EFDA–JET–CP(09)07/03

A. Murari, G. Vagliasindi, E. Arena, P. Arena, L. Fortuna
and JET EFDA contributors

On the Importance of Considering Measurement Errors in a Fuzzy Logic System for Scientific Applications in Nuclear Fusion

On the Importance of Considering Measurement Errors in a Fuzzy Logic System for Scientific Applications in Nuclear Fusion

A. Murari[1], G. Vagliasindi[2], E. Arena[2], P. Arena[2], L. Fortuna[2]
and JET EFDA contributors*

JET-EFDA, Culham Science Centre, OX14 3DB, Abingdon, UK

[1]*Consorzio RFX-Associazione EURATOM ENEA per la Fusione, I-35127 Padova, Italy.*
[2]*Dipartimento di Ingegneria Elettrica Elettronica e dei Sistemi-Università degli Studi di Catania, 95125 Catania, Italy*
** See annex of F. Romanelli et al, "Overview of JET Results",*
(Proc. 22[nd] IAEA Fusion Energy Conference, Geneva, Switzerland (2008)).

Preprint of Paper to be submitted for publication in Proceedings of the
4th International Scientific Conference on Physics and Control
(1st September 2009 - 4th September 2009)

ABSTRACT
In Magnetic Confinement Nuclear Fusion, as in practically all fields of science, the measurements are affected by noise, which can sometimes be modelled with an appropriate probability distribution function. The results of the measurements are therefore known only with uncertainties which sometimes can be significant. In many cases the noise sources is independent from the system to be studied and the quantities to be measured. In this paper, a numerical approach to handle statistical uncertainties, due to an independent noise source, in a Fuzzy Logic System is developed. Numerical analysis and various tests with a benchmark show how the statistical error bars can be interpreted as an independent "axis of complexity" with respect to the fuzzy boundaries of the membership functions. The uncertainties in the inputs can be transferred to the output and handled separately from the system intrinsic fuzzyness. The main advantages of this independent treatment of the measurement errors are shown in the case of a binary classification task: the regime confinement identification in high temperature Tokamak plasmas. Significant improvements in the correct prediction rate have been achieved with respect to the classification performed without considering the error bars in the measurements.

1. INTRODUCTION
Fuzzy logic is a very powerful conceptual framework [1] to model complex systems, which present a level of uncertainty too high to be managed with traditional boolean logic. After proving very successful in a series of engineering applications [2], in recent years this form of multivariate logic has become increasingly accepted also in the scientific domain. Its potential has become quite clear even in Magnetic Confinement Nuclear Fusion, where it has provided very competitive results in various fields, ranging from prediction to identification and even physical modelling and interpretation. With regard to the first point, disruption predictors based on Fuzzy Logic have been developed using a wide database of JET discharges [3], improving both the rate of success and the understanding of the phenomenon. In terms of identification, the determination of the plasma confinement regime, the fact whether the plasma is in the L or H mode of confinement [4] has been a particularly interesting application; in this case fuzzy logic identifiers have been compared with all the other major alternatives on a Joint European Torus (JET) database and have provided very competitive results [5]. The L-H transition has also been analysed by generating a Fuzzy Logic System (FLS) automatically on the basis of Classification Trees built with the CART software package [6] and the resulting rules give significant intuitive understanding of this auto-organisation process [7].

On the other hand in Nuclear Fusion, and more generally in the natural sciences, the data available is often affected by significant error bars, which have to be properly taken into account to provide reliable results. These measurement uncertainties are typically expressed in probabilistic terms and are therefore not naturally handled by the Fuzzy Logic intellectual framework. Indeed Fuzzy Logic is a sort of sophisticated multivariate logic, based on sets with fuzzy boundaries, and is therefore meant to be used to model systems, which are inherently characterised by these fuzzy boundaries between various states or configurations. This type of complexity is conceptually different from the errors induced by the measurement process, which introduces probabilistic uncertainties in the inputs and are usually quantified in statistical terms. Since the sources of noise can be very often assumed to be independent from the phenomena to be investigated, in this paper it is shown that the statistical uncertainties can be handled separately from the inherent fuzzy sets used to model the system under study. In this way the effects of the measurement errors can be isolated and given a specific treatment, while the power and richness of the fuzzy treatment are fully preserved. In the next section 2, the effects of the uncertainties in the measurements on the fuzzy inference process are analysed in the case of some simple but basic cases using a typical benchmark system found in the literature. In the following section 3, the same methods to quantify the effects of the measurement errors are applied to a classification problem in Nuclear Fusion, the confinement regime identification in Tokamak plasmas using signals of JET database [8]. The proposed numerical treatment improves the performance of the Fuzzy Logic classifiers and therefore makes them more robust versus the measurements errors. Further developments of the approach and potential interesting new applications are then reviewed in the last section.

2. THE STATISTICAL UNCERTAINTIES DUE TO THE MEASUREMENTS AS AN INDEPENDENT AXIS OF COMPLEXITY

In all the scientific domains the performed measurements are not perfect and their results are only known within specific error bars. In the case of complex physical phenomena modelled with FLSs, the fuzzy logic inference process must therefore be adequately upgraded to cope with the statistical uncertainties due to the measurement noise. To clarify this aspect, let's take as an example the height of people. In some contexts, like for example for the definition of a suitable diet, it can be inappropriate to divide people in the two simple categories of tall and small (see figure 1a) and a more sophisticated classification may be required. A possible membership function with fuzzy boundaries could be the one represented in figure 1b. This more involved membership is adopted to take into account the inherent complexity of the problem at hand and it does not have anything to do with the uncertainties in the measurements. Even in the case of ideal, perfect measurements of people height, a modelling with fuzzy sets could be required.

In the treatment adopted in this paper, the errors in the measurements introduce another dimension of complexity. Not only any input belongs to various sets with a different degree of membership, but also that degree of membership is not absolute but must be expressed by a probability.

This additional dimension in the complexity of the problem is shown pictorially in figure 2, where the two horizontal axes represent the traditional fuzzy boundaries between sets and the third, vertical one the probabilistic distribution of the membership values due to the measurement errors. It is also worth mentioning that this additional axis of complexity is not directly related to the fuzziness of the mfs, as it is the case in type-2 fuzzy systems [9]. In the problems discussed in this paper, the uncertainties affect directly the inputs and are independent from the FLS and its properties.

The error bars in the measurements can affect the operation of a FLS in different ways depending on the form of the membership functions (mfs) and the actual values of the measurements. The three fundamental possibilities are reported in figure 3. In certain eventualities, the uncertainty due to the measurement noise can be such that the input remains in a flat region of the membership function and therefore the output of the FLS is the same as if the noise was not present (top case of figure 3). A second situation is encountered when the noise can bring the input in a region where the same membership function is involved but with a different value (middle case in figure 3).

This of course would change the output of the FLS in a way which depends on the details of the membership functions (mfs) and the fuzzy rules. A last alternative arises when the noise combines with the form of the mfs in such a way as to even involve other membership functions (bottom case in figure 3), which also affects the output of the FLS.

To illustrate the impact of the measurement errors on FLSs, a simple fuzzy system well known in the literature has been used as a benchmark [10]. This FLS was developed to simulate the tipping habits of customers in a restaurant. Only two quantities, the quality of food and service, are supposed to have an influence on the customer tip, which is the output variable. The mfs of the two inputs and the single output are shown in figure 4.

The fuzzy rules implemented are:
1. If **service** is poor then **tip** is cheap
2. If **service** is excellent then **tip** is generous
3. If **service** is good and **food** is rancid then **tip** is cheap
4. If **service** is good and **food** is delicious then **tip** is generous
5. If **service** is good and food is excellent then **tip** is generous

In many cases typical of the exact sciences, even if the noise cannot be reduced at will, it is very often possible to derive detailed information about its probability distribution function. A typical realistic assumption, for example, is for the noise to obey Gaussian statistics. This additional information can help to handle the impact of the noise on the operation of a FLS. To this end it must be remembered that, in the interpretation proposed in this paper, the statistic of the noise remains completely independent from the fuzzy mfs and the fuzzy rules. Therefore, if the probability distribution of the actual inputs can be calculated, the corresponding distribution of the outputs can be derived numerically. Since the FLS can be interpreted as a nonlinear transfer function, the output calculated assuming perfect measurements in general does not coincide with the peak of the output probability distribution function. When this is the case, the value calculated without noise does not represent the most likely value of the output and can therefore be misleading.

As an illustrative example, this aspect has been investigated for the case of the benchmark previously described. For simplicity sake, the noise has been assumed to have a uniform probability distribution within the error bars but this does not affect the quality of the results in any significant way. For the cases discussed in the following, the two inputs, food and service, have considered known with an uncertainty of 40%. The probability of the FLS outputs has been tested numerically. The inputs have been scanned and, for each individual value, one hundred points have been selected, within the confidence interval, and then randomly chosen with equal probability. Therefore for each couple of inputs, a statistics of ten thousand combinations has been analysed. The output values have been determined with the FLS and compared with the single value obtained neglecting the noise and assuming a unique, exact value for each input. In general the output of the FLS without noise can be very different from the most likely one, once the perturbation due to the noise is taken into account. This discrepancy is general and has been confirmed using different distribution functions of the noise.

The impact of the noise in the inputs is easy to quantify in the case of binary classification problems. To this end, it has been assumed that the output value of 0.76 indicates the propensity of the customer to give a tip; for values below this threshold the costumers tip cheaply and above it they tip generously. Again it has been assumed that the inputs are affected by an uncertainty of 40% and the same number of one hundred test points has been chosen as in the previous case. Both inputs have integer values in the range [1,7] and all possible combinations have been taken in account, producing a total of 49 outputs. Now for each of the input combinations, the average of the ten thousand outputs has been calculated to determine whether it is below or above the threshold of 0.76. This average value has then been compared with the one obtained assuming the input to be perfectly known. The difference between the decisions to tip or not for the the cases with and without noise is plotted in figure 5.

The main added value of the treatment described in this section is that a probability can be attributed to each output of the FLS. The interpretative powers of both fuzzy logic and probability can therefore be combined. The outputs are calculated according to the rules of fuzzy logic but their probability can also be determined on the basis of the statistics of the noise. The additional representational power of this combined approach can be very useful in the exact sciences as illustrated by the application to nuclear fusion described in the next section. The approach of computing only a limited number of strategically chosen activation values of the rules must be adopted with extreme caution because it can provide significantly distorted outputs.

3. A CASE STUDY: CONFINEMENT REGIME IDENTIFICATION IN JET

The methodology described in the last section has been applied to the case of confinement regime identification at JET. The main objective of the analysis is a typical classification problem: to determine whether a discharge is in the L or H mode of confinement for every point in time.

In Magnetic Confinement Fusion (MCF), plasmas of hydrogen isotopes are heated to high temperatures in order for the nuclei to overcome the Coulomb barrier and produce a sufficient rate of fusion reactions. To increase the net energy output, the density and confinement time of the plasma have to be increased as much as possible at temperature of tens of keV's. Today the most studied configuration in the world is the Tokamak, a Russian acronym for "Magnetic Bottle". In the ASDEX device it was then discovered in 1982 that, increasing sufficiently the input power and if appropriate conditions were met, the plasmas tended to transit spontaneously from the low confinement mode (the L mode) to an enhanced confinement mode called the High confinement or H-mode [4]. The H-mode is characterized by the presence of a thin region of very low transport situated at the edge of the plasma. Steep gradients in the density and temperature profiles are observed across this region.

To study this phenomenon, a database of 53 discharges has been verified by experts at JET, who have determined specifically the times of the L to H and H to L transitions for each discharge. The most relevant signals to be used have been derived with exploratory analysis techniques and in particular with the CART algorithm [5] within a wider class indicated by the experts. In the end, the six signals summarised in table I have been retained for the purpose of studying the LH transition. They are the most relevant ones for the database, whose results are presented in this paper and their information content is estimated to be about two orders of magnitude higher than the next most important one (the first one discarded).

In order to investigate the potential of the proposed approach to treat the error bars, a previously automatic method to generate FLSs from CART trees has been used [7]. The mfs adopted are of the general form shown in figure 6 and they are particularised for each of the relevant variables of table I. This choice of the mfs is a good trade-off between complexity and interpretability of the final results. The number of rules equals the number of terminal nodes of the chosen CART tree. Therefore selecting the number of nodes to consider implies fixing the complexity of the final FLS. Indeed, each rule is determined automatically from the inequalities at each node of the tree branch leading from a terminal node to the root. In table II the number of terminal nodes selected for each tree is reported in the first column to provide information about the complexity of the corresponding FLSs. All the details of the original procedure devised to directly obtain the FLSs from the CART tree are provided in [7]. For the purpose of the discussion in this paper, the most important point to retain is that the results do not depend on the details of the FLSs but they are really due to the uncertainties in the inputs (the error bars).

With the automatic technique just described, a series of Fuzzy classifiers of different level of complexity have been derived and their success rate verified. A first classification has been performed without considering the measurement errors. The FLSs have been trained with a random set of 38 discharges and then tested on the remaining ones (not used for the training). For each discharge, 200 time slices, one every ms, are available and therefore the FLS has been trained with 7600 examples. For each FLS, the best threshold has then been determined in an empirical way by choosing the value which provided the best performances.

The same training procedure has then been performed taking into account the uncertainties in the measurements. The assumed error bars are also reported in table I. Two different estimates of the error bars have been considered. Within the error bars, four different values of the signals have been selected and then the combinations of the various inputs have been randomly chosen. Given the complexity of the FLSs under test and the number of inputs, handling more alternatives is prohibitive in terms of computational time required. For each combination of these inputs the value of the output has been calculated. The plasma is then assumed to be in the state of the higher probability, i.e. in the state (L or H) which is more likely on the basis of the statistical distribution of the outputs. The results of the two categories of FLSs, with and without the error bars, are compared in table II for the time interval [−100 ms, 100ms] around the transition. The first column in this table reports the number of nodes in the CART tree which have been retained in the automatic construction of the FLSs. A higher number of nodes implies higher complexity in the structure of the FLSs [7].

A significant improvement in the success rate is always achieved by taking into account the most likely output instead of the simple value calculated assuming perfect measurements. In some cases the success rate of the classifier increases of more than 12% points. Moreover the positive effect of taking into account the statistics of the noise is significant also for the most performing FLS. It is important to note that this interval [−100ms, 100ms] just around transition is the most challenging from the point of view of the classification and therefore it is the most vulnerable to noise.

This is confirmed by the analysis of time intervals further away from the transition time, for which the improvement in the performance is lower because the influence of the noise is expected to be less critical. It is also worth noticing that the improvement is in general more evident for higher level of the assumed noise on the input signals as shown by the results of case two. On the other hand it has been verified that the increase in the success classification rate tends to saturate if the assumed noise becomes unrealistically high. This suggests a possible way to determine whether the error bars assumed are correct by increasing the noise level up to the point where the performance of the FLS does not improve.

For some of the most performing FLS the same test has also been carried out for a higher number of input combinations corresponding to different noise values, in order to increase the statistical basis of the results. In general increasing the number of input values of the signals has a positive effect on the classification capability of the FLS, because of course the statistical basis for the decision becomes more solid. On the other hand, already for less than ten different values of the inputs the computational time becomes prohibitive (several days).

4. PROSPECTS OF FURTHER DEVELOPMENTS

A specific numerical strategy has been tested to address the issue of taking into account the error bars in the measurements used as inputs to Fuzzy Logic Systems. The approach of considering the uncertainty of the measurements as an independent axis of complexity is conceptually sound for many real system and provides a clear improvement in the performance of the FL classifiers when the statistic of the noise is properly taken into account. Indeed the advantages of the adopted approach have been tested using a JET database to classify plasma depending on their mode of confinement.

These results confirm the importance of the noise for the time interval close to the confinement regime transition. Moreover the proposed analysis can provide an independent confirmation of the error bars estimated for the various measurements.

With regard to future developments, it would be nice to confirm the potential of the strategy for a concrete regression problem. Also the case of a FLS with more than one output could be analysed.

ACKNOWLEDGMENTS

This work, supported by the European Communities under the contract of Association between EURATOM and ENEA was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES

[1]. L.A. Zadeh, “Fuzzy sets”, Infor. Control, 8, pp.338-353, 1965

[2]. P. Bonissone, “Soft computing applications in Prognostics and Health Management (PHM)”, in Proceedings of the 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid, Spain September 21-24, 2008, pp. 751-756.

[3]. A. Murari et al Nuclear Fusion **48** (2008) 035010

[4]. Wagner, F. et al., Physical Review Letters **49**, 1408 (1982).

[5]. A. Murari et al. IEEE Transactions on Plasma Science, Vol. **34**, No. 3, June 2006

[6]. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J. 1984 Classification and Regression Trees (Belmont, CA: Wadsworth Inc.) (1993, New York: Chapman and Hall)

[7]. G. Vagliasindi, P. Arena, A. Murari, “CART data analysis to attain interpretability in a Fuzzy Logic Classifier” accepted to International Joint Conference on Neural Networks.

[8]. Y. Andrew, N.C. Hawkes, Y.R. Martin, K. Crombe, E de la Luna, A. Murari, I. Nunes, R. Sartori “Access to H-mode on JET and implications for ITER” accepted for publication in Plasma Physics and Controlled Fusion

[9]. http://www.type2fuzzylogic.org/publications/

[10]. © 1984-2007 The MathWorks, “Fuzzy Inference System”, Fuzzy Logic Toolbox of Matlab®.

SIGNAL NAME	Acronym	Unit	Uncert. Case 1	Uncert. Case 2
Magneto-hydrodynamic energy	Whmd	[J]	±15%	±20%
Toroidal Magnetic Field	BT80	[T]	±2%	±2%
Electron Temperature	Te	[eV]	±10%	±20%
Beta Normalized	Bndiam		±15%	±20%
X-point Radial Position	Rxpl	[m]	±0.01	±0.01
X-point Vertical Position	Zxpl	[m]	±0.01	±0.01

Table I: List of the signals used as predictors for the classification trees.

Term. Node Num.	Thr.	Perform without noise (%)	Uncert. CASE 1		Uncert. CASE 2	
			Perform (%)	Improv. (%)	Perform (%)	Improv. (%)
20	0.45	55.17	60.9	5.73	65.2	10.03
20	0.5	57.43	60.5	3.07	61.67	4.24
18	0.45	55.57	59.73	4.16	65.37	9.8
18	0.5	57.93	65.73	7.8	68.7	10.77
16	0.45	56.27	59.97	3.7	63.7	7.43
16	0.5	59.6	69.1	9.5	71.83	12.23
14	0.45	56.2	58.43	2.23	61.5	5.3
14	0.5	73.43	76.83	3.4	77.67	4.24
12	0.45	68.7	72.4	3.7	73.77	5.07
12	0.5	73.2	76.4	3.2	77.5	4.3
4	0.45	60.93	72	11.07	76.33	15.4
4	0.5	69.23	74.53	5.3	75.6	6.37
2	0.45	61.67	67.43	5.76	69.63	7.96
2	0.5	67.2	70.1	2.9	69	1.8

Table II: The performances of the various FLSs, with and without the noise on the inputs. Thr. indicates the threshold in the output adopted to discriminate between the L and H mode of confinement.

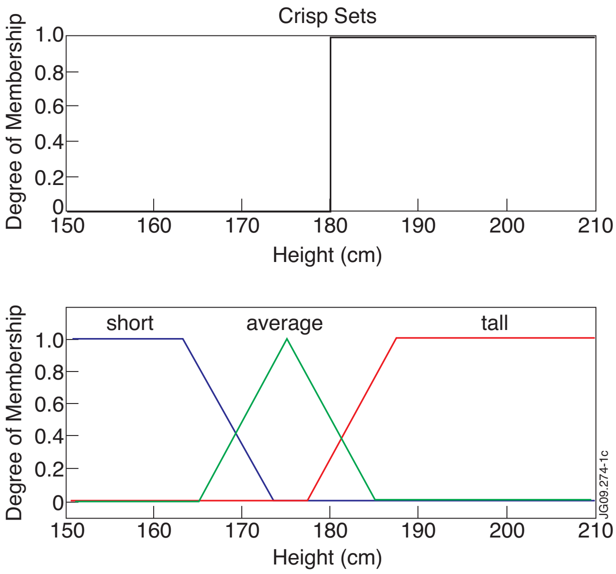

Figure 1: a) modelling people height with crisp sets b) modelling people height with fuzzy sets without any provision for uncertainties in the measurements.

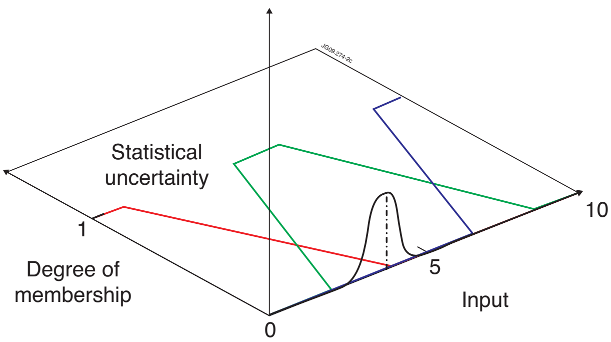

Figure 2: Statistical errors in the measurements of the inputs to the FLS create a second axis of uncertainty. The x axis represents the input value, the y axis the degree of membership, the vertical dimension the statistical uncertainties in the inputs and therefore the consequent probability distribution of the membership values.

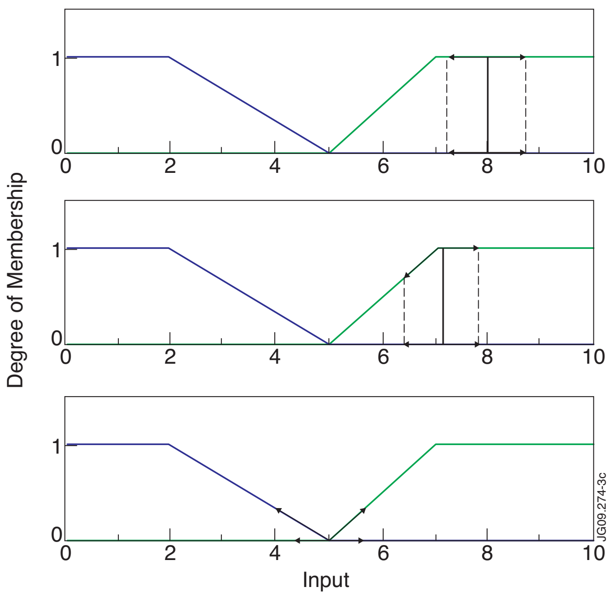

Figure 3: Three main alternatives encountered when adding noise to the inputs of a FLS.

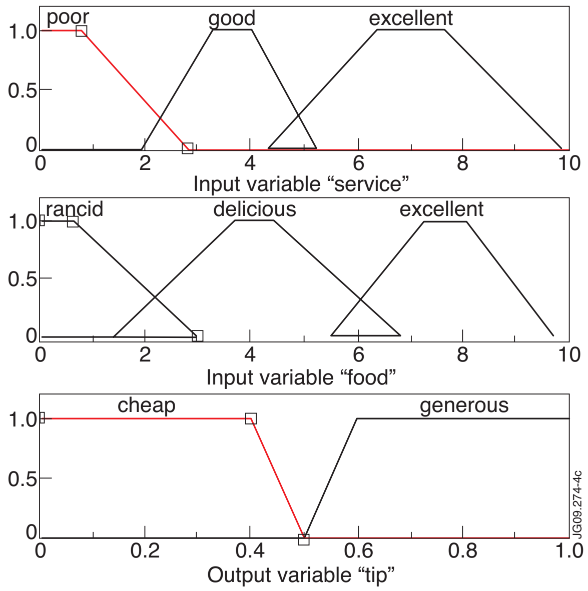

Figure 4: Membership functions of the inputs and the ouput of the simple model for the tipping behaviour of customers in restaurants. From top to bottom: a) the input food b) the input service c) the output tip.

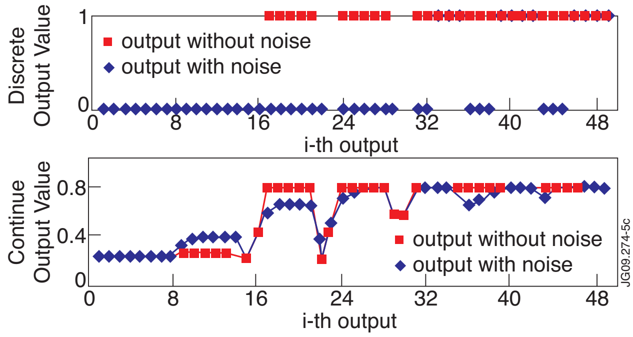

Figure 5: Top, output with binary classification; bottom, representation of the continuous output values. The value 1 indicates a generous tip and the value 0 a cheap tip. The difference in the classification of the tipping benchmark for the cases with and without noise can be observed. Indeed, there is a clear region of the input space where the most likely output for the noisy case is different from the one calculated without taking the noise into account. Form of the membership functions adopted in the automatic method to generate FLSs.

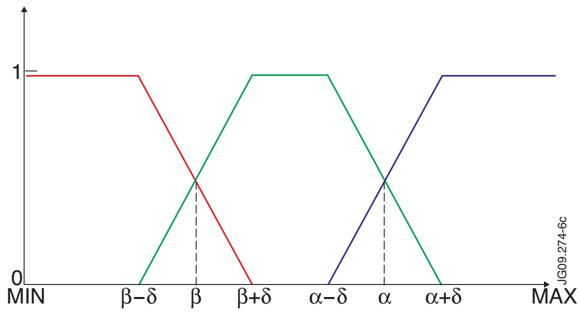

Figure 6: Form of the membership functions adopted in the automatic method to generate FLSs.