Statistics 29(3):221–240
116. Näther W (2006) Regression with random fuzzy data. Comput Stat Data Anal 51(1):235–252
117. Näther W, Albrecht M (1990) Linear regression with random fuzzy observations. Statistics 21(4):521–531
118. Nguyen-Trang T, Vo-Van T (2017) A new approach for determining the prior probabilities in the classification problem by Bayesian method. Adv Data Anal Classif 11(3):629–643
119. Okuda T, Tanaka H, Asai K (1978) A formulation of fuzzy decision problems with fuzzy information using probability measures of fuzzy events. Inform Contr 38:135–147

120. Parchami A, Taheri SM, Mashinchi M (2009) Fuzzy p-value in testing fuzzy hypotheses with crisp data. Stat Pap 51(1):209–.226
121. Pedrycz W (1998) Shadowed sets: representing and processing fuzzy sets. IEEE Trans Syst Man Cybern Part B-Cybern 28:103–109
122. Pedrycz W, Bezdek JC, Hathaway RJ, Rogers GW (1998) Two nonparametric models for fusing heterogeneous fuzzy data. IEEE Trans Fuzzy Syst 6(3):411–425
123. Pham DL (2001) Spatial models for fuzzy clustering. Comput Vision & Im Underst 84:285–297
124. Puri ML, Ralescu DA (1985) The concept of normality for fuzzy random variables. Ann Probab 11:1373–1379
125. Puri ML, Ralescu DA (1986) Fuzzy random variables. J Math Anal Appl 114:409–422
126. Ramos-Guajardo AB, Colubi A, González-Rodríguez G, Gil MA (2010) One-sample tests for a generalized Fréchet variance of a fuzzy random variable. Metrika 71:185–202
127. Ramos-Guajardo AB, Lubiano MA (2012) K-sample tests for equality of variances of random fuzzy sets. Comput Stat Data Anal 56:956–966
128. Rocci R, Vichi M (2005) Three-mode component analysis with crisp or fuzzy partition of units. Psychometrika 70(4):715–736
129. Ross TJ, Booker JM, Parkinson WJ (eds) (2002) Fuzzy Logic and Probability Applications: Bridging the Gap. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia
130. Ruan JH, Wang XP, Chan FTS, Shi Y (2016) Optimizing the intermodal transportation of emergency medical supplies using balanced fuzzy clustering. Int J Prod Res 54(14):4368–4386
131. Runkler TA, Bezdek JC (2003) Web mining with relational clustering. Int J Approx Reas 32:217–236
132. Ruspini EH (1969) A new approach to clustering. Inform Contr 15:22–32
133. Ruspini EH (1970) Numerical methods for fuzzy clustering. Inform Sci 2:319–350
134. Shan J, Cheng HD, Wang Y (2012) A novel segmentation method for breast ultrasound images based on neutrosophic l-means clustering. Med Phys 3:5669–5682
135. Silva L, Moura E, Canuto AMP, Santiago RHN, Bedregal B (2015) An interval-based framework for fuzzy clustering applications. IEEE Trans Fuzzy Syst 23:2174–2186
136. Sinova B, Gil MA, Van Aelst S (2016) M-estimates of location for the robust central tendency of fuzzy data. IEEE Trans Fuzzy Syst 24(4):945–956
137. Skala HJ (1975) Non-Archimedean Utility Theory. Series Theory and Decision Library, Vol. 9. D Reidel Pub Co, Dordrecht
138. Singpurwalla ND, Booker JM (2004) Membership functions and probability measures of fuzzy sets. J Amer Stat Assoc 99(467):867–877
139. Son LH (2015) DPFCM: A novel distributed picture fuzzy clustering method on picture fuzzy sets. Exp Syst Appl 42:51–66
140. Statistical Science Issue on Artificial Intelligence and Expert Systems. Stat Sci 2(1):3–44
141. Tamura S, Higuchi S, Tanaka K (1971). Pattern classification based on fuzzy relations. IEEE Trans Syst Man Cybern 1:61–66
142. Tan T, Suk HW, Hwang H, Lim J (2013) Functional fuzzy clusterwise regression analysis. Adv Data Anal Classif 7(1):57–82
143. Tanaka H, Uejima S, Asai K (1982) Linear regression analysis with fuzzy model. IEEE Trans Syst Man Cybern 12(6):903–907
144. Tanaka H, Watada J (1988) Possibilistic linear systems and their application to the linear regression model. Fuzzy Sets Syst 27(3):275–289
145. Theodorou Y, Drossos C, Alevizos P (2007) Correspondence analysis with fuzzy data: The fuzzy eigenvalue problem. Fuzzy Sets Syst 158:704–721
146. Tokushige S, Yadohisa H, Inada K (2007) Crisp and fuzzy k-means clustering algorithms for multivariate functional data. Comput Stat 22:1–16
147. Viertl R (2006) Univariate statistical analysis with fuzzy data. Comput Stat Data Anal 51(1):133–147
148. Wang D (2004) A note on consistency and unbiasedness of point estimation with fuzzy data. Metrika 60:93–104
149. Watanabe N, Imaizumi T (1993) A fuzzy statistical test of fuzzy hypotheses. Fuzzy Sets Syst 53:167–178

150. Wu H-C (2005) Statistical hypotheses testing for fuzzy data. Inform Sci 279:446–459
151. Wu K-L, Yang M-S (2002) Alternative c-means clustering algorithms. Patt Recogn 35(10):2267–2278
152. Yamashita N, Mayekawa S-I (2015) A new biplot procedure with joint classification of objects and variables by fuzzy c-means clustering. Adv Data Anal Classif 9(3):243–266
153. Yang M-S, Hwang P-Y, Chen D-H (2004) Fuzzy clustering algorithms for mixed feature variables. Fuzzy Sets Syst 141:301–317
154. Yang M-S, Nataliani Y (2017) Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. Patt Recogn 71:45–59
155. Yang M-S, Pan J-A (1997) On fuzzy clustering of directional data. Fuzzy Sets Syst 91:319–326
156. Zadeh LA (1965) Fuzzy sets. Inform Contr 8(3):338–353
157. Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23:421–427
158. Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Part 1: Inform Sci 8:199–249; Part 2: Inform Sci 8:301–353; Part 3: Inform Sci 9:43–80
159. Zadeh LA (1995) Discussion: Probability theory and fuzzy logic are complementary rather than competitive. Technometrics 37(3):271–276
160. Zadeh LA (2004) Comment: Membership functions and probability measures of fuzzy sets. J Amer Stat Assoc 99(467):880–881
161. Zadeh LA (2015) Fuzzy logic–a personal perspective. Fuzzy Sets Syst 281:4–20
162. Zhou J, Hung CC, Wang X, Chen S (2007) Fuzzy clustering based on credibility measure. In: Proc 6th Int Conf Inf Manag Sci, Lhasa, China, pp 404–411